

Intersections and differentiations: a corpus-assisted discourse study of gender representations in the British press before, during and after the London Olympics 2012

Article

Accepted Version

Jaworska, S. ORCID: <https://orcid.org/0000-0001-7465-2245> and Hunt, S. (2017) Intersections and differentiations: a corpus-assisted discourse study of gender representations in the British press before, during and after the London Olympics 2012. *Gender and Language*, 11 (3). pp. 336-364. ISSN 1747-633X doi: 10.1558/genl.28858 Available at <https://centaur.reading.ac.uk/56337/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1558/genl.28858>

Publisher: International Gender and Language Association

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other

copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Intersections and differentiations: a corpus-assisted discourse study of gender representations in the British press before, during and after the London Olympics 2012

1 Introduction

Sport is undeniably one of the most pervasive areas of human activity spanning all social strata and widely present across media and education. Apart from being a form of entertainment and physical activity, it is a common cultural and symbolic resource deployed to aid group solidarity and identity formation (Meân and Halone 2010). Given that sport is so omnipresent in everyday life and intersects so many layers of society, it is an appropriate lens through which to study culture and social relations, including gender relations. As Beard (1998: 17) observes, if there are any gender issues in society, these are very likely to be reflected in sport.

While some studies show an increase in the amount of coverage of women in sport, and the 'disturbing' and challenging effect this can have on the traditional perception of sport as a male domain (Messner et al. 1993), this has not necessarily led to gender parity, but rather to different ways of constructing sportswomen and sportsmen discursively (Messner et al. 1993). Overt expressions of sexism seem nowadays to be largely absent from sport news and commentators appear to be more aware of gendered language (Duncan et al. 2005), but gendered hierarchies are still maintained, arguably through much subtler patterns of representations, which can be more pernicious than blatant sexism (Duncan 2006). Given that most people experience sporting events through media only, sports media are a powerful tool in disseminating such biases and influencing, directly and indirectly, public attitudes. Because journalists do not just reflect their own views, but also draw heavily on perceptions and ideas that they believe are widely accepted (Cotter 2010), studying representations as constructed through the language of sports media can reveal discourses and biases widely distributed in society. With a few exceptions (Meân 2001; Sznycer 2010; McDowell and Schaffner 2011; Aull and Brown 2013; Caple 2013), linguists have, to date, paid little attention to the role of sport discourse in the construction of gendered identities and biases. It is for these reasons that a number of sports scholars call for greater attention to the discourse of sports media (Kassing et al. 2004).

The present study responds to this call in that it aims to examine the impact of a global sports event on the discursive construction of gender representations in media reporting.ⁱ Whereas previous research on gender and sport has been mainly concerned with sports events in the North American or Australian context (e.g. Billings and Eastman 2002; Duncan et al. 2005; Bruce et al. 2010), this study investigates British media reporting surrounding the London Olympics in 2012. Also, most studies on sport and gender have focused on the representations of sportswomen and sportsmen during sports events and paid little attention to the impact on gender representations beyond these events. Given the current commitment of most international sports organisations including the International Olympic Committee (IOC)ⁱⁱ to promote gender equality in sport and beyond, examining the impact of sports events on gender representations seems a timely endeavour. By investigating patterns of gender representations before, during and after the London Olympics 2012, this study aims to assess whether this global sporting event had an impact on the construction of gender in media reporting and if so, what the nature of this effect was and whether it continued in subsequent reporting. Furthermore, most studies on the subject tend to treat gender as an isolated category, which does not adequately reflect the social 'reality' of gender, especially in sport. The notion of intersectionality reminds us that socially constructed categories such as gender cannot be studied in a one-dimensional way, as they are always bound with other social constructs, for example, ethnicity, age or nationality (Crenshaw 1991). Sport in

general, but global sports events in particular have traditionally emphasised masculinity, whiteness and nationalistic ambitions and hence, they present a highly relevant context in which to examine the intersections of gender, race and nation (Spracklen 2013; Billings and Angelini 2007). With the exception of Billings and Eastman (2002), there has been very little research investigating the effects of race and nation on gender representations in sports media. This study aims to go beyond the usual one-dimensional treatment of gender in sport and examines the extent to which gender representation intersects with nationalism and ethnicity. The specific questions which this research addresses are:

Q1: which gender representations are foregrounded and which are backgrounded or silent during major global sports events?

Q2: do national and racial identities intersect with gender representations and if so, what are the implications of such intersections?

Q3: do global sports events have an impact on the discursive ways in which gender representations are constructed in sports reporting, and, if so, what is the nature of this effect?

Most studies on the relation between sport and gender adopt the methodology of content analysis, which has been widely used since 1950s in studying mass media communication. Despite its many strengths, content analysis is generally conducted manually and it is based on categories that are normally set *a priori*. In this way, a study examining gender and using content analysis might be based on judgements about what is gendered and what is not, even before the data analysis (cf. Macalister 2011). Thus different researchers may arrive at different judgements when studying the same data sets. Some degree of subjectivity and *a priori* judgements are always involved in any kind of research and are impossible to completely remove (cf. Baker 2006). However, we can reduce the bias somewhat when we begin with the data and try to identify patterns and categories of representations as they emerge from the data *a posteriori*, as is possible with Corpus Linguistics methods.

A corpus-based approach was adopted by McEnery et al. (2013) to investigate the representations of disability and of UK, London especially, before, during and after the London Olympics. Gender was not of concern to this study. Similar to our undertaking, McEnery et al. (2013) were interested in examining the impact of the event, but the analysis was based on media data produced until December 2012. In the concluding remarks, the researchers suggest that in order to understand better the nature of the impact, a follow-up study one year on would be needed (McEnery et al.: 47). We follow this recommendation and investigate the representations of gender one year after the London Olympics.

Our study follows the approach of Corpus-Assisted Discourse Studies (CADS) which uses large data sets (corpora) and both quantitative and qualitative research procedures (Partington et al. 2013). Answers to Question 1 and 2 are offered by examining a large corpus of newspaper data that comprises all articles published in six major British national newspapers during the London Olympics 2012. To test the impact of this global sport event on the construction of gender representations (Question 3), data from two control corpora encompassing articles from the same publications with the same time span, but one year before and one year after the global event have been collected for comparison. By scrutinising frequency lists and studying collocational profiles of lexical items denoting gender, we provide attested linguistic evidence of salient discursive representations of gender and the impact of the sporting event on these.

This paper begins with a brief explication of our understanding of gender, followed by an overview of research on the relation between sport and gender, focusing especially on media reporting (Section 2). Section 3 outlines the research methodology and the data sets under scrutiny. Section 4 focuses on gender representations and intersections in media reporting on

the London Olympics 2012 by discussing the key results that emerged from the study. We conclude our paper with observations regarding the strengths and weaknesses of the methodology used.

2 Gender Representations in Sports Media

In studies of language and gender, neo-Marxist critical analysis and post-structuralism are the dominant theoretical approaches to the analysis of gendered texts. Significant differences between them include a movement from a tendency to use binary terms and thinking, for concepts like sex and gender, towards a recognition that these aspects of identity constitute a diverse set of performances with various social meanings (cf. Butler's performativity theory (1990, 1993)). A further important shift concerns the concept of sex, which is no longer viewed as innate or rooted in the physical; rather it is seen as socially constructed in much the same way as gender has been. While the terms 'social constructivism' is often associated with postmodern/poststructural approaches, Cameron (2005: 483) explains that "all feminist theory since Simone de Beauvoir's *The Second Sex* (first published in French in 1949) has been committed to the proposition that gender is socially constructed; while recent theorists may have interpreted that proposition in more radical ways than their predecessors, it is inaccurate to suggest that feminist social constructionism did not exist before the 1990s". As theoretical, and often political, positions in research into language and gender, neo-Marxism and poststructuralism resemble in practice more of a continuum than a binary choice, with many authors taking predominantly one view but appropriating elements of the other, or even failing to align themselves explicitly with one perspective but selecting useful elements from each, partly due to there being a distinct lack of unified views within the two approaches (Cameron 2005). In this paper, we take a social constructionist view of gender, from a generally critical stance and focusing on how the writers of these texts construct the world for the readers, but we do incorporate elements from post-structuralism, insofar as it helps us to illuminate how the texts under analysis have social effects via their representations of males and females. Our analysis of their representation, therefore, incorporates the assumption that gender and sex are socially constructed, and that the sexes and genders of the human beings represented in the media are not singular or homogenous, but fluid, and contextually responsive. More importantly, inspired by the notion of intersectionality (Crenshaw 1991) which is now increasingly adopted in linguistic research on language, gender and sexuality (e.g. Levon and Mendes 2016; Milani 2015), we view such constructions as multi-dimensional constructs influenced by many different social categories. It is our intention to demonstrate how intersections of gender with ethnicity and nationality are, in the context of sport reporting, purposefully deployed to create representations in support of certain hegemonic discourses and social orders.

