



REPORT

Jessica Ennis, Mo Farah and Identity Language
in the British Press: A Case Study in Monitoring and
Analysing Print Media

Published: 11/12/2012



www.migrationobservatory.ox.ac.uk

Introduction and context

Since July 2012, the Migration Observatory has been building the framework for a Media Monitoring Project. Its aim is to improve understanding of the coverage of migration and related issues in the British press. We are gathering a comprehensive set of articles from Britain's national newspapers beginning in 2005 and to be continuously updated to the present on a weekly basis. These articles will include all mentions of migrants, refugees, and asylum seekers in British newspapers. From this large database (or 'corpus' of texts), we will get a sense not only of how much attention the press devotes to migration, but also the nature of coverage. This will include the general tone of coverage and the specific ways in which migrants are portrayed. We are interested in knowing, for example, if press is currently contributing to the widespread public perception of immigrants as asylum seekers (see the Migration Observatory report - Thinking behind the Numbers). This image may stem from high levels of asylum applications in the early 2000's, or it may be partly the product of continued media coverage even with asylum numbers declining. Of course, simply describing and monitoring press coverage does not demonstrate a connection to public perceptions, but it can help us determine whether or not such a connection is plausible.

The media project will also be designed to respond flexibly to other questions, including those raised by organisations working on migration or related issues, from a wide range of perspectives. In this document we present results from the first such effort. The Observatory was commissioned by the think-tank British Future to investigate media use of languages of identity and origins in association with Jessica Ennis and Mo Farah.

Ennis and Farah were among the most discussed and admired British gold medallists in the Games. While clearly they were discussed mainly as athletes, their racial, ethnic, and religious background and relationships to migration were sometimes a matter of public discussion as well. Ennis is the British-born child of a white British mother and father of Jamaican/Afro-Caribbean origins (thus sometimes referred to as 'mixed race', although this term like many racial categories is inherently difficult to define precisely and may or may not be frequently used as a self-description). Farah, meanwhile, was born in Somalia and came to Britain as a child. He is also known to be Muslim, whereas Ennis' religion does not appear to be a matter of public discussion. In the context of the London Olympics, a period widely thought to have produced an outpouring of national pride, their backgrounds seemed to figure in some discussions of the relationships among race, ethnicity, religion, national origins and British/English national identities.

The Migration Observatory was commissioned to attempt to quantify these trends in press coverage of both athletes, to help in discerning what sorts of identity language were most frequently used in connection with each of them. In particular, in commissioning this research, British Future were interested in finding out whether Ennis was described more in terms of her local origins (i.e. the 'girl from Sheffield') than her racial/ethnic background, and whether Farah described more as Somali-born than in terms of his more local origins after arriving in Britain as a child. Therefore, quantifying the presence of certain kinds of words in different types of coverage could help indicate the nature of discourses surrounding identity in British public life. The results presented below come from an analysis of the frequency of a set of identity-related words in press coverage mentioning Ennis and/or Farah. Although the words chosen were specified in advance by British Future to represent their hypotheses about the public identities of these two figures, the analysis was conducted independently by the Migration Observatory.

The analysis highlighted a few basic findings. In articles mentioning Ennis, her local origins in Sheffield were mentioned more frequently than her ethnic background, whether captured in terms of her father's origins in Jamaica or in racial/ethnic terms such as Afro-Caribbean, 'black', or 'mixed race'. In articles mentioning Farah, Somalia was indeed much more common than any local origin terms. Notably, explicitly racial or ethnic terms were quite rare in these sets of articles, relative to other sorts of identity terms. There was some discussion of the so-called 'mixed race' category in articles mentioning Ennis, while race—at least as identified by the term 'black'—did not arise in any significant measure in describing Farah. National identity terms appeared frequently in articles mentioning either or both athletes: 'British' was used in numerous ways, while 'English' often referred to the English language rather than

English national identity, in relation to Farah's arrival in Britain with no knowledge of the English language. Even in the absence of positive net migration, the population is projected to grow significantly in the future. Assuming net migration of zero at every age, the UK population is projected to reach 66 million by 2035 an increase of 6% from the 2010 level.

