BRITISH COUNCIL

UNIVERSITY *of* **STIRLING**

Newspaper discourses on Andy Murray's performance at Wimbledon:

A contrastive corpus-driven investigation of successful and unsuccessful tennis match reports

Author's Name: Ben Dickson

University of Stirling

Newspaper discourses on Andy Murray's performance at Wimbledon: a contrastive corpus-driven investigation of successful and unsuccessful tennis match reports

Abstract

Over the last ten years of UK newspaper sports reporting, much of the media's attention has been focused on the efforts of the tennis player, Andy Murray. While the Dunblane-born athlete is termed a Scot by some and a Brit by others, the language with which the media refers to Murray's national identity has become a source of some debate. Arising from this has been the belief that Murray is referred to as British when he wins, and Scottish when he loses by the UK press (Clarke, 2013; Mesure, 2012; Viner, 2010).

Drawing on Corpus Linguistics, this study aimed to investigate the language used by the UK media when reporting on Andy Murray tennis matches in order to determine whether people's intuitions on this topic have any basis in actual language use.

A corpus of Wimbledon tennis match reports charting Murray's entire career was first of all created. Articles were taken from the online editions of four newspapers which represented a cross-section of the UK press: *The Guardian,* a broadsheet published nationally across the UK; *The Mirror,* a tabloid, also published nationally; *The Scotsman,* a broadsheet published locally in Scotland; and *The Daily Record,* a tabloid also published locally in Scotland. Corpus analysis software was used to compare the language of the unsuccessful versus successful match reports of each of these newspapers in order to identify the keywords used within the articles. These keywords were then analysed and their significance discussed.

While the results produced no evidence to support the claim that the media identifies Murray differently depending on the result of his match, the findings did show that the newspapers differ in terms of how they portray these matches. Much of these findings were well beyond the expectations of this study, which is one of the strengths of corpus research.

It was concluded that people's intuitions about language are often inaccurate and that this investigation had helped reconsider one such belief. Further research could examine how the language of other types of news media within the UK, such as the BBC, would compare to these findings.

List of figures and tables

Table of contents

Chapter 1 - Introduction

In Section 1.1 of this introduction I shall detail the background and context of the dissertation. The next section shall briefly summarise a selection of the background research on the areas covered in this study (Section 1.2). Next, Section 1.3 makes a statement of problem, while Section 1.4 looks at the purpose for this study. The introduction is concluded by providing a brief outline of each chapter in this paper.

1.1 - Background and context

When Andy Murray won the Wimbledon men's singles title in 2013, the tennis player from Dunblane had finally reached the pinnacle of his sport. His achievement had not been matched by any UK man since the pre-war days of Fred Perry. He was also the first person from Scotland to win Wimbledon since Harold Mahoney in 1896.

Murray's victory capped a sporting journey that began back in 2005, with his very first Wimbledon appearance. Ever since that debut, Murray's progress in tournaments from around the world has been covered by the various section of the UK press. Murray's relationship with these newspapers has been somewhat eventful at times. For example, he experienced the hostility of the national tabloid press when he quipped that he would be supporting '*anyone but England*' at the football World Cup before losing in a fourth round match at Wimbledon in 2006 (Mail Online, 2006). However, by 2012 he was proclaimed a '*hero*' by the same newspaper after winning the gold medal in tennis at the London Olympic Games (Harris, 2012).

Between 2005 and 2014, the language used by the UK newspapers when reporting Andy Murray's various victories and defeats, soon became the source of debate in the UK, particularly concerning Murray's national identity. A perception developed among some readers that the press tended to describe Murray as British when he won and as Scottish when he lost (Clarke, 2013; Mesure, 2012; Viner, 2010). Intuitions about language use are common, but can often be incorrect. Through my studies on Corpus Linguistics, I realised there was a statistical methodology with which I could analyse the language of these unsuccessful and successful match reports by the media. By doing so I could determine whether or not this type of intuition held any basis in the language actually used by the UK newspapers.

1.2 - Background research

The background research suggested that the subject of national identity within a sports article can often be influenced by two external factors relating to the newspaper; the location of publication, i.e. whether a local or a national paper; and the type of the newspaper, i.e. whether a broadsheet or a tabloid.

Firstly, the literature showed that the area where a newspaper was published often had an impact on the narrative of articles. Local publications, sold only in Scotland, displayed many Scottish frames of reference and stories tended to be explicitly Scottish (Kielty et al., 2006; Rosie et al., 2004), while those sold nationally throughout the UK, showed elements of British triumphalism, particularly when reporting sport (Thomas and Anthony, 2015).

The literature also showed that the broadsheet and tabloid media approached the subject of national identity differently. The tabloids tended to show explicit nationalism regardless of their political alignment, while the broadsheets were more subtle in their approach to national identity (Conboy, 2007; Uribe and Gunter, 2004).

1.3 - Statement of problem

People often have their own intuitions about language. However, as these ideas are not based on any empirical evidence, they can often be incorrect and misleading. The claim that the media's portrayal of an individual's national identity is determined by the result of a tennis match, if true, would be significant in analysing the language of the UK press. However without any basis, the intuition bares no weight and so underlines the need for research in this area.

Furthermore, within a devolved United Kingdom, where nations exist within other nations, the concept of a nationwide press acting in unison in this way is unlikely. The UK newspaper are comprised of local, national, tabloid and broadsheet publications, all aiming to serve the various different niches that exist within UK society. Thus, to analyse the language of the UK media, newspapers from each corner of the market must be investigated individually.

1.4 - Purpose of the study

The purpose of this study is to contrast the language used by the UK media when reporting on the unsuccessful and successful tennis matches of Andy Murray. Using corpus-driven

investigation, it aims to considering the follow two aspects:

(i) the relationships between how Murray's national identity is referred to in match reports when he loses compared to when he wins; and

(ii) the language of the local, national, broadsheet and tabloid media in relation to whether Murray loses or wins.

1.5 - Dissertation structure

Following the introduction, this dissertation shall include four additional chapters. In Chapter 2, a literature review shall take place which discusses previous research on the topics of national identity in the UK, the local and national media in the UK and finally the broadsheet and tabloid media in the UK. In Chapter 3 the methodology for this dissertation is explained. It details the characteristics of corpora in general, how the corpus for this study was compiled and also the procedures used for data analysis. In Chapter 4 the results of keyword analysis are displayed and discussed, with a separate section for each newspaper that was investigated. Finally, in Chapter 5, a summary is made of the findings from this investigation, before suggesting the limitations of this study and any further research that could be made.

Chapter 2 - Literature review

Chapter 2 shall examine the literature that relates to the main areas of research by this dissertation. The first section shall discuss the literature concerning the issue of national identity in the UK. It shall also look into the portrayal of national identity within sports journalism as well as the presence of nationalism in sports articles. In Section 2, the literature examined looks more closely at the differences and similarities between the local media in Scotland and the national media of the UK. These two groups are discussed in relation to how they tackle the issue of national identity. Finally, in Section 3, literature on how identity is represented within the discourse of the broadsheet and tabloid newspapers shall be examined. The extent to which elements of nationalism and personalisation exist within these types of papers shall also be discussed.

2.1 - National identity in the UK

The United Kingdom of Great Britain and Northern Ireland is both one country and a complex political arrangement of various governments, assemblies, nations and peoples. Over the last 500 years, wars, unions, separations, acts and bills have shaped the complicated constitutional set up of this country. With the 2014 independence referendum for Scotland fresh in the memory, the UK continues to be a fluid and ever changing amalgamation of socio-political identity.

When sport is added to this mix, the issue of national identity in the UK can become rather divisive. MacRury and Poynter (2010) explored one problematic example which occurred during the planning of the London Olympic Games in 2012. The main issue of contention came when Prime Minister Gordon Brown called for a 'Team GB' soccer team to take part in the games. In the past, the UK Olympic committee had chosen not to enter a soccer team for the Olympic Games as, like rugby, soccer is a devolved sport and competed internationally by the separate football associations of England, Scotland, Wales and Northern Ireland. British identity faced a 'paradox' with the government in London asking football fans to 'come together' while the government in Edinburgh called the move an 'own goal' and a 'threat to the status of the independent national teams' (MacRury and Poynter, 2010, p. 2960).