Women have come a long way in sport. Despite the initial exclusion of women from international sports events – allegedly the founder of the first Olympic Games Pierre de Coubertin was of the opinion that women “should not soil the Olympic Games with their sweat” (Bruce et al. 2010: ix) - female participation in national and international tournaments has increased significantly since the 1950s. For example, at the 1964 Olympic Games in Tokyo female athletes accounted only for 13% of all participants. In 1984 in Los Angeles, the proportion rose to 23%, reaching 44% in 2012 at the London Olympics.ⁱⁱⁱ This process has been accompanied by marked changes in social attitudes towards women in sport. Pursuing a career in sport, for example, is now seen as a legitimate option for young women.

However, despite the increased visibility of sportswomen and a wider social acceptance of female participation in sport, the media have been rather reluctant to reflect this trend (Messner et al. 1993; Bruce et al. 2010). As research on gender in sports media reporting over

the last three decades has demonstrated, media across Western countries have been systematically underrepresenting female sports and sportswomen. An extensive overview of routine female coverage in the major newspapers published in 15 countries conducted by Bruce et al. (2010) reveals that on average sportswomen receive less than 10% of newspaper coverage. Such systematic underrepresentation of women is, in the view of the authors, yet another piece of evidence which ‘convincingly demonstrates that women do not matter in our culture’ (Bruce et al. 2010: 5).

Although women are routinely underrepresented in sports media, research concerned with female coverage in print media during international sport events, especially the Olympic Games, has found significant increases since the 1980s (Pemberton et al., 2004). Currently, women receive on average one third of the total Olympic coverage, which is still much less than men. However, research also indicates that female coverage is relative to the proportion of women participating in the national Olympic teams (Pemberton et al., 2004). Thus, some researchers have optimistically suggested that print media coverage of international sport events seems to be less marked by gender biases and appears to be more orientated towards performance focusing on ‘athletes as athletes’ (Pemberton et al. 2004: 95; Wensing and Bruce 2003). However, the increased coverage of female athletes during international sports events may not be necessarily due to conscious efforts to achieve gender parity in media reporting. As a study by Wensing and Bruce (2003) suggests, other factors, but especially nationalism, can result in a greater visibility of women. Participation in international sports events, which are largely driven by nationalistic ambitions, means that women are members of national teams and as such are also seen as symbols of national identity and indicators of ‘national’ performance. On examining the representations of the Australian athlete Cathy Freeman in the major Australian newspapers during the Sydney Olympic Games in 2000, Wensing and Bruce (2003) conclude that the nation’s sense of self may override the gender factor. The increase in female coverage during international tournaments might thus serve more nationalistic tendencies rather than gender equality.

Although there are some reasons for optimism, Bruce et al. (2010) advise against interpreting the raw coverage data as a sign of a more equitable treatment of women during Olympic Games. They argue that in addition to measuring the amount of coverage, researchers need to give more attention to the content of media reporting, especially the types of women’s sports in which female athletes are featured. Research in this area provides plenty of evidence for gender-biased representations. For example, Duncan et al. (2005) show that despite the increased female coverage on televised sport news during Olympic Games, the TV reporting tends to depict women mostly in disciplines traditionally seen as feminine or female-appropriate such as gymnastics or ice-skating. These disciplines focus more on aesthetics as opposed to strength and contact – attributes associated with male’s sports. Earlier research by Vincent et al. (2003) on photographic newspaper coverage in Britain, US and Canada points to the same tendency: here too the focus appears to be on ‘female appropriate’ sports. Thus, some researchers are sceptical about the increased female coverage in media, arguing that instead of achieving gender equality, it further perpetuates the ideology of biological difference which views women as suitable for certain sports only and denies them the possibility of demonstrating their prowess in others (Duncan et al. 2005; Bruce et al. 2010).

Although gender may be less marked during major sport events, there is still a compelling amount of evidence suggesting that media reporting perpetuates gender biases in more subtle ways. This has been evidenced by studies that examined the ways in which female and male athletes are depicted. For example, sportswomen are more frequently described by references to their age, emotions, dating habits or family – attributes that are rarely used to refer to sportsmen (Eastman and Billings 2002). Successful female athletes are also often compared

to famous sportsmen (*ibid.*). In contrast, successful male athletes are never compared to famous sportswomen, but to religious, mythical or fictional figures of power such as Jesus or Superman. Eastman and Billings (2002: 208) conclude that such comparisons place sportsmen several levels above sportswomen and fix the male athlete as the prototypical athlete. Early research by Daddario (1994) also shows that sportswomen are likely to be represented as sex objects or in heterosexual roles, pointing to the role of media in perpetuating the norm of compulsory heterosexuality.

Research discussed thus far points to biased gender representations in sport reporting. Although there is some evidence that journalists and sports commentators are aware of gendered language, and instances of overt sexist or negative judgments are rare (Duncan et al. 2005; Eastman and Billings 2002), gender biases are still perpetuated through the ambivalent ways in which female athletes are discursively constructed vis-à-vis male athletes (Wensing and Bruce 2003; Duncan 2006). Increased coverage during international sport events and positive descriptions of female performance are offset by descriptions that demean, eroticise or trivialise women's efforts and accomplishments (Duncan 2006). In the view of Wensing and Bruce (2003), this ambivalence appears to be the dominant framing technique in media coverage of sportswomen, only suspended during international tournaments when nationalism may override gender.

While these content analyses point to the significance of language in constructing and disseminating gender biases in sports media, most of the studies are not systematic linguistic investigations and thus, little is known about specific linguistic strategies and patterns underlying gender representations in this context. Despite the vast body of research on gender and language, there exist only a few linguistic studies that look at linguistic strategies adopted to construct gender identities in sport (Meân 2001, Szyycer 2010, McDowell and Schaffner 2011, Caple 2013). These linguistic studies focus mostly on the talk of the players themselves and provide invaluable insights into the ways in which gender is discursively constructed and contested in speech. With reference to written usage, Aull and Brown (2013) consider the representation of gender in the media coverage of two particular events in basketball matches in the USA, using a small corpus of just over 112 000 tokens, and conclude that the linguistic patterns they identify contribute to the 'othering' of women and female athletes. Using the social semiotic multimodal analysis, Caple (2013) investigates the text-image relations in the representations of female athletes during the regular everyday reporting in the Australian media and concludes that the sexualised and emotional images of sportswomen in earlier reporting has been replaced with discourse representing female athletes as elite sports player much in control of their emotions. However, to our knowledge, there is no linguistic study that systematically examines gender representations in the context of sport and in large collections of written texts, media texts especially. In contrast to previous and mostly content-analytic studies, we adopt the novel methodology of Corpus-Assisted Discourse Studies (CADS) and use a large corpus of newspaper data. In doing so, our study endeavours to provide attested linguistic evidence of salient discursive representations of gender in media reporting produced during the London Olympics 2012. Before addressing our specific research questions, the next section discusses in more depth the methodology and the data sets under scrutiny.

3 Research Methodology and Data

This study utilises the methodology of Corpus Assisted Discourse Studies (CADS) that has been increasingly used to examine media representations (Partington et al. 2013). In contrast to qualitative approaches to discourse, CADS utilises quantitative tools and methods developed in Corpus Linguistics. A CADS project is also normally based on a large amount of data, which is interrogated by studying word frequencies and collocations. In the simplest

terms, frequency can be defined as the number of times an item occurs in a corpus, where 'item' can include a word, word form, part of speech (if a corpus was annotated accordingly), keyword or cluster (Baker 2006). Frequencies can be gleaned from frequency lists that normally rank all items (words, keywords, etc.) by how often they occur. Such lists can be generated within seconds by concordancers such as the one used in this study, that is WordSmith Tools (Scott 2008). Frequency lists are useful tools in determining the focus of a given data set and are good pointers to features that are typical of a particular genre, register or domain. They also allow the researcher to notice regularities and lexico-grammatical patterns that are salient but not immediately visible to the naked eye. Collocations are understood as frequent or strong word associations retrieved automatically within a certain span, in which span is defined as the number of words to the left and to the right of the search term. Collocations are normally established by using a measure of statistical significance, of which the most widely adopted are Mutual Information (MI), Log-Likelihood and T-score. It needs to be noted that the tests yield slightly different results because they favour different types of words.^{iv} Word associations retrieved in such a way point to recurrent lexical preferences that are often not just a matter of individual choice, but reflect established practices and evaluations and are a means by which communities express, interpret and evaluate people and actions. For example, Stubbs (2002) shows how the word 'gossip' is mostly used with terms denoting or implying a female speaker, highlighting the stereotype that only women gossip. Frequent patterns of co-occurring lexical choices can, therefore, indicate how a phenomenon, group of people or event are persistently framed in discourse (Baker et al. 2013). Flowing from this, patterns in discourse function to foreground or background particular identities, depending on the frequency with which they are invoked and reiterated through reference to their main elements and clusters of related collocations, and the positive or negative values associated with each. Thus we see mainstream or hegemonic forms of femininity or masculinity as typically supported by frequent, positive reiteration resulting in foregrounding, while more peripheral conceptions of identity are backgrounded or omitted entirely via less frequent, and often less positive, reference.