Research questions

The Migration Observatory was commissioned to examine the use of words relating to multiple kinds of identity in relation to Ennis and Farah. Specifically, we investigated terms associated with national identity, local origins within or outside Britain, racial/ethnic background, and religion to answer the following questions:

1. How often do origin- and identity-related words appear in coverage mentioning either Mo Farah, Jessica Ennis, or both?
2. Does origin-related or race-related language feature more often in coverage mentioning Jessica Ennis compared to Mo Farah?
3. Within coverage mentioning each athlete, which types of terms are more relevant?

Methodology

Building the corpus

The methodological approach was based on the field of corpus linguistics. Corpus linguists assemble large bodies of text, and then use a combination of quantitative and qualitative techniques to analyse those texts. The Migration Observatory project is in the process of building a larger corpus of British newspaper stories mentioning migration or asylum/refugees, which will date back to 2005 and run through the present. However, a smaller bespoke corpus was required to investigate British Future's questions, since most coverage of Ennis and Farah would not appear in a corpus focused on migration. We named this corpus 'FEO' to reflect the topical search for Farah, Ennis, and the Olympics.

FEO was built by searching the NexisUK database for newspaper articles mentioning Ennis, Farah, or both, between the dates of 29 June and 9 September 2012. This period corresponded with one month of coverage before the Olympic Opening Ceremony, the period of the games themselves, and one month after the Olympic Closing Ceremony. To rule out articles mentioning different people coincidentally named Ennis or Farah, we discarded articles that did not also include the term 'Olympics' or a variant such as 'Olympic', 'Olympian', or 'Olympism'. In database terminology we used the following search strings:

- (a) Farah AND Olymp! AND NOT Ennis
- (b) Ennis AND Olymp! AND NOT Farah
- (c) Farah AND Ennis AND Olymp!

The ! character (as in 'Olymp!') indicates that all words beginning with the letters 'Olymp' were searched, including 'Olympic' and 'Olympian'.

The search was conducted on 16 major national weekly and Sunday newspapers: The Sun, the Daily Mirror and Sunday Mirror, The Express, Sunday Express, Daily Mail and Mail on Sunday, The Times, Sunday Times, The Guardian, The Observer, The Independent, Independent on Sunday, Daily Telegraph, Sunday Telegraph, and the Financial Times. Highly similar articles as defined by NexisUK were filtered out of the results, which yielded 2,429 articles totalling 1,972,971 words. Text was stored in nine separate folders ('subcorpora') divided by period (pre-Olympics, during Olympics, post-Olympics) and by athlete mentioned (Ennis only, Farah only, both Ennis and Farah). Table 1 describes how the nine subcorpora emerged from the search strategy. In the analysis that follows, references to the 'Farah

subcorpora', 'Ennis subcorpora', and 'both subcorpora' point to coverage before, during, and after the Olympics that mentioned the respective athletes.

Table 1. The nine subcorpora in the FEO corpus

	Before Olympics	During Olympics	After Olympics
Farah AND Olymp! AND NOT Ennis	F _a	F _b	F _c
Ennis AND Olymp! AND NOT Farah	E _a	E _b	E _c
Farah AND Ennis AND Olymp!	FE _a	FE _b	FE _c

Table 2 provides a descriptive overview of the number of articles that appeared in each subcorpus. It also divides FEO into tabloid and broadsheet coverage. The Farah-only corpus was the largest (867 articles) but only by a small margin over the Ennis-only group (838), while a substantial share (724) included both names. It is important to recognise that these articles were not all 'about' Farah or Ennis; many had other main topics and mentioned one of the two athletes in a peripheral way. Nonetheless, they are all included in the corpus, as even a peripheral mention can contribute to the larger narrative about a person or event.