It is within similarly contentious settings that we find the current world number 3 in tennis, Dunblane's Andy Murray; considered both Scottish and British, representing both Scotland and Great Britain at tennis, and being reported on by the Scottish and British media, both of whom have readerships from all parts of Scotland and Great Britain (Meech and Kilborn, 1992).

2.1.1 - The portrayal of national identity within sports journalism

The public image of politicians, musicians and celebrities has long been at the hands of the media. It has often been they who have determined the popularity of such individuals, rather than the success that that individual has achieved in their own field (Van Dijk, 1998). Whilst public image aspects like personality, intelligence and talent are subjective and open to manipulation by the press, the national identity of an individual is not usually open for discussion and debate (Van Dijk, 1998, p. 30). Except however, in the case of sports and its stars, where nationalism and the idea of a nation versus another nation remains at the essence of every global sporting event (Bairner, 2015).

In the of case Andy Murray, being the only man from the UK to win any major tennis tournament in the last 77 years, the local and national media within the UK have been waiting a long time to celebrate this success with their readers. The problem is, Murray is not just a British success, but a Scottish success too, which means that the media have a conundrum of how to portray him to their readership. For example, national papers wishing to share his success may feel that, by calling him Scottish, they risk alienating readers from other parts of the UK, while to ignore his Scottish identity would lack detail and credibility. Thus, the debate surrounding national identity and sports journalism has been discussed by the following researchers.

In 2010, Bloyce et al. completed a sociological analysis of the British media's coverage of the 2005 British and Irish Lions Rugby tour of New Zealand. This study is of particular relevance as international rugby union is usually played by the individual nations of England, Scotland, Wales and an all-Ireland team. However, once every four years, these four teams combine to form one team; the British and Irish Lions, meaning that questions of national identity are taken on by the press. Bloyce et al. (2010) found that the sports journalists from Scotland, Wales and Ireland became "bound up in the dual-sided production of either national difference or sameness" towards the British team. Welsh and Scottish journalists were constrained to 'Brit-bashing' and anti-English rhetoric, which Bloyce et al. (2010) attributes to feelings of London domination within Britain (p. 464). On the other hand national 'sameness' was displayed by Irish journalists, who positively associated the Irish players with the English. The authors believe this self-confident stance may be down to Ireland's political maturity, having gained independence from the UK in 1919 (Bloyce et al., 2010, p. 464). Overall though, each local media found it difficult to avoid having a strong opinion of national identity resonate through their discourse.

National identity was seen to be similarly divisive during the media's coverage of the London Olympic Games in 2012. When analysing the newspapers reactions to the games' opening ceremony, Thomas and Anthony (2015) stated that the journalists rendered kaleidoscopic visions of national identity in the UK (p. 501). For example, one author from the national broadsheet *The Daily Telegraph* said the ceremony represented UK identity by highlighting British historical achievements, while an author from the national tabloid *The People* suggested it was represented by the self-depreciative humour and eccentricity (Thomas and Anthony, 2015). Thomas and Anthony (2015) concluded that there was no single identity displayed by the media and that "any linear narrative of British identity was both impossible and misrepresentative," (p. 500).

Overall these examples have shown that when portrayed through sports journalism, national identity within the UK is often multi-faceted. Further research into this area using corpus analysis would allow for a more accurate study of the media's language.

2.1.2 - Nationalism in sports articles

While Bale (1986) argues that regionalism and nationalism are built into sports teams, the more current literature on national identity argues strongly that the sports journalists are often responsible for stoking the flames of nationalism within the UK (Bairner, 2015; Rosie et al., 2004). This is believed to have been pursued in two ways by the media.

First of all, Rosie et al. (2004) suggest that the authors of sports articles tend to position their readership on national terms, meaning they encourage and remind their readers to think patriotically about sports. Nationalism is used to by these journalists to press a point of identification with their readership and Bairner (2015) argues that the presence of various sporting nations within one country makes the UK the perfect place for this implementation of nationalism. According to Billig (1995), the media often claim to stand for the country and represent the nation, and so sports coverage is usually the means with which this is done.

Secondly sports journalism has begun mythologizing nationalism using sporting contexts (Bloyce et al., 2010). Bale (1986) believes that for many people in the UK, after war, sports are the principle means of collective national identity in modern life, and so the media look to exploit this popular sentiment. Bloyce et al. (2010) states that nationalism in the UK seemed to be growing and that the media had begun to use sport to symbolize it.

Overall, sports, media and nationalism seem to be interlinked. This dissertation will use the information gathered from this literature to better understand its own findings. The further research of this dissertation into this area, will likely help confirm how nationalism and national identity affect sports journalism in the UK.

2.2 - Local and national media in the UK

In the UK, defining what exactly is meant by 'local' and 'national' media can be quite complicated (MacInnes et al., 2007). Throughout the UK, there are towns, cities, counties and regions all producing what they would consider to be their 'local' newspaper. Meanwhile, Scotland, Wales and Northern Ireland all produce newspapers that are sold nationally, but are still considered 'local' within the context of the entire United Kingdom.

To complicate things further, there are UK-wide newspapers that have both a UK and a Scottish edition of their paper, further blurring the lines between 'local' and 'national' media (Meech and Kilborn, 1992). MacInnes et al. (2007) studied the circulation densities of newspapers throughout the UK and concluded that the term, *the British national press*, was highly problematic as it under-theorized the national identity and diversity of its readership. The authors advised caution when using newspapers to represent a place on the map.

In this dissertation, the term *local* media refers to newspapers that have been produced and circulate only in Scotland, but does not include the Scottish editions of UK newspapers. *National* media are those produced elsewhere in Britain, predominantly in England, but circulated throughout the entire United Kingdom, including Scotland. Again, the Scottish editions of these newspapers are not included as part of this group.

2.2.1 - Local media

Since this dissertation examines the references to Andy Murray's national identity by the media, it is of interest to note whether this may be shaped by how newspapers reference their own national identity. The literature suggests that local newspapers lack consistency in this area. For example, in Higgins' (2004) corpus analysis of the discourse of Scottish newspapers, the author noted that the sub-editor of the *Glasgow Herald* had asked his writers not to refer to Scotland in stories where the national context was obvious. However, corpus findings showed that nearly 55% of sentences containing 'place' formulations, referred to Scotland (Higgins, 2004, pp. 479-480). Additionally, when Petersoo (2007) examined the discursive construction of national identities in the media, he discovered in one section of *The Scotsman* newspaper, the word *we*, had no less than 3 different frames of reference. Naming it the 'wandering *we*,' (Petersoo, 2007, p. 431) the author found that Scottish media in general used the word interchangeably, allowing it to mean British, Scottish or the newspaper, depending on the context (Petersoo, 2007).

One area of consistency concerning the local media is the content of these newspapers, which MacInnes et al. (2007) describe as 'ostentatiously Scottish' (p. 195). Rosie et al. (2004) compare the '*implicitly* English' nature of the UK national papers to the '*explicitly* Scottish' nature of the local media, and suggest that, since devolution in 1999, the Scottish papers carry far fewer stories about the rest of the UK (pp. 454-455). Kiely et al. (2006) refer to this as 'home-nationalisation' (p. 489) of the UK media and mentions that English

people living in Scotland find themselves surrounded by Scottish national frames of reference with very little British perspective. Meech and Kilborn (1992) state that the marketing of all Scottish papers, whether tabloid or broadsheet, is heavily laden with Scottish national imagery and rhetoric, ranging from the thistle on *The Scotsman* logo, to the 'Edited, printed and published in Scotland' slogan found on every front page of *The Daily Record* (p. 257). Overall, further research is required to discover how the language used by the local media may be shaped by the results of sporting events and whether the issue of national identity is affected by this.