While corpus linguistics tools and analytical procedures are a critical part of CADS's methodological repertoire, the approach extends the CL methodology by integrating techniques commonly associated with qualitative discourse analysis. Having said that, CADS is not focused on a specific type of discourse analysis: it is much more eclectic in nature, encouraging the researcher to draw on as many analytical tools, frameworks and contextual information as it may be required to obtain the most complete picture (Partington et al. 2013). Furthermore, CADS researchers prioritise a comparative approach. As Partington et al. (2013: 12) highlight: 'we are not deontologically justified in making statements about the relevance of a phenomenon observed to occur in one discourse type unless, where it is possible, we compare how the phenomenon behaves elsewhere.' Comparisons can take different forms. We can compare and contrast one discourse type in different sources, for example, national print newspapers vs. political speeches. We can also examine the same discourse produced under different circumstances, for example, during different events. Another type of comparison is a diachronic one, which involves studying a specific type of discourse or representations at different points of time (Marchi 2010; Taylor 2010). This, in turn, can reveal changes and shifts in representations. The latter is the type of comparison which this study follows.

CADS and similar approaches that combine Corpus Linguistics with (Critical) Discourse Analysis have been successfully used to study a variety of media discourses including discourse of science (Taylor 2010), morality (Marchi 2010), nationalism and language ideologies (Vessey 2013), the representations of Muslims (Baker et al. 2012) and immigrants (Taylor 2014), and general gender representations, including representations of gendered

language (Johnson and Ensslin 2007, Caldas-Coulthard and Moon 2010, Jaworska and Krishnamurthy 2012, Baker 2014, Hunt 2015). The impact of specific sport events on media representations of gender has not thus far been examined by using this methodology. Thus, this study adds yet another context to this growing body of research.

Our study is based on a large corpus of newspaper articles collected from the time of the London Olympics 2012: 27 July to 12 August 2012. The sources included the following four national British newspapers together with their Sunday publications: *Daily Mail/Mail of Sunday*, *Guardian/Observer*, *The Times/Sunday Times* and *The Sun*. The choice of national newspapers was grounded in the assumption that they are designed for mass audiences and can potentially shape the views of large numbers of people. In addition, because newspapers are driven by commercial goals, they often draw on popular discourses assumed to be widely shared in order to respond to and attract more readers (Cotter 2010). Studying media representations surrounding events can, therefore, offer considerable insights into how events are constructed and shed light on discourses, stances and ideologies that underpin such constructions (cf. Baker et al. 2013). The four national newspapers selected for analysis include the most read tabloid-type newspapers, that is, the best-selling tabloid *The Sun* and the middle-range tabloid *Daily Mail* with its Sunday publication. The two others, *The Guardian* and *The Times*, are the most read *quality* newspapers, often referred to as broadsheets.^v The selection thus contains two sets of newspapers that differ in focus and style and have a wider circulation (cf. Baker et al. 2013).

The data was retrieved using Nexis UK, which includes the full print versions of the four national newspapers except for photographs and classifieds. Instead of focusing solely on texts that were specifically about the London Olympics, the decision was made to retrieve all article types that were published in the four newspapers during this period. This allowed us to better assess the extent to which gender discourses are foregrounded and backgrounded across all types of news reports, both sport-related and others, in this context. In order to assess the impact of the London Olympics on gender representations, we created two control corpora encompassing articles from the same publications with the same time span, but one year BEFORE and one year AFTER the event. In doing so, we were able to evaluate the impact of this sports event on the discursive ways in which gender or gender identities are constructed in the media. Table 1 shows the corpus data. The corpora were interrogated using the software programme WordSmith Tools 5 (Scott 2008).

Table 1: Corpus Data

Corpus	BEFORE	DURING	AFTER	Words total
Daily Mail/Mail on Sunday	1,135,264	1,114,492	1,087,458	3,337,214
Guardian/Observer	1,484,057	1,758,387	2,121,963	5,364,407
Times/Sunday Times	1,783,546	2,285,999	2,112,689	6,182,234
Sun	597,676	618,901	616,911	1,216,577
Total	5,000,543	5,777,779	5,322,110	16,100,432

We begin our analysis by studying frequency lists retrieved from the three corpora. Since we are concerned with discursive representations, and content words are more likely to reveal aspects of discourse (Baker 2006), our focus is on the content words in these lists. Because personal pronouns are used to signify gender (cf. Pearce 2008), some of these are also included in the analysis. We specifically examine the differences and similarities in the frequencies of gendered pronouns and gendered words such as ‘woman’, ‘women’, ‘man’ and

‘men’ BEFORE, DURING and AFTER the London Olympics 2012. We also investigate the collocational profiles of ‘woman’, ‘women’, ‘man’ and ‘men’, and compare those across the three data sets. The collocations were calculated using t-score and the span was -5 and 5+.^{vi} In our analysis, we focus on the 20 strongest collocations with t-score greater than 2.0^{vii}. After studying and comparing the general patterns, we move to a case study of the representations of two particularly prominent sportspeople and Olympic champions Jessica Ennis and Mo Farah. These reveal interesting patterns of parallels but also intersections of gender with nationality and race. In line with CADS principles, additional sources of data including statistics and press interviews are drawn on to interpret and contextualise the results. The next section presents the major findings.

4 Results and Analysis

Table 2 shows the 20 most frequent content words and pronouns in all three subcorpora. At the first glance, the lists seem similar: they contain mostly the same words with comparable rankings. In the three data sets, the subject pronoun ‘he’ is either the first or the second most frequent item followed by the possessive ‘his’. The subject pronoun ‘she’ occurs high on the lists too, but always after the masculine pronouns. Overall, the frequencies of the gendered pronouns point to a greater proportion of masculine items and hence to a male bias. This is consistent with previous corpus research that revealed a significantly higher proportion of male than female terms in general language usage in English (e.g. Romaine

Table 2: the 20 most frequent content words and pronouns across the three subcorpora

BEFORE				DURING			AFTER		
Rank	Word	Freq.	NF ^{viii}	Word	Freq.	NF	Word	Freq.	NF
1	HE	33,356	667.0	I	39,572	684.9	I	38,610	725.5
2	I	29,579	591.5	HE	38,413	664.8	HE	38,610	725.5
3	HIS	26,763	535.2	HIS	30,403	526.2	HIS	30,403	586.4
4	THEY	18,299	365.9	THEY	20,387	352.9	THEY	21,206	398.5
5	YOU	16,179	323.5	HER	19,681	340.6	SAID	18,635	350.1
6	SAID	15,509	310.1	YOU	19,235	332.9	YOU	18,310	344.0
7	THEIR	15,235	304.7	SHE	18,614	322.2	WE	17,356	326.1
8	WE	14,827	296.5	SAID	18,265	316.1	THEIR	17,032	320.0
9	HER	12,120	242.4	WE	17,388	300.9	HER	13,757	258.5
10	SHE	11,316	226.3	THEIR	16,171	279.9	SHE	13,057	245.3
11	YEAR	9,753	195.0	MY	11,650	201.6	YEAR	11,816	222.0
12	MY	9,296	185.9	YEAR	10,716	185.5	MY	9,874	185.5
13	LAST	9,245	184.9	LAST	10,425	180.4	TIME	9,508	178.7
14	FIRST	7,969	159.4	FIRST	10,186	176.3	NEW	9,244	173.7
15	TIME	7,865	157.3	TIME	10,177	176.1	PEOPLE	8,466	159.1
16	NEW	7,463	149.2	LIKE	8,433	146.0	THEM	8,319	156.3
17	PEOPLE	7,212	144.2	YEARS	8,267	143.1	YEARS	8,048	151.2
18	THEM	7,000	140.0	HIM	8,036	139.1	HIM	7,916	148.7
19	HIM	6,864	137.3	THEM	7,852	135.9	ME	6,371	119.7
20	YEARS	6,429	128.6	OLYMPIC	7,672	132.8	OUR	6,142	115.4

2000; Sigley and Holmes 2002; Pearce 2008; Baker 2014) and also in the British media specifically (Caldas-Coulthard and Moon 2010).

When we compare the normalised frequencies of the female pronouns across the three subcorpora, a clear difference emerges (see Table 3). In the Before Corpus, ‘she’ occurs 226.3 times per 100,000 words. In the During-Corpus, the number rises to 322.2 and in the After Corpus declines again to 245.3. The same pattern applies to the use of the pronoun ‘her’. The difference between the Before and During Corpus is very significant, statistically speaking, for both pronouns at $p < 0,0001$ (for ‘she’ $LL = 899.26$, $df = 1$; for ‘her’ $LL = 888.47$, $df = 1$)^{ix}. Also, the difference between the DURING and AFTER corpus is statistically significant (for ‘she’ $LL = 577.01$, $df = 1$; for ‘her’ $LL = 324.82$, $df = 1$), implying that indeed the rise in the DURING corpus was considerable.