Table 2. Dimensions and breakdown of the FEO corpus

TOTAL COVERAGE	Before Olympics	During Olympics	After Olympics	Total Coverage
Farah AND Olymp! AND NOT Ennis	95	343	429	867
Ennis AND Olymp! AND NOT Farah	152	520	166	838
Farah AND Ennis AND Olymp!	84	382	258	724
Total coverage	331	1245	853	2429
A. TABLOIDS				
Farah AND Olymp! AND NOT Ennis	40	154	210	404
Ennis AND Olymp! AND NOT Farah	81	233	93	407
Farah AND Ennis AND Olymp!	41	177	111	329
Total coverage	162	564	414	1140
B. BROADSHEETS				
Farah AND Olymp! AND NOT Ennis	55	189	219	463
Ennis AND Olymp! AND NOT Farah	71	287	73	431
Farah AND Ennis AND Olymp!	43	205	147	395
Total coverage	169	681	439	1289

Analytical procedures

With FEO built, an analytical approach was required. Corpus linguists might investigate specific questions—a 'top down' or 'corpus-based' approach—or examine the most frequent words and combinations of words appearing in a corpus to see what emerges—a 'bottom up' or 'corpus-driven' approach (Baker and McEnergy, 2005). In this case, to investigate specific questions about Ennis and Farah and terms associated with identities, a top down approach was most appropriate.

The goal of the project was to examine language associated with the identities of the two athletes. Particularly, to investigate the questions raised by British Future, the aims were to see if Ennis was more described in terms of national and local origins rather than her racial/ethnic background, while Farah’s coverage emphasised not local roots but foreign birth and racial and religious identities. To this end, groups of words related to origin and identity were selected for analysis. For the purposes of this report, when a word in the text appears in SMALL CAPS, it is being referenced as a search term.

The following terms were used: BRITISH and ENGLISH for national identities; SHEFFIELD for Ennis’ place of origin, HOUNSLOW/ ISLEWORTH/ TEDDINGTON for local UK origins and SOMALIA for birthplace for Farah; BLACK, MIXED-RACE, CARIBBEAN (including AFRO-CARIBBEAN), and JAMAICA for race and ethnic background for both (Ennis’ father is Afro-Caribbean and from Jamaica); and MUSLIM for Farah’s religious identity. (Ennis’ religious faith and background, if any, does not appear to be a matter of public information). These words are listed, by category, in Table 3 below. Again, the ! character (as in SOMALI!) indicates that words beginning with ‘Somali’ were searched, including SOMALIA and SOMALIAN.

Table 3. Origin and identity terms used to analyse the FEO corpus

General terms	Origin terms - Ennis	Origin Terms - Farah	Race & religion terms
British	Sheffled!	Somali!	Black
Britishness	Caribbean	Hounslow	Mixed
English	Jamaica!	Teddington	Muslim
		Isleworth	

Analysis and results

Analysis began by simply counting the frequency of each selected word in the corpus of texts mentioning Ennis only, Farah only, and both Ennis and Farah, respectively. Computer software, in this case WordSmith Tools (Smith, 2012) can conduct these counts automatically in a matter of moments.

However, a computerised tabulation does not account for different meanings of words in different contexts. Realising that some uses of the searched words might actually be irrelevant to our purposes, we also examined a ‘concordance’ of each word to see how it was being used. A concordance displays each use of a word of interest in the context of the words surrounding it. Examining the concordances, we detected a number of irrelevant results that were subtracted from the initial counts. For example, some uses of HOUNSLOW (a term referring to Farah’s local origins) were actually referring to Richard Hounslow (a British Olympic canoeist). These mentions were excluded. Other irrelevant and subsequently excluded uses included BLACK as a non-racial descriptive adjective, MIXED in any context other than ‘mixed-race’ or ‘ethnically mixed’, and ENGLISH appearing as part of the meta-data in the articles downloaded from NexisUK when it referred to the language of publication. In addition, concordances enabled us to verify that certain instances should be included: for instance, although mentions of the football club ‘Sheffield United’ might seem to be irrelevant, it turned out that all but one of these instances referred to the club’s decision to name a stand after Jessica Ennis.