2.2.2 - National media

When London successfully hosted the Olympic Games in 2012, Thomas and Anthony (2015) noted certain linguistic features present within the British national media's reporting of the opening ceremony. First of all, they highlighted the triumphalism and self-congratulatory nature of the national press' coverage, in particular the regularity of word plays involving 'great' and Britain. Secondly, while the ceremony highlighted both the achievements and failings of Britain's imperial past, the national media concentrated their attention on the former. Thomas and Anthony (2015) also argued that the British media supported a strong British national identity through criticizing the ceremonies of previous host countries. This research perhaps indicates how the language of the national press can be influenced by the success of a sporting event and so these findings are relevant to the research of this dissertation. Nevertheless, more investigation is required to find out if the findings of Thomas and Anthony (2015) are also witnessed within the local media in Scotland. A Keyword analysis of the unsuccessful and successful Andy Murray matches would shed some light on this area.

2.3 - Broadsheet and tabloid media in the UK

An important area of study relevant to this investigation is the characteristics of broadsheet and tabloid media in the UK. Since the present dissertation shall examine these two forms of media concerning the way they report on Andy Murray's tennis matches, an understanding of their differences and similarities must first of all be made. This shall be done by reviewing the relevant literature on this topic.

There have been arguments made over recent years that the differences are narrowing between the broadsheet and tabloid newspapers. However, while Franklin (as cited in Uribe and Gunter, 2004), suggests that the content of serious journalism and tabloid

journalism is becoming more similar, the majority of literature states that broadsheets and tabloids have become homogenised in form and style only, with their range of content differing greatly. For example, having converted in recent years to a compact newspaper format, some of Britain's oldest broadsheet newspapers such as *The Times* and *The Scotsman* are now no longer printed 'broadsheet' in the most literal sense of the word. Nevertheless, the content of broadsheets and tabloids remains strikingly different. Uribe and Gunter's (2004) research explored over 10 years of news coverage by *The Sun* and *The Mirror* newspapers and found that they were dominated by entertainment and less serious domestic stories, relying heavily in the presence of headlines and visuals. They use the term 'tabloidization' to describe a newspaper which follows in this style, and use *The Sun* and *The Mirror* as examples (p. 388). Other evidence of tabloidization includes discourse simplification whereby less space on the page is given to information and text, which are replaced instead with pictures and headlines (Connell, as cited in Uribe and Gunter, 2004; Djupsund and Carlson, as cited in Uribe and Gunter, 2004; Rooney, as cited in Uribe and Gunter, 2004). Less attention is devoted to foreign news stories, and there is often feminization of certain stories according to Van Zoonen (as cited in Uribe and Gunter, 2004) and Winston (as cited in Uribe and Gunter, 2004).

With particular significance to the areas of investigation in this dissertation, the literature suggests there are three two ways in which tabloids and broadsheet differ in content and language: nationalisation and personalisation.

2.3.1 - Nationalisation

In 2007, Conboy compared the language of UK tabloid newspapers and found strong evidence to suggest that these tabloids often distorted stories so as to promote modern day nationalism and patriotism. His study looked at the tabloid reporting of the D-Day and Battle of Trafalgar Anniversaries, which took place in 2004 and 2005 respectively. The language used in tabloids was heavily laden with national sentiment and each paper exploited the occasions for their own editorial agendas. For example, *The Daily Mail* referred to the British army veterans attending the ceremony as 'British heroes,' but then went on to state the failures in government bureaucracy reporting that some 'heroes' had not received invitations (Conboy, 2007, p. 259). Overall, despite the varying political allegiances of the different newspapers, each tabloid adopted a 'consensual approach to patriotic populism' (Conboy, 2007, p. 257). He argues they did this because, despite the social-class and political differences of the readership, to be patriotic and nationalist is seen

as a civic duty by many in the UK and so will always be favourable with readers. Conboy (2007) concludes that, while this central patriotic approach is shared by most media outlets in the UK, including very subtlety in the broadsheets, it is far more evident in the 'explicitly nationalist' tabloid press (p. 257).

Overall, this evidence is important in understanding the difference between the broadsheet tabloid press on the subject of national identity and the impact any engrained nationalism may have when reporting on Andy Murray.

2.3.2 - Personalisation

With regards to why tabloid newspapers may choose to pay particular attention to the national identity of sports professionals, literature suggests it may be because they are more personally rather than professionally focussed (Uribe and Gunter, 2004). Studies by Franklin (as cited in Uribe and Gunter, 2004) and Uribe and Gunter (2004) found that *The Mirror* newspaper's coverage of people's private lives increased, with personal news items acquiring more space within the paper than professional-focused items. Sparks (2000) found that tabloids placed more emphasis on the private lives of those within their news stories, while Van Zoonen (2000, as cited in Uribe and Gunter, 2004) claims that tabloid style is defined by personalization. However, there was little research to suggest that this was the case in broadsheet newspapers.

Using an analysis of keyword lists, this dissertation will be able to substantiate some of these claims made in the literature. It may also discover the extent to which the personalization of sports personalities occurs within the tabloid and broadsheet press, which up until this point, has not been properly researched.

Chapter 3 - Methodology

This chapter on methodology shall first of all introduce corpora and the characteristics that define a corpus (Section 3.1). In section 3.2 I shall detail the purpose and rationale for compiling a corpus for this dissertation. Section 3.3 explains the design of the corpus, looking in particular at which tennis matches and which newspapers were selected. Section 3.4 shall outline the step-by-step process of how the corpus was compiled. Finally, in

Section 3.5 I shall describe the procedures for data analysis, including the production of word lists and keyword lists and how these keyword lists were analysed.

3.1 - Corpora

The second half of the 20th century saw the development of corpus linguistics and the wider accessibility of computerized corpus software and it was this that allowed linguists to analyse vast amounts of digital forms of lexical data, which in turn led to the development of corpus linguistics (McCarthy and O'Keefe, 2010, pp. 5-6). Containing over 450 million words, one of the largest examples of a general corpus has been the creation of the Corpus of Contemporary American English (available at http://www.corpus.byu.edu/coca/). Similarly, much smaller, more specialised corpora have also been created with useful findings resulting from their analysis (Koester, 2010). The examination and manipulation of findings from these different types of corpora have been useful in the advancement of research and application into such linguistic areas as lexicography, language learning, socio-linguistic studies, corpus-based grammars and the historical evolution of language over time (Bowker and Pearson, 2002).

3.1.1 - Characteristics of corpora

The initial design of a new corpus must adhere to certain characteristics that are required to produce data reliable enough for linguistic analysis. There are four features of corpora that literature on this topic consistently highlights.

Firstly, there must be a framework of external criteria for which the corpus is being produced (Sinclair, 2005). This could include the aims of a study or the purpose for which a corpus is compiled. It is this framework which will ultimately define the design of the corpus as it will help the analyst decide which texts are relevant, how large the corpus should be and whether the texts should be spoken or written.

Secondly, a corpus should be representative, meaning that the texts selected for the corpus must be relevant to the use of language being investigated (Tognini-Bonelli, 2001). For example, if examining the language of tennis match reports, it would be unrepresentative to include reports from other sports or contain tennis articles which are not reporting a match, like blogs or editorials. The representativeness of a corpus can also be improved by increasing the number of texts included in the corpus, so long as the relevancy of each article is adhered to.

Thirdly, a corpus must be a compilation of naturally occurring, written or transcribed spoken texts, or a combination of both (Conrad, 2002). Naturally occurring texts are vital to the reliability of a corpus for reasons of authenticity. Texts cannot be purposely created for the inclusion in a corpus. Instead they should be genuine language produced according to their communicative function in the community in which they arose (Bowker and Pearson, 2002; Sinclair, 2005).

Finally, Conrad (2002) makes the point that a corpus must be available to access electronically. In doing so, various types of software can be used to compute, examine and decipher the discourse more readily and more accurately.

3.2 - The purpose and rationale of corpus compilation

Andy Murray is a successful tennis player who has won several major titles including the US Open, Wimbledon and an Olympic gold medal. He has also lost in major finals including once at Wimbledon, once at the US Open and four times at the Australian Open. With both his unsuccessful and successful matches widely reported in the UK, there is a growing public perception that the media refer to Andy Murray as British when he wins and Scottish when he loses (Clarke, 2013; Mesure, 2012; Viner, 2010). People often have intuitions about language use such as this one, but these perceptions are sometimes not reliable. The main purpose for creating this corpus is to challenge the accuracy of these ideas and discover whether there is evidence to support this public perception.