Table 3: Frequencies of gendered pronouns

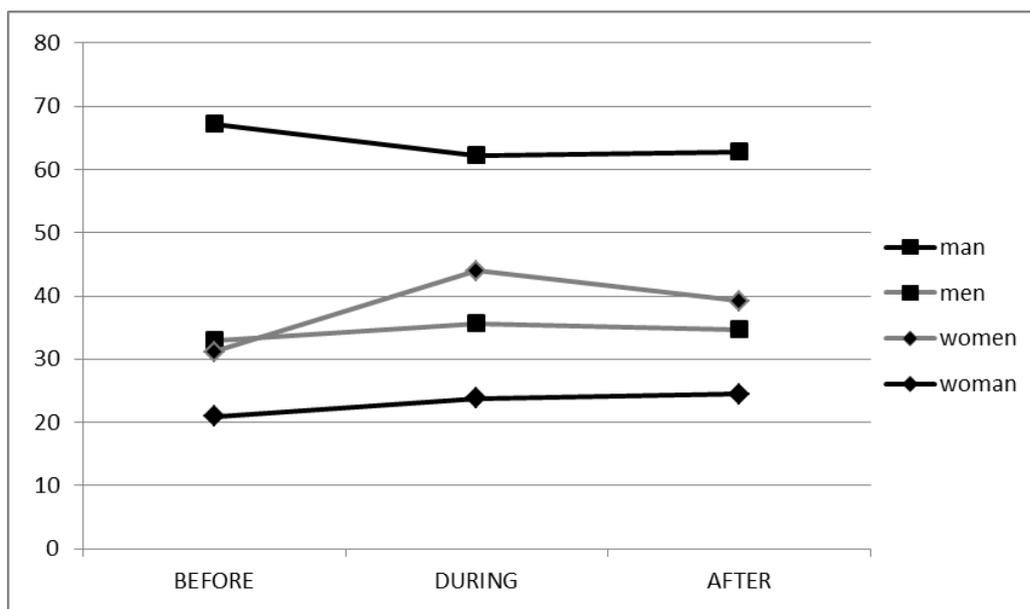
BEFORE			DURING			AFTER		
Pronoun	Freq.	NF	Pronoun	Freq.	NF	Pronoun	Freq.	NF
he	33,356	667.0	he	38,413	664.8	he	38,610	725.5
his	26,763	535.2	his	30,403	526.2	his	30,403	586.4
him	6,864	137.3	him	8,036	139.1	him	7,916	148.7
she	11,316	226.3	she	18,614	322.2	she	13,057	245.3
her	12,120	242.4	her	19,681	340.6	her	13,757	258.5

Although female pronouns are less frequent than their male equivalents in our data sets, the London Olympics time period coincides with an apparent increase in media attention paid to women. We tested this observation by examining the use of other gendered terms such as ‘woman’, ‘women’, ‘man’ and ‘men’. Here too, the same pattern emerged. Table 4 shows the frequencies of ‘woman’, ‘women’, ‘man’ and ‘men’ in each subcorpus. The singular ‘man’ is much more frequently used than the plural ‘men’ across the three sub-corpora. For ‘woman’ and ‘women’, the order is reversed, suggesting that women are more likely to be referred to collectively. This confirms a tendency observed in previous research on general English language use (Pearce 2008). It is the ‘collective’ usage of ‘women’ which has significantly increased in the During Corpora ($LL = 116.11$, $p < 0,0001$). It decreases in the After Corpus, but not to the same level as in the Before Corpus (see Figure 1).

Table 4: Frequencies of the items ‘woman’, ‘women’, ‘man’ and ‘men’

BEFORE			DURING			AFTER		
Word	Freq.	NF	Word	Freq.	NF	Word	Freq.	NF
woman	1,045	20.9	woman	1,377	23.8	woman	1,303	24.5
women	1,562	31.2	women	2,542	44.0	women	2,124	39.2
total	2,607	52.1	total	3,919	67.8	total	3,709	64.4
man	3,359	67.2	man	3,594	62.2	man	3,340	62.8
men	1,652	33.0	men	2,063	35.7	men	1,847	34.7
total	5,011	100.2	total	5,657	97.9	total	5,187	97.5

Figure 1: Distribution of ‘woman’, ‘women’, ‘man’ and ‘men’ (per 100,000 tokens)



Although the overall frequencies for male terms are substantially higher across all three data sets, the significant rise of female terms, especially the plural ‘women’ in the During Corpus indicates that media reporting during the London Olympics 2012 did indeed pay more attention to females. This, in turn, could suggest that the event may have had an impact on the representations of gender. In order to explore the nature of this impact, the most frequent collocations of the gendered terms ‘woman’, ‘women’, ‘man’ and ‘men’ were investigated and compared across the three data sets. Table 5 shows the 20 strongest collocations of the singular ‘woman’ and Table 6 those of the plural ‘women’.

Table 5: the 20 strongest collocates of ‘woman’ across the three subcorpora

BEFORE	DURING	AFTER
HER, SHE, YOUNG, OLD, YEAR, MAN, SAID, FIRST, BEAUTIFUL, HIS, LIFE, WORLD, OLDER, TOLD, LOVELY, MOTHER, ATTRACTIVE, AGED, BLACK, MARRIED	HER, FIRST, SHE, YOUNG, MAN, BRITISH, WIN, OLD, OLYMPIC, BECAME, YEAR, HIS, BECOME, ME, MARRIED, HISTORY, HE, LIFE, SAID, GOLD	HER, SHE, YOUNG, HE, MAN, OLD, HIS, SAID, YOU, CHILD, AVERAGE, TIME, PREGNANT, TOLD, LOOK, WORK, YEARS, MUSLIM, MARRIED, WHITE

Table 6: The 20 strongest collocates of ‘women’ across the three subcorpora

BEFORE	DURING	AFTER
MEN, THEIR, YOUNG, CHILDREN, SAID, AGED, NUMBER, THINK, OLDER, PREGNANT, LOOK, MAKE, GIRLS, AGE, BEAUTIFUL, MIDDLE, LOOSE, TIME, WANT, OUR	MEN, YOUNG, FIRST, BRITAIN'S, BRITISH, SAID, SHE, CAN, SPORT, TEAM, GAMES, OLYMPIC, FINAL, MAKE, HER, GOLD, HE, YOU, TIME, GROUP	MEN, YOUNG, CHILDREN, SAID, HE, CAN, WORK, NUMBER, SHE, HAIR, RAPE, VIOLENCE, PREGNANT, SAY, GIRLS, SEX, OLDER, BRITISH, GROUP, THREATS

As Tables 5 and 6 demonstrate, there are a number of similarities and differences in terms of the collocational profiles of the terms ‘woman’ and ‘women’ across the three data sets. Firstly, there are a number of collocates that seem to be consistently associated with both items, including terms which refer to males such as ‘men’ and ‘his’; female terms like ‘she’

and ‘her’, as well as the adjectives ‘young’ and ‘married’. ‘Young’ also collocates strongly with ‘man’ and ‘men’ (see Tables 7 and 8), and, hence, it seems that youth is foregrounded for both genders. ‘Married’, on the other hand, is a strong collocate for female referential terms, especially the singular ‘woman’, but not for male ones. This shows that women seem to be more strongly associated with marital status than men, which is also consistent with previous research (Pearce 2008).

However, during the London Olympics such associations seem weaker and new sets of collocates for female terms appear. For example, collocations pointing to appearance (‘beautiful’, ‘attractive’), reproductive status (‘pregnant’) and maternal responsibilities (‘children’) are more prominent in the Before and After Corpora than in the During Corpus. In contrast, in the During Corpus new collocates emerge including references to nationality: notably ‘British’, the possessive form of the country ‘Britain’s’ and items referring to sport and sporting success (‘first’, ‘gold’). These are not usually associated with the lemma ‘woman’ (cf. Pearce 2008). Some of these ‘new’ collocates (‘Olympic’, ‘win’, ‘gold’ ‘team’ and ‘history’) are also strongly associated with ‘man’ (see Table 7) and ‘men’ (see Table 8). Similar to men, then, in the During Corpus, women are represented as part of teams, as achieving sporting success and making history. Thus, it seems that the media reporting at the time of the Olympic Games had some positive influence on gendered representations in that the more traditional associations with ‘women’ were suppressed and replaced by elements referring to sporting success. Thus some parity between the representations of men and women is apparent.