After completing these steps of designing and building the corpus, counting instances of certain words, and verifying their relevance through manual concordance analysis, we can make the following observations which are displayed below. Table 4 displays both raw counts of each word and uses of each word as a percentage of the total words in each sub-corpus. This is done to allow comparisons across subcorpora as needed, although many of the results of interest in this commissioned project involved comparing words within each corpus to ask, for instance, if Ennis is more often associated with her Sheffield background or her racial/ethnic background.

Table 4. Frequency of origin- and identity-related words in newspaper articles mentioning Mo Farah and/or Jessica Ennis

National identity	Farah only	% of words ¹	Ennis only	% of words	Farah and Ennis	% of words
British ²	1110	0.172%	1095	0.178%	1770	0.247%
Britishness	23	0.004%	6	0.001%	17	0.002%
English ³	129	0.020%	58	0.010%	93	0.013%
Origin terms - Farah						
Somali!	305	0.047%	3		136	0.019%
Hounslow ⁴	14	0.002%	2		12	0.001%
Teddington	11	0.002%	0		10	0.001%
Isleworth	2		0		2	
Origin terms - Ennis						
Sheffield! ⁵	16	0.002%	281	0.046%	128	0.018%
Caribbean	11	0.002%	2		16	0.002%
Jamaica! ⁶	1		14	0.002%	10	0.001%
Race & religion terms						
Black	5	0.001%	0		3	
Mixed ⁷	1		6	0.001%	29	0.004%
Muslim	41	0.006%	0		18	0.003%
Total Words	643,618		613,741		715,612	

[1] Percentages below 0.001% are left blank

[2] Includes all instances of BRITISH, not just those which refer to FARAH or ENNIS

[3] Excludes instances where ENGLISH refers to the language of publication

[4] Excludes references to RICHARD HOUNSLOW, a canoeist for Great Britain

[5] Includes instances of SHEFFIELD UNITED where referring to ENNIS

[6] Excludes references to USAID BOLT or other JAMAICAN athletes

[7] Includes only uses of MIXED as 'MIXED RACE' or 'ETHNICALLY MIXED'

SOMALIA and its variants were, unsurprisingly, present in the Farah-only subcorpora and the Ennis-and-Farah subcorpora but almost not at all in the Ennis-only subcorpora. SOMALIAN origins decisively outweighed local origin terms in the coverage of Farah, as measured by word frequencies. SOMALI and its variants were much more frequent in the Farah subcorpora than local origin terms for Farah; HOUNSLOW, TEDDINGTON, and ISLEWORTH are all fairly low-frequency words. Second, it is approximately as frequent in Farah's subcorpora as SHEFFIELD is in Ennis' subcorpora. Finally, it is more frequent in the Farah-only subcorpora than in the corpus of coverage including both Farah and Ennis.

These differences—as well as others discussed below—are all statistically significant according to log-likelihood tests which, roughly, assess the odds of a difference in word usage (either a word appearing more often in one corpus than another, or a word appearing more often than another word in the same corpus) occurring by random chance. In such large sets of words, even small differences are likely to be statistically significant at least at the 1% level, meaning roughly that we can be 99% confident that the differences we observe are not due to random chance.

Origins of Jessica Ennis

Ennis' coverage showed a predominance of local origin terms over racial or ethnic terms, or any reference to the Jamaican or Afro-Caribbean side of her family origins. SHEFFIELD is most frequent in the Ennis subcorpora, and somewhat frequent in the subcorpora containing references to both athletes, while appearing rarely in the Farah subcorpora. As noted above, coincidentally, SHEFFIELD is about as common in the Ennis-only coverage as SOMALIA is in the Farah-only coverage. On the other hand, CARIBBEAN in reference to her father's origins is not more likely to appear in the Ennis subcorpora. JAMAICA is more likely to appear in the Ennis subcorpora, but is much less frequent than SHEFFIELD, evidence that local roots rather than more distant ones were more prominent in coverage surrounding Ennis.