The corpus compiled for this dissertation is thus a collection of online texts written by a selection of UK newspapers. Each text is an article reporting on an Andy Murray tennis match at Wimbledon between 2005 and 2014. Linguistic analysis of this specialized corpus aims to examine the language of unsuccessful match reports against the language of successful match reports. The following two aspects shall be considered:
(i) the relationships between how Murray's national identity is referred to in match reports when he loses compared to when he wins; and
(ii) the language of the local, national, broadsheet and tabloid media in relation to whether Murray loses or wins

The rationale for creating such a corpus was simply that there are none available which meet the needs of this study. There have been other corpora which have been created to assess newspaper discourse. For example Scott and Tribble (2006) created a corpus of articles from *The Guardian* newspaper and sought to investigate the general themes of the

publication. They used key word searches to examine gender balance and the shifting news agenda of the paper over time. However there are no corpora of this specialised nature which would suit this study.

3.3 - Designing the corpus

As mentioned in Section 3.1.1, the initial design of a corpus is often defined by the purpose for which it is compiled. This study aimed to investigate the language of newspaper tennis match reports and so in this section I shall detail the selection process regarding which tennis matches were to be included in the corpus and from which newspaper titles the texts should be taken from.

3.3.1 - Selecting which tennis matches

I decided to build my corpus using 182 Andy Murray tennis matches played exclusively at the Wimbledon Championships, ranging from the start of his career in 2005 until 2014. This decision was made for three main reasons. Firstly, perceptions of Murray's identity are likely to have been influenced by the match reports which are of the most interest to the public. These are likely to have come from his matches at Wimbledon as they take place in London and receive live coverage on terrestrial television through the BBC, thus generating a far bigger UK audience than tournaments held abroad. Secondly, by including all of Murray's matches from Wimbledon, a mixture of both unsuccessful and successful reports could be included in the corpus without having to be selective about which matches to keep and which to disregard. Finally, by including match reports from 2005 through to 2014, the corpus would be more representative of the language that will have developed and evolved throughout Murray's entire career.

3.3.2 - Selecting which newspapers

I selected four newspapers to represent the four spectrums of the UK media (local, national, broadsheet and tabloid) that were specified in the research questions: The national broadsheet is represented by *The Guardian*; the national tabloid is represented by *The Mirror*; the local broadsheet is represented by *The Scotsman*; and the local tabloid is represented by *The Daily Record*.

The selection process used to determine these papers was closely scrutinised in order to ensure a fair representation of the UK's local, national, broadsheet and tabloid media.

Firstly, I took into account the complicated nature of the UK's media. According to Rosie et al. (2004), technological advancements over the last twenty years have allowed the news industry to make advancements in news gathering, editing and printing on a far more local level throughout the UK. This has resulted in the production of more geographically niche editions of what were once nationwide papers, including *The Scottish Daily Mail* and *The Scottish Sun*. However, the online content of these newspapers takes on the national form and so differs from the printed Scottish editions. For this reason, I avoided the inclusion of such titles as they would be unrepresentative of a local newspaper for this investigation.

Secondly, I considered the readership numbers of the UK press. For consistency, I aimed to include papers that acquired a relatively similar proportion of readers relative to its market. Regarding the local papers, *The Daily Record* in the most popular tabloid sold in Scotland and acquires a 642,000 net reach across its print and online versions of it newspaper every day, according to the National Readership Survey (2015). *The Scotsman* meanwhile receives 120,000 daily readers and is the second most popular broadsheet sold in Scotland (National Readership Survey, 2015).

The national broadsheet *The Guardian* has a net reach of 1,494,000 across both print and online platforms, while the national tabloid, *The Mirror* acquires 2,292,000 readers daily (National Readership Survey, 2015). While both of these newspapers are among the leaders in their specific markets, there are further reasons for selecting these two newspapers. For example, *The Mirror*, whilst sold throughout Britain, receives its strongest readership numbers in the north of England. It would be wrong then to pair this newspaper with *The Daily Telegraph* on a national level, as its strongest readership is concentrated to London and the south of England. It would be inconsistent to select 'national' newspapers that received readership numbers from geographically different regions of UK. *The Guardian* on the other hand, is more similar on a national level to *The Mirror* as it is similarly popular in the north of England and so these two papers included together would make a more consistent corpus (Rosie et al., 2004).

3.4 - Corpus compilation

I shall now outline the step by step process of how the corpus was compiled.
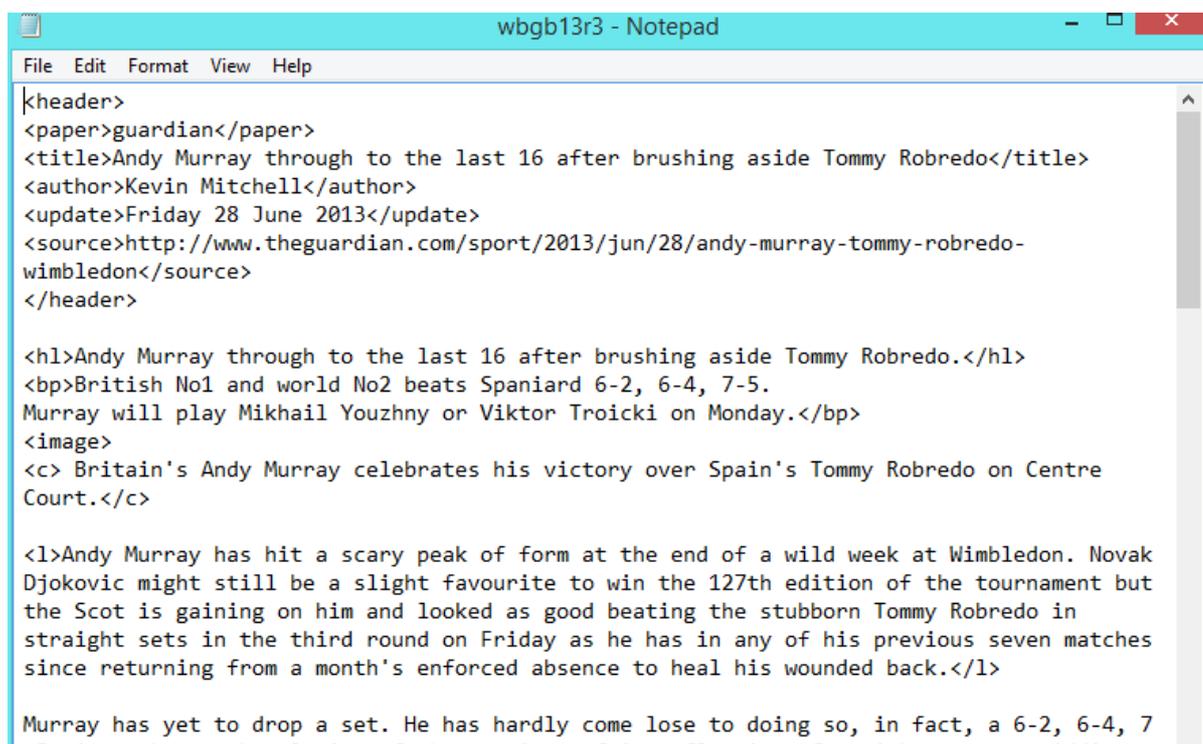
3.4.1 - Locating the match reports

Having decided upon exactly which match reports the corpus would be made up of, the next step in the process was locating these reports on the internet. As the media coverage of Andy Murray has increased the longer his career has gone on, I initially decided to locate the URL's of the 2014 match reports of all four newspapers first, with a plan to then work backwards year by year. However, beginning with *The Mirror* newspaper website, I soon discovered that it would be quicker to familiarise myself with one newspaper's search engine interface and locate as many URLs as I could before switching to another website. Each webpage ULR was saved to a bookmark folder located on the browser. The same process was then repeated for the three other newspapers.

One issue that was addressed during this process was that *The Guardian* website often produced as many as three different articles on the same match; an 'as it happened' match report, an initial match report, and a more reflective and opinionated match report later in the day. To ensure consistency, due care and attention was taken to select the 'initial' match report as it was most similar in genre and form to the match reports from the other newspapers.