Table 7: The 20 strongest collocates of ‘man’ across the three subcorpora

BEFORE	DURING	AFTER
HIS, CITY, YOUNG, OLD, YEAR, STAR, KEY, SAID, MATCH, MADE, HER, HIM, SHE, BEST, WOMAN, MY, FIRST, NEW, GREAT, GOOD	YOUNG, FASTEST, SAID, HIM, HER, WORLD, YOU, SHE, WOMAN, WIN, OLYMPIC, BECAME, BRITISH, GREAT, BEST, STAR, HISTORY, WHITE, RIGHT, TEAM	HIS, HE, OLD, YOUNG, HER, HIM, WOMAN, SAID, MATCH, CITY, ARRESTED, SQUAD, BEST, BIG, RIGHT, CALLED, BLACK, GOOD, FOUND, KNOWN,

Table 8: The 20 strongest collocates of ‘men’ across the three subcorpora

BEFORE	DURING	AFTER
WOMEN, YOUNG, MAD, HIS, HER, HE, GODS, BLACK, AGED, BRITISH, OLDER, SHE, MIDDLE, GROUP, NUMBER, MADE, ARMED, WHITE, ASIAN, ARRESTED	WOMEN, THEIR, YOUNG, THEY, FINAL, GOLD, BRITISH, MEDAL, SAID, HER, FINALS, TEAM, GROUP, OLYMPIC, BRONZE, WORLD, RACE, SILVER, GAY, FREESTYLE	MEN, WOMEN, THEIR, YOUNG, SAID, MADE, AGED, HIS, HER, SHE, THEM, MIDDLE, THINK, TIME, OLD, GAY, MAD, WAY, GET, WORLD

However, when we compare the collocates of ‘woman/women’ and ‘man/men’ in the During Corpus in more detail, our hopeful statement about parity needs to be qualified. For example, the collocations associated with ‘man’ include strongly positive modifiers (e.g. ‘great’), some of which occur in the superlative form (‘fastest’). These are absent from the list of female collocates, despite the fact that a number of female athletes broke world records during the London Olympics 2012, including the Jamaican runner Shelly-Ann Fraser-Pryce.

Another difference is reflected in the fact that ‘men’ collocates strongly with all three classes of medals, whereas ‘women’ only with gold, despite the fact that female athletes also naturally received silver and bronze. The British athletes received in total 69 medals, of which one third (24 medals) were won by female athletes. Women received overall 11 gold, 11 silver and 7 bronze medals. It appears that in the context of Olympic Games, women are more frequently referred to and framed positively when they win the highest possible prize (gold). This suggests that winning gold, for a female athlete, is seen as worthy of reporting, whereas lesser achievements are less likely to be recorded. The fact that men are more often associated with other types of medals could also be due to the fact that they received more medals overall.

In terms of the intersection of national identity and gender, the association with ‘British’ is stronger for ‘women’ (t-score 7.423) and ‘woman’ (t-score 7.164) than it is for ‘man’ (t-score 5.323) and ‘men’ (t-score 5.001). The plural ‘women’ also collocates strongly with the national term ‘Britain’s’ (t-score 7.832). Overall, it appears that the reporting about women in the context of Olympic Games represents females as members of a nation first, whereas this association is weaker for men. This pattern is consistent across the studied national newspapers. There are in total 61 occurrences of the collocate pair ‘British’ + ‘women’, of which 11 occur in the *Daily Mail/Mail on Sunday*, 5 in *The Sun*, 22 in the *Guardian/Observer* and 23 in *The Times/Times on Sunday*. The raw hits point to differences, which could be due to the different amount of tokens collected from each newspaper (see Table 1). Normalised frequencies per 100,000 show that the collocate is used with a fairly similar proportion in each source (*Daily Mail/Mail on Sunday* 0.9, *The Sun* 0.8, the *Guardian/Observer* 1.25 and *The Times/Times on Sunday* 1.0). The strong association between ‘British’ and ‘women’ comes into view when we do the reverse and study the strongest collocations of ‘British’ in the During Corpus. As Table 9 shows, ‘women’ is also a frequent collocate of ‘British’ alongside items pointing to the Olympics and various sporting disciplines. ‘Men’ appears on the list as well, but only in 42nd position.

Table 9: The 20 strongest collocates of ‘British’ in the During Corpus

BRITISH
TEAM, OLYMPIC, CYCLING, RECORD, ATHLETES, SWIMMING, MEDAL, SPORT, GOLD, AIRWAYS, WOMEN, ATHLETICS, ROWING, GOVERNMENT, GAS, SWIMMER, ECONOMY, BANKS, MEDALS, FANS

Of all the instances of the collocation pair ‘British’ + ‘women’, 53 refer specifically to the female athletes participating in the London Olympics. Figure 2 shows the first 10 examples of concordance lines with the collocation pair. As we can see, words in the vicinity of this combination point mostly to positive features and achievements, for example, British women are described as fabulous, as having things under control and as making history.

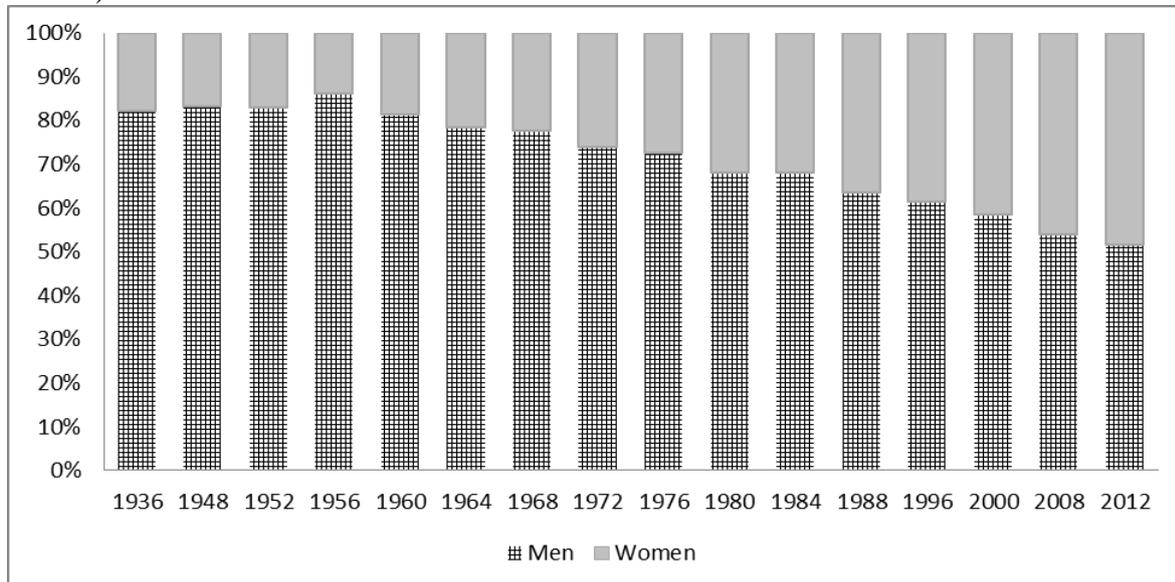
Figure 2: Examples of Concordance lines of the collocation pair ‘British’ + ‘women’

N Concordance

1 a country gave thanks to Heather Stanning and Helen Glover. The first British women ever to win Olympic gold in rowing had done so with
2 back the inexorable passage of time. And in purely national terms, the British women have been fab, medal-winners and also-rans and all. As I
3 to the skin and a brightness to the eyes ...that a new survey says that British women wear the highest heels in Europe. Every heel aficionado
4 track cycling Victoria Pendleton's retirement will leave a gap for speedy British women. Becky James from Abergavenny has showed the
5 time soon. Winning an Olympic medal is the headiest cocktail of all. BRITISH women have a distinguished pedigree in modern pentathlon.
6 opponents to row what Hosking later described as a "time trial". The British women had the race under control from a long way out, rowing
7 the history books as they triumphed by a whole length - the only British women ever to take Olympic rowing gold. Army officer Heather,
8 as Stanning and English pairs partner Helen Glover became the first British women rowers ever to be crowned Olympic champs. They
9 captain triumphed with former PE teacher Helen Glover as the only British women to ever take rowing gold at an Olympics. The jubilant pair
10 TASHA DANVERS (right) won bronze for GB in 2008, one of only two British women to have won a medal in the 400m hurdles at the Games.

By studying the representations of one female athlete, Wensing and Brunce (2003) concluded that an increase in positive female coverage during international tournaments is often ‘hijacked’ to celebrate the nation’s sense of self and thus, serves more to satisfy nationalistic aspirations. The above data confirms this pattern across a larger data set. Indisputably, the increase in the use of female terms and positive associations similar to those describing men points to some parity and thus, a positive impact of international tournaments on gendered representations. This is possibly due to the increased female participation in sport, especially the Olympics Games (see Figure 3). Whereas in the Berlin Olympics in 1936 there were only 37 British female athletes (17% of the team), in 2012 the number increased to 262 or 48% of the team. Historically, this is quite an achievement. However, the numbers are still far from equal. In the London Olympics, there were 1,216 more male than female participants.^x

Figure 3: The proportion of British female and male athletes in the Summer Olympics (1936 – 2012)



At the same time, such representations are, to borrow from Gabrielatos and Baker (2008: 11), ‘seasonal’. There seems to be little carry-through of the ‘gains’ to the following year, when the more traditional and stereotypic representations are again more salient. Also, the strong association of ‘women’ with ‘British’ suggests that, in the context of the Olympic Games, women, specifically gold medal winners, are overwhelmingly ‘used’ to celebrate nationalistic aspirations. As shown in a study by Jaworska (2016), women who ‘only’ received a bronze medal, were less frequently referred to and not strongly linked with being ‘British’.