Race and religion

The results suggest that explicitly racial/ethnic discourse and mentions of religion were relatively rare, but associated more with one athlete than the other. First, in terms of religion, the word MUSLIM did appear at a noticeably higher rate in the Farah subcorpora than in the 'both' subcorpora, and was essentially absent in the Ennis subcorpora. However, SOMALIA and its variants were significantly more frequent; religion did not appear to be as prominent as national origins in coverage mentioning Farah by this measure. No religious term appeared at a significant rate in articles mentioning Ennis only.

In racial terms, BLACK as an identifier was virtually absent from all coverage of the athletes. MIXED-RACE did appear in the subcorpora mentioning both athletes, although not very frequently. This suggests that explicitly racialised discussion of either athlete was rare—and if it occurred it was more likely to refer to the notion of a mixed background than to refer to either athlete as black.

National identity

British Future were also concerned with detecting discussion of Ennis and Farah as representatives of new forms of British identity. Here, the analysis was less conclusive. BRITISH was used many times in the coverage involving both athletes, although concordance analysis revealed that it referred to many other objects, events, and people aside from the athletes' own identities. Meanwhile, the explicit identity term BRITISHNESS appears most often in the Farah subcorpora, followed by the 'both' subcorpora, but was mentioned relatively few times in the entire set of texts examine, with less than 50 mentions in total. ENGLISH appears most often in the Farah subcorpora. However, this did not mean that Farah was commonly identified as English in the sense of national identity. Concordance analysis revealed that the majority of these uses refer to the English language, as it was common to discuss Farah's arrival in the UK with no knowledge of how to speak English.

Change over time

Additional analysis displayed in Table 5 considers change in coverage over time, comparing articles in the time period leading up to the Olympics with coverage during and then after the Games. As noted above, BRITISHNESS was mentioned relatively few times, but it did show a marked increase over time: this term was hardly used at all in pre-Games articles mentioning Ennis or Farah, but rose in the post-Olympic coverage mentioning Farah. To the extent that BRITISHNESS indicates explicit discussions of national identity in connection to these two athletes, it seemed to increase during and immediately after the Games. This might suggest that the athletes generated some level of reflection on British identity itself. Further analysis could be directed at testing this hypothesis beyond simple counts of appearances of BRITISHNESS. Each athlete's most-used origin term of those examined here—SHEFFIELD for Ennis and SOMALIA for Farah—was used more during and after the Olympics than before. This may indicate that coverage before the Games was somewhat less focused on Ennis' and Farah' background stories, although both terms were present in the pre-Olympics corpora as well.

Table 5. Frequency of origin- and identity-related words before, during, and after the Olympics

	Before Olympics	% of total words	During Olympics	% of total words	After Olympics	% of total words
Farah-only coverage						
English	21	0.023%	48	0.018%	60	0.021%
Britishness	1	0.001%	8	0.003%	14	0.005%
Somali!	32	0.035%	148	0.057%	125	0.043%
Total words	90,341		260,842		292,435	
Ennis-only coverage						
English	18	0.017%	30	0.008%	10	0.009%
Britishness	1	0.001%	3	0.001%	2	0.002%
Sheffield	23	0.022%	192	0.049%	66	0.057%
Total words	106,870		390,577		116,294	
Farah & Ennis coverage						
English	7	0.007%	53	0.015%	33	0.012%
Britishness	0		12	0.003%	5	0.002%
Somali!	6	0.006%	103	0.030%	27	0.010%
Sheffield	17	0.017%	92	0.026%	19	0.007%
Total words	98,607		348,487		268,518	