3.4.2 - Creating a header

For the language of the online reports to be processed by computerised corpus software, the article must first of all be converted into a text file. The first step in this process was to choose which information from the webpage should be contained in the header of the text file and which should be kept in the main body for corpus analysis. As seen in Figure 1, the header contains the name of the paper, the title of the article, the author, the last update of the publication and the report's URL address, as these are all useful for referencing and cataloguing the match report.

Figure 1: Screenshot of a text file

With the exception of the title though, this information was not included in the main body of the text as it was not relevant to the language which was to be analysed for this study. A sample of a completed text file can be found in Appendix 1.

3.4.3 - Tagging the text

One of the main structural features of media discourse is to have a large print headline, followed by a subheading or bullet points in smaller print, then a lead paragraph followed by a body of text (Biber and Conrad, 2009). According to Delin (2000), trainee journalists are taught to follow an *inverted pyramid* where the headlines and lead mention the most crucial events, while the information in the body gradually decreases in value (p. 18). This also means that there is a departure from chronology, preferring instead to use an order of importance to structure the news story (Delin, 2000, p. 19). Events from the future are thus brought to the beginning of the article. As the information contained in these sections is of particular relevance when analysing the discourse, these areas have been tagged in the text file. As can be seen in Figure 1, <hl> tags the headline used in the article, <l> represents the lead paragraph, and <bp> tags the use of bullet points.

Furthermore, when a photograph has been removed from the article, it has been replaced with the tag <image>, however the caption supporting the image has been kept as part of the body and tagged with <c>. All of these linguistic devices are important to the purpose

of this study as headlines, captions, bullet points and lead paragraphs contain language relevant for analysis. By appearing in the main body of text, the tags provide a useful reference to these sentences during the analysis stage of the study.

3.4.4 - Filing of the Text files

Also seen in Figure 1 is an example of a file name. The file naming system is detailed below in Table 1.

Feature	Type	Code
Match result	Win	w
	Loss	l
Type of publication	Broadsheet	b
	Tabloid	t
Newspaper	*The Guardian*	g
	The Scotsman	s
	The Mirror	m
	The Daily Record	r
Area of publication	Local	s
	National	b
Year	20_ _	13
Round	Round 1	r1
	Round 2	r2
	Round 3	r3
	Round 4	r4
	Quarter-Final	qf
	Semi-Final	sf
	Final	f

Table 1: File naming system

The purpose of setting out a file naming system is to allow a large number of files to be sequenced, arranged and categorised into useful lists that correspond with the research questions. As the main aim of this study is to investigate the language of Andy Murray matches when he loses compared to when he wins, the loss/win notation appears first on the file name. Win or loss files will automatically be grouped together and easily located when needed for creating word lists. Other features which can be used to differentiate the files like publication type, paper name, area of publication, year and round are also included so that the exact content of the files can be known from their file name.

3.5 - Procedures for data analysis

This section will describe the process of how the data was analysed from the corpus. This section shall explain the production of word lists and how these are then used to produce the keyword lists from which the language analysis takes place.

3.5.1 - The production of word lists

In order to analyse and compare the language of losses with wins from the local, national, broadsheet and tabloid media, a word list for each of these variables was compiled using the computer software package *WordSmith Tools* (Scott, 2012). WordSmith is a corpus analysis program that allows users to input a corpus and, using the data, produce word lists, concordance lines, key words, word clusters and other corpus linguistic functions and techniques (Scott, 2012). A word list contains a list of every word that appears in a given corpus, along with vast amounts of statistics connected to each word. Whilst it has various functions for analysts, its primary role in this investigation was to provide the word list required to produce keyword lists.

3.5.2 - The production of keyword lists

Keyword lists are compiled when two word lists are contrasted to produce a new list of words that are deemed unusually frequent in a given set of texts (Scott, 2010). Usually, a keyword list is produced by comparing a specific word list with a more general, reference corpus, like the British National Corpus. However, the purpose of producing keyword lists for this dissertation was to investigate the linguistic differences between the unsuccessful and successful tennis match reports of the local, national, broadsheet and tabloid media. Thus, the unsuccessful word lists representing each of these categories were compared with the successful word lists to produce a keyword list for each newspaper. Again, the

computer software package WordSmith Tools was used for this process (Scott, 2012). The resulting data is a new, much shorter list of words which the software has determined statistically key to the language of either the unsuccessful texts or the successful texts. It is with these keywords on which the analysis of this investigation is based on. The four keyword lists produced for this dissertation can be found in the appendices 2, 3,4 and 5.

3.5.3 - Keywords

The words contained on a key-list are ranked in order of their keyness. The notion of keyness represents the sum of a probability calculation made by WordSmith Tools that takes into account sub-corpus size and word frequency (Baker, 2006). The words that are produced represent the vocabulary that has a strong significance due to their saliency in the discourse. Keywords with a positive keyness are significant to the sub-corpus, which for this study was the unsuccessful reports. The words which appear on the bottom of a keyword list are those with a negative keyness, and this means they are significant to the language of the successful match reports.

One important alteration that had to be made to WordSmith during the production of these keyword lists was to increase the default probability value (p) from <0.000001 to <0.001. As the original setting produced too few results, this change allowed words with a slightly lower probability of keyness to be included on the list. Although Baker (2006) states that by increasing the (p) value of a keyword list, you increase the chance that a word's keyness may be due to chance rather than the author's conscious word choice, the setting of 0.001 was still within the range commonly used in corpus linguistics.

3.5.4 - Keyword analysis

With a final list of keywords for each of the local, national, broadsheet and tabloid match reports, I then went through the process required to analyse each word. The first step was to take a keyword list and then categorize each keyword from it into a specific semantic group. The topics of these groups included *identity, the crowd, sympathy for Murray, criticism for Murray, loss of a match, tennis lexis* and others and a full list can be found throughout Chapter 4. The main purpose of dividing the keywords into semantic groups was so that the words within these groups could be analysed with those from the corresponding groups of other keyword lists.

The second step was to analyse each keyword so as to determine what it reveals about the language. This was primarily done by examining the concordance lines for each word i.e. the language which precedes and follows a keyword. Using these concordance lines, various information can be determined about a keyword and how it has been used in the texts. For example, they can show the most typical meaning, collocates and phraseology of a particular word. A greater understanding of a keyword's context can also be ascertained, which allows the analyst to notice subtle forms of language, like inferences. Aside from the concordance lines, further useful information about a keyword can also be gained from the statistics shown on the keyword list, like its frequency, its normalised frequency and the number of texts it appears in. These statistics can be seen in the keyword lists shown in Appendix 2-5.

This process was carried for the four keyword lists produced for this study, meaning that the language significant to the unsuccessful versus unsuccessful match reports by the local, national, broadsheet and tabloid newspapers was analysed. The results of this are shown in Chapter 4.

Chapter 4 - Results and discussion

This section shall examine the language of the Andy Murray match reports from a losses versus wins perspective. To do this, a keyword analysis of the unsuccessful and successful match reports of each of the four newspapers was made and discussed.

4.1 - Unsuccessful versus successful match reports in the national broadsheet
(The Guardian)

Section 4.1 shall analyse the keyword list of unsuccessful versus successful match reports taken from the national broadsheet newspaper, *The Guardian*. Each key word shall be discussed within its specific semantic group. The keyword list from which these results are taken can be found in Appendix 2.

4.1.1 - Identity

In this semantic group, an analysis of the keywords relating to how people are identified shall be made.

The most obvious results displayed on *The Guardian*'s keyword list relate to the names of Andy Murray's opponents during the matches he lost. Five of the six men who have registered victories against Murray at Wimbledon appear with high keyness values: '**Dimitrov**' defeated Murray in 2014, '**Nadal**' in 2008, 2010 and 2011, '**Federer**' in 2012, '**Nalbandian**' in 2005 and '**Roddick**' in 2009. These proper nouns likely have high keyness values because of two factors. Firstly, aside from Roddick, Andy Murray has never recorded a victory over any of these players at Wimbledon, meaning their names would be unlikely to appear in the wins corpus. Secondly, due to the smaller size of the unsuccessful match reports corpus, the names of Dimitrov, Nadal, Federer, Nalbandian and Roddick will likely have generated a higher normalised frequency compared to the names of opponents found in the wins corpus. These two factors would also account for the appearance of the possessive proper nouns '**Nadal's**' and '**Federer's**' on the keyword list. Finally, the absence of Marcos Baghdatis, the sixth man to have beaten Murray, may be explained by the fact that Murray also defeated Baghdatis in a subsequent year. This meant his name featured in both successful and unsuccessful match reports, lessoning the keyness value of his name.