The analysis has so far been concerned with representations surrounding gendered terms. In order to investigate whether these patterns framed the representations of sportswomen and sportsmen, we examined references used to describe the most frequently mentioned athletes in the During Corpus, namely the two British Olympic champions Jessica Ennis and Mo Farah.

We started the analysis by examining the most common naming practices centred on proper names and used to refer to the two athletes. The way proper names are used can reveal gendered practices. For example, Romaine (2000) suggests that, in contrast to men, women are more likely to be referred to by their first name, which creates a degree of familiarity and at times displays trivialising tendencies. We wanted to test whether such practice is prevalent when naming famous athletes such as Jessica Ennis and Mo Farah. We did so by checking all instances of the following combinations: first name + surname, surname only, first name only.

In the During Corpus, Jessica Ennis is referred to mainly by her surname or the combination of her first name and surname. There are only 70 instances of the use of the first name alone including the abbreviated form ‘Jess’ (see Table 10). A similar pattern can be observed in references to Mo Farah, who, like Ennis, is also mostly referred to by his surname and the combination of his first name and surname. Some researchers argue that the use of first names points to familiarity, which is less prevalent with men, who tend to be described by references to their surnames (e.g. Romaine 2000). In the context of references to these successful athletes in our data set, this pattern cannot be confirmed, as both Jessica Ennis and Mo Farah are mostly described by their surnames or a combination of the first name and surname. Interestingly, it is the male athlete who is more frequently referred to by his first name. Thus, naming practices of successful female and male Olympic champions seem to point to some parity. This results is also consistent with findings obtained by Caple (2013) who too found that female players were mostly refer to by their surnames.

Table 10: Naming practices (proper names) in the During Corpus

	Jessica Ennis	%	Mo Farah	%
Surname only	366	48.09	440	53.14
First Name + Surname	325	42.54	261	31.52
First Name only	70	9.55	127	15.34
Total	761	100	828	100

Subsequently, we examined the collocations occurring in the vicinity of ‘Ennis’ and ‘Farah’ – the two most frequent types of proper name references. Table 11 shows the 20 strongest collocates for each (only content words). As can be seen, both ‘Ennis’ and ‘Farah’ share a number of collocates, especially items pointing to their success (‘gold’, ‘won’, ‘Olympic’) as well as names of other British Olympic champions such as Greg Rutherford. But there are

Table 11: Collocates of ‘Ennis’ and ‘Farah’ in the During Corpus

ENNIS	FARAH
JESSICA, HEPTATHLON, FARAH, MO, GOLD, PERSONAL, OLYMPIC, SAID, BESTS, JUMP, GREG, RUTHERFORD, GIRL, STADIUM, OLYMPIC, HURDLES, CHRIS, BRITISH, WON, STORMED	MO, ENNIS, LAST, GOLD, SAID, WON, OLYMPIC, NIGHT, RUTHERFORD, GREG, BORN, FINAL, BRITAIN, CAME, GO, MOVE, BRITAIN’S, SOMALIA, DISTANCE, MOVED

some striking differences too. For example, ‘Ennis’ collocates strongly with the item ‘girl’ (19 times in total), whereas we have no instances of ‘boy’ in the vicinity of ‘Farah’. More than half of the instances of ‘girl’ in the vicinity of ‘Ennis’ are used in the phrase ‘poster girl’ (11 occurrences) reflecting the fact that Ennis was selected to represent the Team GB on its official placard. The other examples include a phrase ‘half-girl half-gazelle’ repeated 3 times in *The Sun* only and ‘golden girl’ used 5 times in *Sunday Times* and *The Sun*. These references suggest that Ennis is not seen as an ordinary ‘girl’, and her achievements are not necessarily possible for other females, though the combination of ‘golden’ and ‘girl’ might be used for alliterative effects as suggested by Caple (2013). Nevertheless, referring to an adult woman as a girl can be seen as patronising or trivialising (Romaine 2000). Previous corpus research has noted the tendency for the terms ‘woman’ and ‘girl’ to be used synonymously, whereas this does not seem to apply to ‘man’ and ‘boy’ (Taylor 2013). The association identified above echoes this tendency and supports a claim suggested by Bolinger (1980: 100) that ‘a female never grows up’, meaning that women do not achieve adult status but remain perpetual minors. This shows that certain traditional gender norms are maintained, even if the general picture seems to be more balanced.

The analysis has delivered results pointing to parallels and differences in the ways in which the female and male athlete were described. However, beyond this, the intersectionality of identity as represented in the data is also worthy of comment. The collocation lists contain ideologically significant types which show how nationality is also a particularly salient dimension of identity in the case of these two prominent athletes. Two collocation pairs are especially relevant: the association of ‘Ennis’ with the modifier ‘British’ (22 times, t-score 4.477) and that of ‘Farah’ with the noun ‘Britain’ (15 times, t-score 3.875). ‘British’ also occurs in the vicinity of ‘Farah’ (12 times), but is placed much lower on the list with a t-score of 3.031, whereas ‘Britain’ is not a collocate of ‘Ennis’. As Figure 4 and 5 show, the concordance lines of the patterns ‘Ennis + British’ and ‘Farah + British’ reveal notable differences. Jessica Ennis, who is of mixed-race origin (white and Jamaican) but was born and grew up in Britain, was clearly represented as a British athlete. We see this emphasised in phrases such as ‘British heptathlete’, ‘the best of British’ or ‘the British face’ (see Figure 4). Out of the 22 instances of the collocation pair Ennis + British 11 occurred in *The Times/Sunday Times*, 2 in *The Sun*, 6 in *The Guardian* and 3 in the *Daily Mail/Mail on Sunday*. Normalised frequencies per 100,000 point to almost the same distribution across the sources implying that each newspaper was likely to see Ennis as British (*The Times/Times on Sunday* 0.3, *The Sun* 0.3, the *Guardian/Observer* 0.3 and the *Daily Mail/Mail on Sunday* 0.3). In contrast, the representations of Farah are more varied or ambiguous. The concordance lines in Figure 5 show that the link between him and ‘British’ is not as obvious as it was in the case of Ennis. 4 instances of the collocation pair refer to other people or aspects such as ‘British public’, ‘British multiculturalism’ and ‘the British’. We also find here phrases that frame him clearly as an immigrant, for example, ‘the Somalian-born British’ or ‘imported British athlete’. Interestingly, Farah is associated with Britishness when mentioned together with other British athletes (for example Paula Radcliffe), which suggests that he is able to come in under the mantle of their Britishness when the team is mentioned as a group.

Figure 4: Concordance lines of the collocate pair ‘Ennis’ and ‘British’

N Concordance

1 taking the indoor title. There will be huge expectations of Ennis from the British public and she will be aware of the
2 refugee? - is lauded as the best of British and Jessica Ennis, draped in the Union Jack, jogged a lap of honour to
3 "The loss of her world title will, Chernova believes, spur on Ennis. A thumping victory - and a British record - by Ennis
4 has such consistent success been achieved. Last night, Ennis became the third British woman to climb to the top
5 by winning the greatest prize in women's multi-events. Ennis broke the British record for the second time this
6 are not being considered." But British heptathlon star Ennis insisted yesterday: "It's all about the medal and that
7 are all enjoying the Olympics. The big night with Jessica Ennis and the other British athletes was unbelievable. "We
8 equalled only by that given to British heptathlete Jessica Ennis moments earlier. Overcoming a slow start, the
9 awestruck, sofabound desire to weep along with Jessica Ennis, the British face of the Games, whose astonishing
10 , who is ranked 34th. Captions: Off to a flying start Jessica Ennis set a British record of 12.54sec in the 100m hurdles,

Figure 5: Concordance lines of the collocate pair 'Farah' and 'British'

N Concordance

1 track down this past week. But his pride in the performances of Farah, his imported British athlete, and Galen Rupp, the boy he
2 the games itself. The extent to which the British embraced Mo Farah, the son of refugees from Somalia, as one of their own,
3 young. But I am sure he will be amazing." In March last year Farah broke the British half-marathon record when he ran 60min
4 Ethiopians and the Kenyans had all run the 5,000m faster than Farah. British fans didn't need to be pessimistic to fear this was
5 10,000M At the end of a perfect day for British athletes, Mo Farah saved almost the best till last. It was touch and go. But
6 now so famous for its distance running that British stars Mo Farah and Paula Radcliffe spent time there last winter. Now 63,
7 Londoners," she added. SOMALIA Men's 10,000m final Mo Farah has been the epitome of British multiculturalism at
8 have enough." Patrick Wintour John Kampfner, page 31 >= Mo Farah, the British 10,000m gold medal winner, is expected to
9 next Games, there will be 17 million fewer starving children. Mo Farah, the Somali-born British 10,000m gold medallist, is
10 but he then burst into tears. As Somali-born British athlete Mo Farah closed in his gold medal for the 10,000 metres, a

The frame of Farah as an immigrant is reinforced by the other strong collocates such as 'born', 'Somalia' and 'came'. In fact, the phrase 'Farah who came to Britain' was repeated 18 times in the During Corpus. Out of the 18 instances 8 appeared in *The Times/Sunday Times*, 6 in the *Guardian/Observer*, 2 in *The Sun* and 2 in the *Daily Mail/Mail on Sunday*. Normalised frequencies indicate that the reference occurred with almost the same proportion across the four sources (*The Times/Times on Sunday* 0.3, *The Sun* 0.3, the *Guardian/Observer* 0.3 and the *Daily Mail/Mail on Sunday* 0.2). The same pattern applies to the collocation pair of 'Farah' + Somalia', which occurs 10 times in the During corpus, of which 5 times in *The Times/Sunday Times*, 3 in the *Guardian/Observer*, 1 in *The Sun* and 1 in the *Daily Mail/Mail on Sunday* (*The Times/Times on Sunday* 0.2, the *Guardian/Observer* 0.2, *The Sun* 0.2, and the *Daily Mail/Mail on Sunday* 0.1). Figure 6 shows all the instances of this collocation pair.