Limitations of the study

This research and its findings should be interpreted with the following limitations in mind. First, as a corpus-based or top-down approach, the study design began with explicit decisions about which questions would be investigated and, further, which terms to analyse. In this case, these decisions were fundamentally shaped by the questions that British Future wished to ask and the terms in which they were interested, although the Observatory designed and executed the methodology independently. Second, the examination of word frequencies is only one technique among a number of possible ways of investigating language use that could include close readings of text, manual content analysis, or the use of automated computer algorithms designed to detect the appearance of frequent patterns, phrases, or topics (Quinn et al, 2010). Therefore, other modes of investigation into this corpus are certainly possible, and might turn up different ways of identifying identity language in the context of Ennis, Farah, and the Olympics. Third, the choice of articles hinged on the NexisUK database, and specifically its ability to filter out duplicate articles, which is very helpful but may be imperfect. Finally, the size of FEO—in terms of both words and number of articles—is relatively small compared to other datasets used by corpus linguists. Since one of the advantages of using corpus techniques is their ability to uncover patterns which may infrequently occur, a small corpus may not capture these less common observations.

Conclusions

This study has shown that press coverage mentioning Jessica Ennis was more likely to refer to her local place of origin in Sheffield rather than her racial or ethnic background, including her father's birthplace of Jamaica. Meanwhile, press coverage mentioning Mo Farah was more likely to refer to his birthplace of Somalia rather than an upbringing in Hounslow. Farah's religious group identity, represented by mentions of the word MUSLIM, garnered more mentions than his local British origins, but fewer than mentions of SOMALIA. Trends over time suggested an increasing use of some forms of identity language from the pre-Games coverage to the during- and post-Games coverage.

This work is a key early step in the development of the Media Monitoring Project at the Observatory. Corpus linguistic methods enable researchers to efficiently handle large amounts of textual data while also providing scope

for qualitative reliability and validity checks. Yet, as demonstrated in this report, it is important to be clear about the steps and decisions surrounding corpus-building and analysis. By applying this branch of linguistics to social scientific questions—particularly in the context of media outputs—we aim to shed new light on the nature of portrayals surrounding migrants and refugees. Although additional research will investigate core concerns of the Observatory’s researchers, such as describing how migrants have been portrayed and with key actors in stories about migration, the corpus and linguistic tools will be available for additional work on behalf of a wide range of organisations with interests in examining print media coverage of migration and related issues.

References

- Baker, P. and T. McEnery. “A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts.” *Journal of Language and Politics* 4 no. 2 (2005): 197–226.
- Quinn, K.M., B.L. Monroe, M. Colaresi, M.H. Crespin, and D.R. Radev. “How to analyze political attention with minimal assumptions and costs.” *American Journal of Political Science* 54 (2010): 209–228.
- Smith, M. *WordSmith Tools*. Liverpool: Lexical Analysis Software, 2012.



The Migration Observatory

Based at the Centre on Migration, Policy and Society (COMPAS) at the University of Oxford, the Migration Observatory provides independent, authoritative, evidence-based analysis of data on migration and migrants in the UK, to inform media, public and policy debates, and to generate high quality research on international migration and public policy issues. The Observatory's analysis involves experts from a wide range of disciplines and departments at the University of Oxford.



COMPAS

The Migration Observatory is based at the ESRC Centre on Migration, Policy and Society (COMPAS) at the University of Oxford. The mission of COMPAS is to conduct high quality research in order to develop theory and knowledge, inform policy-making and public debate, and engage users of research within the field of migration.

www.compas.ox.ac.uk

Press contact

Rob McNeil

Senior Media Analyst

robert.mcneil@compas.ox.ac.uk

+ 44 (0)1865 274568

+ 44 (0)7500 970081

Suggested citation

Allen, William and Scott Blinder. "Jessica Ennis, Mo Farah and Identity Language in the British Press: A Case Study in Monitoring and Analysing Print Media." Migration Observatory report, COMPAS, University of Oxford, December 2012.