Regarding the question of whether Murray is British when he wins and Scottish when he loses, there is nothing on this keyword list to suggest that this is the case with national broadsheets. There are no direct references or inferences towards either of these national identities, which perhaps suggest that *The Guardian* newspaper does not overtly choose one phrasing over another.

However, the pronouns '**he**' and '**I**' do appear. The greater prevalence of '**he**' in unsuccessful match reports perhaps suggest that the authors prefer to limit their use of identity markers when Murray loses. This may be because as Murray was unsuccessful, there is no need to proudly proclaim or emphasize a British or Scottish national achievement. The literature suggested that nationality was often used by the media to cement its identification with its readership (Bale, 1986). However, when the story in question does not positively reflect that identity, the newspapers may be less inclined to promote it. This may account for 'he' being the preferred form of reference for Andy Murray when he losses.

'**I**' meanwhile, is found to have a negative key value, which means it is more specific to the corpus of successful match reports. The concordance line for this word shows that it most commonly occurs within a set of quotation marks. Interviews with Andy Murray are

common within the successful tennis match articles. The reason that 'I' may be particularly key to successful match reports is because when Murray wins, he will likely talk about himself and the match from his perspective.

> '"**I** played well throughout," said Murray.'

However, in defeat, the concordance lines reveal that Murray tends to reference his opponents performance, rather than his own.

> '"**He** played so much better than me," the Scot admitted.'

This results in a decrease of the word 'I' during unsuccessful reports, while also an increase in 'he', which further explains why it is a key word in loss articles.

Furthermore, the fewer occurrences of 'I' in unsuccessful matches may also result because the journalists opt to give less of a voice to Andy Murray when he loses. Analysis of the loss articles themselves show that in some articles, no interview with Murray is referenced at all.

4.1.2 - The crowd

In this semantic group, an analysis of the keywords relating to the crowd watching the match shall be made. First of all, the noun '**applause**' is used when detailing the reaction of the Wimbledon crowd when watching a match. For example,

> 'The crowd were entranced and slowly, slowly, the **applause** grew in volume.'

However, this word is perhaps not that significant to the language of unsuccessful match reports as analysis shows that four of its five occurrences are contained within only one of *The Guardian's* articles. This would suggest that 'applause' may only be reflective of that particular match report, and less key when comparing losses versus wins.

On a similar theme, the **crowd** at Wimbledon seem to be mentioned far more often in the unsuccessful match reports. The concordance lines show that the authors tend to reference the crowd in order to highlight the level of attention, expectation and pressure that surrounds Murray when he plays at Wimbledon;

> 'The **crowd** on Centre Court came to witness a moment in history.'

Since 2008, Murray has reached at least the quarter finals of every Wimbledon championship. This means that Murray's defeats always come when the level of public interest in the tournament is peaking and the pressure on him is at its greatest. Compare this to his 1st, 2nd, 3rd and 4th round matches which make up the majority of the wins corpus article. These matches take place during the quieter first week of Wimbledon meaning the reports would likely speak far less of the large crowd expectation and anticipation on Murray.

4.1.3 - Sympathy for Murray

There are suggestions that the national broadsheet authors are generally sympathetic and quite positive about Murray when he is defeated. This semantic group shall analyse the keywords related to this area.

One example of this is where the adverb '***simply***' is used to describe a loss for Murray, but at the same time, infer that he has nothing to be ashamed about. For example,

'*Baghdatis…had **simply** been the steadier.*'

In this case Murray is not portrayed negatively. Instead, his opponent receives praise and the implication is made that Murray was not at fault, but simply beaten by someone better on that day. This praise of opponents on the day is also responsible for the appearance of '**rarely**' on the key word list,

'*The American can **rarely** have played such a tactical game.*'

This sentence again shows the authors choosing to heap praise on Murray's opponent, whilst inferring from this that Murray was not accountable for the loss.

The word '**age**' is also used sympathetically towards Murray. The data shows that the word is used particularly prominently when Murray was defeated at his first ever Wimbledon in 2005 when he was 18 years old.

'*Murray's potential was evident from a young **age.***'

'*He is almost preternaturally confident for his **age**, but…*'

The authors make the positive point that, being so young and with such a promising future ahead, there was certainly no disgrace in defeat.

There is one final example highlighting compassion for Murray. The concordance lines for the modal verb '**could**', show that its most common collocates in the L1, R1 and R2 positions create the phrase '*he could not quite*'.

> '*What Murray needed was a sustained phase of magic, something to feed the doubt again into Federer's mind, as he had done at the beginning. He **could** not quite find enough moments.*'

This type of phrasing infers that Murray fought valiantly and was in touching distance of victory but lost by only the narrowest of margins.

4.1.4 - Criticism of Murray

This semantic group shall analyse two examples in which the authors appear less sympathetic towards Murray. Firstly, the concordance lines for the word '**wanted**' show that it is often followed by a 'but' or an 'although', and in these cases, the authors contrast the positivity of the crowd to the negative performance of Murray. For example,

> '*The crowd desperately **wanted** him to survive, but he was doing little to encourage their slim hopes.*'

The wishful sentiment of the word 'wanted' is perhaps key to the unsuccessful match reports as it likely contrasts with the materialisation of this wish in the successful articles.

A second example of a more critical approach by the authors is shown by the inclusion of the word '**another**' in the keyword list. This keyword is most commonly used in the unsuccessful articles to summarise and contextualise the defeat.

> '*Just **another** British loss in a Wimbledon men's semi-final.*'

This negative language is perhaps not aimed so much at Andy Murray, but instead highlighting the many years of continual disappointment that UK tennis fans have had to endure. This example is taken from a year prior to Murray finally winning Wimbledon in 2013. Before then, the UK was without a male winner for over 75 years. This contextual type of lexis is likely not to appear in the successful match reports, for as long as Murray is winning, the authors need not to summarise past failures.

There is a semantic group of words which clearly define the unsuccessful nature of the match from Andy Murray's perspective. The word '**defeat**' exemplifies the general tone of the articles reporting a loss:

> '*Murray, meanwhile was as devastated in **defeat** as he was after losing to Roger Federer.*'

Meanwhile the keywords of '**disappointment**', '**tears**' and '**aggression**' detail the emotions of Andy Murray during and after a defeat. For example:

> '*Young Murray could not hide his **disappointment** at losing to Marcos Baghdatis.*'

> '*Andy Murray driven to **tears** by Roger Federer.*'

> '*"Raise you game Andy!" came the shout [from Murray], an edge of **aggression** in the call as Murray struggled.*'

The concordance lines reveal that the inclusion of the word 'tears' refers solely to when Murray cried during an on-court interview after his 2012 final defeat, which explains why its three instances were all found in a single text. Without any other noticeable incidents involving tears during Murray's career at Wimbledon, this lexis is thus unlikely to be referenced in the corpus of successful match reports. Similarly, as the authors tend not to give a voice to the losing opponents of Murray, 'disappointment' and 'aggression' are also unlikely to be used as often in the successful articles.

Finally, he word '**cracks**' is used to describe the gradual breakdown in Murray's game, shown best in the following example,

> '*The first **cracks** began to appear.*'

Overall, each of these words is used to tell the story of the loser of the match rather than the victor. Because the words appear with high keyness values in the unsuccessful match reports, it means that the successful match reports do not tell the story of the losing player (i.e. the one who has been beaten by Murray) in this same way. As Murray is always the losing player in the unsuccessful match reports and always the focus of these words, it would suggest that, win or lose, he remains the main subject of every article. With the narrative focused on Murray and not his opponent, the authors seem more inclined to tell

Murray's story than report each match from a more neutral perspective. This is understandable given UK based readership of this newspaper.