Figure 6: Concordance lines of the collocate pair 'Farah' and 'Somalia'

N Concordance

1 medals, including 25 gold, with two days still to go. Farah, born in Somalia but raised in London, will be
2 nature and nurture, a bit like The Tempest. Here's Farah, Somalia-born, with all the right genes to
3 competitive sport in schools. Try telling that to Mo Farah, who arrived from Somalia as a refugee and
4 nature and nurture, a bit like The Tempest. Here's Farah, Somalia-born, with all the right genes to
5 of the 10,000m podium. Though born in Somalia, Farah was brought up in London from the age of
6 up, this is where I started life." - Somalia-born MO FARAH is asked about representing Britain
7 at 8-13 Born in Mogadishu, Somalia, the young Farah lived what he has since described as a
8 have spent the time since trying to work him out. Farah, who was born in Somalia but moved to
9 in the same way as Ibrahim Jeilan did in Daegu. Farah, who was born in Somalia but moved to
10 let alone Yorkshire or London. Distance runner Mo Farah (born in Somalia), cyclist Bradley Wiggins

Mo Farah immigrated to Britain from war-torn Somalia at the age of eight, which was a significant event in his life. However, we may reasonably ask why this seemed to be of such importance in the context of his win during the London Olympics 2012. After all, Farah is a British citizen, who was trained and made his sporting career in Britain. As a counterpoint,

another Olympic champion Bradley Wiggins, who was born in Belgium, was not referred as Belgium-born even once in the *During Corpus*. In the run-up to the London Olympics 2012, the British middle-range and very popular tabloid *Daily Mail*^{xi} criticised the composition of the Team GB by describing some of its foreign born athletes as ‘plastic Brits’^{xii}. This term is pejorative and was used specifically to question the ‘suitability’ of foreign-born athletes to represent the British identity. Mo Farah took a stance against such accusations in a press conference following his 10,000m win, in which he, in response to a journalist’s question as to whether he would run for Somalia, famously said: “Look, mate, this is my country. This is where I grew up, this is where I started my life”^{xiii}. His response indicates that place of birth is not necessarily central to an individual’s sense of national identity, despite the close connection made between place of birth and national identity in our data set in reporting on Jessica Ennis and Mo Farah.

While the analysis shows a number of parallels in the representations of the female and male athletes, there are striking differences and intersections. The success of the female athlete is celebrated on a par with her male counterpart. At the same time, certain gendered representations, such as the association of Jessica Ennis with ‘girl’, are still preserved. Concurrently, her success is strongly framed in national(istic) terms: Jessica Ennis appears to be considered a representative of British identity possibly due to the fact that she was born in Britain and her mother was (white) British. She is presumably seen as less of a ‘plastic Brit’. On the other hand, Farah does not seem to share the same status. His immigrant background was frequently foregrounded implying some ‘resistance’ to accept Farah as a ‘truly’ British athlete.

5 Conclusions

The analysis of the use of gendered terms in the context of the London Olympics 2012 has shown an increase in attention paid to women during the event. Also, more balanced gender representations could be observed: the usual collocates of ‘woman’ and ‘women’ that point to appearance, marital and reproductive status were suppressed by more positive associations, also shared by the representations of men (cf. Caple 2013). This trend appears to point to greater gender parity and thus one may optimistically conclude that the London Olympics had a positive impact on gender representations in the media. This would also be in line with claims suggested in previous research (e.g. Wensing and Bruce 2003). However, a closer comparison of the collocations used to refer to women and men may caution against such optimistic conclusions. Firstly, there seems to be little carry-through of the ‘gains’ to the following years, highlighting that such representations are only ‘seasonal’ and do not have a long-lasting effect. Secondly, the positive female representations apply mainly to women who achieved the highest success, that is, gold medals, which in turn is ‘hijacked’ to celebrate nationalistic sentiments (cf. Wensing and Bruce 2003; Billings and Angelini 2007). Thus, in this context gender seems to intersect strongly with nationalism, which uses successful female athletes to reinforce the concept of the nation-state. In addition, the relation between sporting success, gender and nationalism is much more complex and nuanced. As the comparison of the representations of Jessica Ennis and Mo Farah has demonstrated, ethnicity and place of birth play a strong role in who might be considered ‘British’ and a representative of British identity. Despite the fact that the success of many foreign-born athletes was celebrated as a national success, there seems to be subtle ‘resistance’ to linking ‘British’ with those who do not share the identity characteristics of being white or having been born in Britain.

This study has also shown the benefits of using a comparative CADS approach to study gender representations. The quantitative, frequency-based analysis was a good point of entry

into the data revealing general salient and systematic patterns in the use of gender pronouns and gendered words before, during and after the London Olympics. The subsequent collocational analysis supported by examination of concordance lines demonstrated patterns of representations that would otherwise go unnoticed, and suggested seams of meaning for further, micro-level analysis. Without the collocation lists it would not have been possible to identify the subtle intersections of gender representations with nationalism and ethnicity. In summary, the trends in frequency and commonly associated meanings are incontrovertible and confirm the perpetuation of certain sexist and racist patterns of representation. These patterns in turn reflect and reinforce dominant discourses of social hierarchies in terms of race and gender which are, or should be, out of step with modern British society.

This study investigated only one global sports event. In order to assess the impact of such events on media representations of gender, further research would need to examine media reporting during other major international tournaments - including sports that are not traditionally associated with women, for example, football. Our research focused on the representations surrounding a limited selection of gendered terms. Future research would need to expand the analysis by studying other terms that connote gender or gender roles, for example, 'lady', 'girl' or 'mother'. Also, our analysis focused on the comparison between one female and one male athlete. To corroborate the relevance of the intersections of nation-state and ethnicity for gender representations in media reporting, it would be necessary to examine media constructions of a larger number of sportswomen and sportsmen from different national and ethnic backgrounds.

6 References

- Aull, Laura L, and Brown, David West (2013) Fighting words: a corpus analysis of gender representations in sports reportage. *Corpora* 8: 27-52.
- Baker, P. (2006) *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker, P. (2014) *Using Corpora to Analyse Gender*. London: Bloomsbury.
- Baker, P., Gabrielatos, C. and McEnery, T. (2013) *Discourse Analysis and Media Attitudes. The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Beard, A. (1998) *The Language of Sport*. London and New York: Routledge.
- Billings, A. and Eastman, S. T. (2002) Selective representation of gender, ethnicity, and nationality in American television coverage of the 2000 Summer Olympics. *International Review for the Sociology of Sport* 37: 351-370.
- Billings, A. and Angelini, J. (2007) Packaging the games for viewer consumption: Gender, ethnicity, and nationality in NBC's coverage of the 2004 Summer Olympics', *Communication Quarterly* 55: 95-111.
- Bolinger, D. (1980) *Language, the Loaded Weapon*. London: Longman.
- Bruce, T., Hovden, J. and Markula, P. (2010) *Sportswomen at the Olympics: A global comparison of newspaper coverage*. Rotterdam: Sense.