4.1.6 - Tennis lexis

This semantic field looks at the keywords which are used in a purely tennis context. Analysis shows that the word '**final**' is key to unsuccessful match reports because firstly, one of Murray's defeats took place in the Wimbledon final. Secondly, three of Murray's eight defeats took place in semi-finals and in these articles, the authors commonly refer to the fact that should Murray have won he would have reached his first final. For example,

> '*Murray had felt he would reach his first **final** but Roddick had not finished.*'

Similar reasoning may account for the appearance or '**round**', where it is key to successful reports. As explained earlier, the majority of Murray's wins takes place in the early rounds of Wimbledon and so the authors often mention that by winning, Murray will progress to the next round, as seen in this concordance line'

> '*Stormed into the fourth **round** with a straight sets win.*'

Finally, '**6-3**' is found with a negative key value meaning it was significant to successful match reports. It is the score of a set in tennis and likely occurs simply because that particular set score occurred on relatively fewer occasions when Murray lost compared to when he won.

> '*Murray stretches for the ball en route to a 6-2, **6-3**, 6-2 victory against Roberto Bautista Agut.*'

However, the concordance lines also reveal that this set score often occurs on its own in successful match reports, but not in unsuccessful reports.

> '*Murray held it together well to take the set **6-3** when Haas pushed a forehand long.*'

This suggests that in successful articles the authors are more likely to report the match in more detail, describing each individual set.

4.1.7 - Other lexis

The concordance lines of the word '**yesterday's**' reveal that on all three occurrences, it has been taken from articles which have been published the day after the match on which they report.

> '*There was a pivot to* **yesterday's** *contest that stands out.*'

Finally, the word '**before**' is used sparingly in the unsuccessful reports and instead has a high keyness value in the wins corpus.

> '*Murray won the first two sets* **before** *cramp severely restricted his movements.*'

Analysis shows it is often used when describing detailed action during the game where the temporal reference within the match is key.

These other lexis seems to show that in unsuccessful match reports, the key temporal reference lies in the relationship between the match and the publication of the report; while in successful match reports, it is the time relationship of episodes within the match that is of relevance.

This completes the keyword analysis for *The Guardian* newspaper.

4.2 - Unsuccessful versus successful match reports in the national tabloid (*The Mirror*)

Section 4.2 shall analyse the keyword list of unsuccessful versus successful match reports taken from the national tabloid newspaper, *The Mirror*. Each key word shall be discussed within its specific semantic field. The keyword list from which these results are taken can be found in Appendix 3.

4.2.1 – Identity

In this semantic group, an analysis of the keywords relating to how people are identified shall be made and expectedly, the keyword list for *The Mirror* includes the names of some of the players who have defeated Murray at Wimbledon. Their surnames comprise the three most salient keywords: '**Roddick**', '**Federer**', and '**Nadal**'. Meanwhile in fifth is '**Rafael**', the forename of Nadal. Further down the list, the possessive form of Andy '**Murray's**' name appears and is used most commonly to make a statement regarding his exit from the tournament.

'*Hearts were once again broken as Andy **Murray's** Wimbledon dream ended.*'

These names all suggest that elements of personalisation are perhaps more apparent in the unsuccessful match reports. The literature suggested this is a form of tabloidization, however similar results were also displayed in the national broadsheet's keyword list in Section 4.1.1 (Uribe and Gunter, 2004). But added to this is the word '**Swiss**', which is used to reference the nationality of Roger Federer. The Swiss player played Andy Murray twice at Wimbledon, winning once and losing once, but curiously his nationality was mentioned 6 times as often in the article in which he won. In this instance, national referencing of the opponent does seem to be linked semantically to matches in which Murray is defeated and is perhaps further evidence of tabloidization of these articles too. There are however, no keywords on the list to suggest that *The Mirror*'s portrayal of Andy Murray's national identity relates to whether he wins or loses.

4.2.2 - The loss of a match

Key to the loss articles are various words used by the authors to summarize the unsuccessful story of Murray's Wimbledon campaign. '**Bidding**' for example, is used when the authors describe how Murray was attempting to become the first British male in over 70 years to either reach or win the Wimbledon final.

'*The word number four – **bidding** to become the first home winner.*'

'*The 24-year old was **bidding** to end a 75 year drought.*'

Meanwhile, analysis reveals that the words '**dream**' and '**hopes**' often occur in the vicinity of 'ended' and are used to conclude Murray's campaign after he has been defeated.

'*Andy Murray's Wimbledon **dream** was **ended.***'

'*Andy Murray's **hopes** of Wimbledon glory **ended** tonight.*'

When Murray is defeated, Wimbledon is effectively over from a UK perspective. Though the tournament continues with the remaining players, the keywords mentioned above suggest that *The Mirror* authors use this opportunity to contextualise and summarise the tournament from a UK point of view. In successful match reports, there is no need for this to happen and so this perhaps explains why they are specific to the losses corpus.

The reaction of the crowd to the loss of a match is also referenced in this keyword list. For each of the three times that the word '**stunned**' appears in the unsuccessful match reports, it is used to describe the reaction of the crowd.

> *'Centre court crowd were left **stunned** from the off.'*

This is further evidence that the crowd seem to play a larger role in the lexis of articles when Murray loses. This may be done so as to connect with the readership of the newspaper. As Murray is Scottish, there may be a need for the national newspapers to show that the English crowds are also part of the article, so that their readership can identify and connect better with the story.

4.2.3 - Tennis lexis

The keywords contained in this semantic group of tennis lexis all focus heavily on the point-by-point details of the tennis match. '**Percentage**' and '**first-serve**' co-occur describing how often Murray faulted while serving, while a '**volley**' is a tennis shot taken before the ball has bounced.

> *'Murray's **first-serve percentage** had dropped below 50%.'*

> *'Finishing it with a drop **volley** to take the set 7-5.'*

Each of these sentences explain very specific aspects of a tennis match and show that the authors have chosen to describe the matches when Andy Murray loses with more detail than when he wins. This may be because the successful articles might skim over the details of the match, preferring instead to look forward and build up his next match.

4.2.4 - Other lexis

The word '**that**' has a high key value in the successful match reports and it is often used as a conjunction so as to express a result.

> *'A forehand so powerful **that** it appeared to knock Robredo of his feet.'*

This completes the keyword analysis for *The Mirror* newspaper.

4.3 - Unsuccessful versus successful match reports in the local broadsheet (*The Scotsman*)

Section 4.3 shall analyse the keyword list of unsuccessful versus successful match reports taken from the local broadsheet newspaper, *The Scotsman*. Each key word shall be discussed within its specific semantic group. The keyword list from which these results are taken can be found in Appendix 4.

4.3.1 - Identity

This semantic group considers the keywords relating to how people are identified within *The Scotsman* articles. The names of Murray's opponents during matches when he lost all appear on *The Scotsman*'s keyword list: '**Nadal**', '**Dimitrov**', '**Roddick**', '**Federer**' and '**Nalbandian**'. The possessive form of **'Nadal's'** name is also found. While these names may seem to suggest that personalisation is more prevalent in unsuccessful match reports, as discussed in Section 4.1.1, the more likely reasoning is that these players (excluding Roddick) have never lost to Murray at Wimbledon and so their names would rarely be mentioned in successful reports. Furthermore, the pronoun '**they**' is key to the successful reports and analysis of this word shows that it often replaces the proper nouns of tennis player names.

> '**They** *were tied on one game each in the third set.'*

The previous literature suggested that an increase in nationalism was a theme of sports reporting and this may be exemplified by the presence of '**Argentine'** on the list (Rosie et al., 2004). David Nalbandian, Murray's opponent in 2005, is the Argentinian player who this term is referencing.

The pronoun '**I**' appears on the list showing that it is key to the wins articles. As discussed in Section 4.1.1, '**I**' is most often located during sections of the article where a player, usually Andy Murray is being interviewed.

> '*Murray said: "***I*** served really great. ***I*** didn't give him many opportunities at all."'*

The decreased prevalence of the word '**I**' in matches lost by Andy Murray suggests that the journalists choose not to give a voice to the player when he is unsuccessful. One could argue that Murray's opponent could have been interviewed when Murray lost, but neither he nor his opponent seems to be given proper space when Murray loses.