- Butler, J. (1990) *Gender trouble: Feminism and the subversion of identity*. London: Routledge.
- Butler, J. (1993) *Bodies that matter: On the discursive limits of 'sex'*. London: Routledge.
- Caldas-Coulthard, C. and Moon, R. (2010) Curvy, hunky, kinky: using corpora as tools for critical analysis. *Discourse & Society* 21: 99-133.
- Cameron, D. (2005) Language, gender and sexuality: Current issues and new directions. *Applied Linguistics* 26(4) 482-502.
- Caple, H. (2013) Competing for coverage: Exploring emerging discourses on female athletes in the Australian print media. *English Text Construction* 6 (2): 271-294.
- Cheng, W. (2012) *Exploring Corpus Linguistics: Language in Action*. London and New York: Routledge.
- Connell, R. W. (1995) *Masculinities*. Cambridge: Polity Press.
- Cotter, C. (2010) *News Talk: Investigating the Language of Journalism*. Cambridge: Cambridge University Press.
- Crenshaw, K. (1991) Mapping the margins: intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43: 1241-1299.
- Daddario, G. (1994) Chilly scenes of the 1992 winter games: the mass media and the marginalization of female athletes. *Sociology of Sport Journal* 11: 275-88.
- Duncan, M. C. (1990) Sports photographs and sexual difference: Images of women and men in the 1984 Olympic Games. *Sociology of Sport Journal* 7: 22-43.
- Duncan, M. C., Messner, M. A. and Willms, N. (2005) *Gender in televised sports: News and highlights shows, 1989-2004*. Los Angeles, CA: Amateur Athletic Foundation of Los Angeles, retrieved on 12 August 2015 from http://www.aafla.org/11pub/over_frmst.htm
- Eastman, S. T. and Billings, A. (2000) Sportscasting and sports reporting. The Power of gender bias. *Journal of Sport & Social Issues* 24: 192-213.
- Gabrielatos C. and Baker P. (2008) Fleeing, Sneaking, Flooding: A Corpus Analysis of Discursive Constructions of Refugees and Asylum Seekers in the UK Press, 1996-2005. *Journal of English Linguistics* 36: 5-38.
- Gries, S. (2010) Useful statistics for corpus linguistics. In A. Sánchez and A. Moisés (eds.) *A mosaic of corpus linguistics: selected approaches* 269-291. Frankfurt am Main: Peter Lang.
- Hunt, S. (2015) Representations of gender and agency in the Harry Potter series. In P. Baker and T. McEnery (eds.) *Corpora and Discourse Studies*. Palgrave Advances in Language and Linguistics Series, pp. 266-284. London: Palgrave Macmillan.

- Jaworska, S. (2016) Using a corpus-assisted discourse studies (CADS) approach to investigate constructions of identities in media reporting surrounding mega sport events: the case of the London Olympics 2012. In I. Lamond and L. Platt (eds.) *Critical Events Studies: Approaches to Research*. Palgrave Macmillan, London (in press).
- Jaworska, S. and Krishnamurty, R. (2012) On the F-word: a corpus-based analysis of the media representation of Feminism in British and German press discourse, 1990-2009. *Discourse and Society* 23: 401-431.
- Johnson, S. and Ensslin, A. (2007) But her language skills shifted the family dynamics dramatically. Language, gender and the construction of publics in two British newspapers. *Gender and Language* 1(2): 229-54.
- Kassing, J., Billings, A. C., Brown, R., Halone, K. K., Harrison, K., Krizek, B., Mean, L. J. and Turman, P. D. (2004) Enacting, (re)producing, consuming, and organizing sport. *Communication Yearbook* 28: 373-409.
- Levon, E. and Mendes, R. (eds.) (2016) *Language, sexuality and power: Studies in intersectional sociolinguistics*. Oxford: Oxford University Press.
- Macalister, J. (2011) Flower-girl and bugler-boy no more: changing gender representation in writing for children. *Corpora* 6: 25-44.
- McEnery, T., Potts, A. and Xiao, R. (2013) *London 2012 media impact study: Report for the Department for Culture Media and Sport*. London: Department for Culture, Media and Sport.
- Marchi, A. (2010) The moral in the story: a diachronic investigation of lexicalised morality in the UK press. *Corpora* 5:161-189.
- McDowell, J. and Schaffner, S. (2011) Football, it's a man's game: Insult and gendered discourse in The Gender Bowl. *Discourse & Society* 22: 547-564.
- Meân, L. (2010) Making masculinity and framing femininity: Fifa, soccer, and world cup web sites. In H. Hundley and A. Billings (eds.) *Examining Identity in Sports Media* 65-86. Thousand Oaks, CA: Sage.
- Meân, L. and Halone, K. (2010) Sport, language and culture: issues and intersections. *Journal of Social Psychology* 29: 253-260.
- Messner, M., Duncan, M. and Jenson, K. (1993) Separating the Men from the Girls: The Gendered Language of Televised Sports. *Gender & Society* 7: 121-137.
- Milani, T. (ed.) (2015) *Language and Masculinities: Performances, Intersections, Dislocations*. New York: Routledge.
- Partington, A., Duguid, A. and Taylor, C. (2013) *Patterns and Meanings in Discourse. Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: Benjamins.

- Pearce, M. (2008) Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine. *Corpora* 3: 1-29.
- Pemberton, C., Shields, S., Gilbert, L., Shen, X. and Said, H. (2004) A look at print media coverage across four Olympiads. *Women in Sport and Physical Activity Journal* 13: 87-99.
- Romaine, S. (2000) *Language in Society*. Oxford: Oxford University Press.
- Scott, M. (2008) *WordSmith tools. Version 5*. Lexical Analysis Software. Retrieved from <http://www.lexically.net/wordsmith/version5/index.html>
- Sigley, R. and Holmes, J. (2002) Looking at girls in Corpora of English. *Journal of English Linguistics* 30: 138-157.
- Spracklen, K. (2013) *Whiteness and Leisure*. Basingstoke: Palgrave.
- Stubbs, M. (2002) *Words and Phrases. Studies in Lexical Semantics*. London: Blackwell.
- Szyncer, K. (2010) Strong female speakers: The resistant discourse of tennis players. *Gender and Language* 7: 303-332.
- Taylor, C. (2010) Science in the news: a diachronic perspective. *Corpora* 5: 221-250.
- Taylor, C. (2013) Searching for similarity using corpus-assisted discourse studies. *Corpora* 8: 81-113.
- Taylor, C. (2014) Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics* 19: 368-400.
- Vessey, R. (2013) Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora* 8: 1-26.
- Vincent, J., Imwold, C., Masemann, V. and Johnson, J. T. (2003) A comparison of selected 'serious' and 'popular' British, Canadian, and United States newspaper coverage of female and male athletes competing in the Centennial Olympic games. *International Review for the Sociology of Sport* 37: 319-335.
- Wensing, E. and Bruce, T. (2003) Bending the rules: media representations of gender during an international sporting event. *International Review for the Sociology of Sport* 38: 387-396.

ⁱ We do acknowledge that other events might have had an impact on the representations of women. Having said that the London Olympics was the major event in Britain in that year and it received a wide coverage including front pages.

ⁱⁱ See the statement on the website of the International Olympic Committee at:

<http://www.olympic.org/news/inspiring-change-and-advocating-gender-equality-through-sport/227373>

ⁱⁱⁱ International Olympic Committee (2014). Factsheet Women in the Olympic Movement,

http://www.olympic.org/Documents/Reference_documents/Factsheets/Women_in_Olympic_Movement.pdf

^{iv} Baker (2006) offers a comprehensive discussion of the differences between these tests and their uses in corpus linguistics.

^v The selection reflects the circulation of the national newspapers as of July 2014, when the data was collected. The circulation statistics were obtained from the Audit Bureau of Circulation (ABC), <http://www.abc.org.uk/>

^{vi} There are over 20 different metrics that can be used to establish the strength of word associations. Currently, the most popular in Corpus Linguistics are Mutual Information (MI) and T-score (Gries 2010). It is known that MI score delivers results that favour low frequency words including idiomatic expressions or technical terms. To put it simply, MI is likely to give high collocation scores to word combinations that are rare or infrequent. In contrast, t-score returns high association scores for pairs of words that are frequent. The drawback is that the results also include function words (because they are frequent). Since we are interested in discourses that are frequently repeated in the media, the decision was made to use t-score as the collocation metrics. Because function words are not likely to tell us much about discourse, these were excluded from the analysis.

^{vii} It is assumed that t-score values greater than 2 are statistically significant (Cheng 2012: 94).

^{viii} The acronym NF stands for normalised frequencies. The base of normalisation used for this study was 100,000 words.

^{ix} The log-likelihood (LL) is a measure of statistical significance commonly adopted in corpus research. In contrast to other often use metrics, for example chi-square, LL does not assume that data is normally distributed and hence, it is seen as a more suitable measure for testing statistical significance of language items (see also McEnery and Hardie 2012). The LL value must be above 3.48 (df=1) for a difference to be statistically significant.

^x See 'Factsheet London 2012 Facts & Figures' published by the International Olympic Committee, http://www.olympic.org/Documents/Reference_documents/Factsheets/London_2012_Facts_and_Figures-eng.pdf

^{xi} The reach of the conservative stance promoted by the Daily Mail is large. The print version of the Daily Mail has the circulation of 1,437,541 making it the second most read print newspaper in the UK. Its online version *MailOnline* is now the world's number one online newspaper with over 11 million visitors per day (*Financial Times* 25.12.2014).

^{xii} "Team GB have 61 'plastic Brits' taking part in London Olympics", *Daily Mail*, 11 July 2012.

^{xiii} "Mo Farah also proved that we can cheer without stopping for 14 minutes", *The Guardian*, 11 August 2012.