Regarding Murray's national identity, the findings from this keyword list do not show any evidence to suggest that it is influenced by whether he wins or loses. Neither his Scottish nor his British nationality occurs on the keyword list, nor any other term or phrase that could be connected to this topic.

4.3.2 - Praise of Murray

In this semantic field, an analysis of the keywords which seem to praise Murray shall be made. The adjective '**amazing**' appears with a high keyness value and highlights some of the positive language used in the unsuccessful reports. The concordance lines show that this word is first used to describe Murray's debut Wimbledon campaign in 2005, which ended in defeat to David Nalbandian. The author calls Murray 'amazing' no less than three times during the report of this match.

> '*Murray will take heart from what has been an **amazing** debut at Wimbledon*.'

However, the rest of the concordance lines are taken from Murray's Wimbledon final defeat in 2013, and all refer to his opponent, Roger Federer.

> '*"It's **amazing**. It equals me with Pete Sampras, who is my hero," said Federer.*'

Similarly the word '**top**' is used to credit both Murray and his opponents.

> '*Murray got on **top** in a rally.*'

> '*Roddick, who hit **top** gear with his service game.*'

Both 'amazing' and 'top' show that the local broadsheet authors use a positive tone during unsuccessful articles, by both praising Murray's effort and the quality of his opponent.

Overall, there is very little negative language to appear on this keyword list. With Murray losing and UK hopes ended, one would expect that the semantic nature of unsuccessful articles would be fairly sombre in tone. However this finding justifies the use of corpus linguistic techniques as it shows that while people have intuitions about language use, these perceptions are not always correct.

4.3.3 - Loss of a match

There is only one word on this list which relates semantically to losing a match. In comparison to the previous section (4.3.2), the tone is seen to be contrasted by the

appearance of '**tears**' on the list, as it suggests that more sombre themes are also present in unsuccessful articles.

> '*While for Murray there were **tears** of despair.'*

However, this word is also used to report Federer's tears of joy at winning.

> '*Federer collapsed to the grass with **tears** in his eyes.'*

In *The Scotsman*, 'tears' refers to both the joy of Federer as well as despair for Murray, whereas in *The Guardian* it relates only to Murray's disappointment (Section 4.1.5). This keyword thus highlights the difference in the language used between these two broadsheets and perhaps suggests that the local media may tend not to focus as much attention on Murray during unsuccessful reports.

4.3.4 - Tennis Lexis

There is an abundance of language present in this keyword list that is used only within the context of tennis matches. A '**tiebreak**' is a fairly infrequent, but crucial part of a tennis match, used to decide who wins a set when the scores are level.

> '*There was little or nothing between the two men all the way to the fourth set **tiebreak.'***

Murray's defeat to Roddick was decided by two of these tiebreaks in one match. With 6 separate references to tiebreaks in this match's article, this explains why it has a high keyness value within the losses corpus.

'**No3**' meanwhile refers to Murray's world ranking and seeding for tennis tournaments.

> '*Four years younger than the **No3** seed, Dimitrov looked, almost all of the match, like the senior player.'*

The world ranking of a player is usually only noteworthy when a highly ranked player is defeated by a player far below them in the rankings, as it can emphasize the surprise of the result. In most tournaments, Murray defeats players ranked below him in the early rounds before losing to a player like Nadal or Federer, who is ranked above him. This perhaps explains why his ranking is mentioned far less in successful articles. However in 2009 and 2014, Murray lost to Roddick and Dimitrov respectively, both of whom were ranked

considerably lower than Murray at that time. The concordance lines reveal that all mentions of 'No3' in unsuccessful articles occurred during these two matches. This would support the idea that the journalists used Murray's ranking to display surprise at these result.

Finally, the word '**round**' appears with a negative keyness value.

> '*Andy Murray has cruised into the fourth **round** at Wimbledon.'*

As explained in Section 4.1.6, this word probably occurs more frequently in the successful match reports because Murray normally wins in rounds 1, 2, 3 and 4, while he usually loses in the quarter-final, semi-final or final of the tournament.

4.3.5 - The crowd

The crowd are once again seen to be a key figure in the reporting of unsuccessful tennis match reports. In *The Scotsman*'s keyword list, the word '**fans,**' is used to reference the Wimbledon crowds.

> '*The 18-year old Scot put on an outstanding display of tennis for the **fans** in Centre Court.'*

While *The Guardian* authors used the crowd to emphasize the pressure Murray was under (Section 4.1.2), the local broadsheet tends to mention the crowd from a more positive perspective, implying the fans were often happy with Murray's performance. This also adds to the evidence that the local broadsheet paper carries a more upbeat tone compared to the national papers, particularly when Murray is defeated. This is perhaps explained by the literature which suggests that the local media tends to be more pro-Scottish in its language (MacInnes et al., 2007; Rosie et al., 2004). To write less positively about Murray may be construed as unpatriotic by their Scottish readership and so even in defeat, a more optimistic perspective is given to Andy Murray by *The Scotsman* newspaper.

4.3.6 - Other lexis

The keyword list shows that the past tense verb '**played**' was specific to unsuccessful articles.

> '*The Bulgarian **played** fearlessly*.'

'Played' is a rather general term to describe a player's action and this perhaps suggests that in successful match reports the authors tend to use greater variation in their language.

Finally, the word '**few**' appears with a negative keyness value and its concordance lines reveal it is most commonly preceded by 'a'.

> '*Obviously a **few** weeks ago I missed the French open.*'

Analysis also shows that this type of phrasing often occurs during spoken text, as an author would likely choose to be more specific. As seen in Section 4.3.1, in successful match reports the authors choose to given Murray more space for spoken text and so this perhaps accounts for why 'few' appears to be key to these articles.

This completes the keyword analysis for *The Scotsman* newspaper.

4.4 - Unsuccessful versus successful match reports in the local tabloid (*The Daily Record*)

Section 4.4 shall analyse the keyword list of unsuccessful versus successful match reports taken from the local tabloid newspaper, *The Daily Record*. Each key word shall be discussed within its specific semantic field. The keyword list from which these results are taken can be found in Appendix 5.

4.4.1 - Identity

This semantic field once again looks at how individuals are identified. The four highest keywords on *The Daily Record's* list are the surnames of the four players who have defeated Murray most recently at Wimbledon: '**Nadal**', '**Roddick**', '**Dimitrov**' and '**Federer**'. Two further identity references are also made to Rafael Nadal, including the possessive proper noun '**Nadal's**' and his nickname, '**Rafa**'. Whilst the appearance of these additional references on the unsuccessful list can be partly explained by the fact that Nadal has defeated Murray on three different occasions at Wimbledon, the use of a nickname within a news article may also suggest evidence of personalisation. The only other newspaper to identify a player on first name terms was the national tabloid, *The Mirror*, which used 'Rafael', and so these results seems to supports the ideas of Van Zoonen (as cited in Uribe and Gunter, 2004) who suggested that personalisation was a characteristic of the tabloid media.

Staying on the theme of identity, the previous literature had also argued that nationalisation was a characteristic of sports reporting by the UK media and of tabloids in general (Bairner, 2015; Conboy, 2007; Rosie et al., 2004). This keyword list for *The Daily Record* seems to support these ideas as it references the nationality of two of Andy Murray's opponents, with '**Swiss**' and '**Bulgarian**' referring to Roger Federer and Grigor Dimitrov respectively. However, nationality references were also found in the keyword lists of *The Scotsman* and *The Mirror* newspapers, suggesting that nationalisation is wide spread within the UK media, particularly when Andy Murray loses. Together these findings maintain Bloyce's (2010) view that sports journalists often adopt a pack mentality, perhaps becoming more nationalistic when Murray is faced with defeat.

Regarding Andy Murray's national identity, there is again no evidence that a loss versus win result has any notable impact on whether he is described as British or Scottish by the media.

A final identity reference found on this keyword list is the name '**Tsonga**', another of Murray's opponents. The findings show the word has a negative keyness value meaning it is specific to the successful match reports. Murray has defeated Tsonga twice at Wimbledon and has never lost to him, thus lessening the likelihood of Tsonga's name appearing in unsuccessful articles.

4.4.2 - The loss of a match

A notable aspect of the keyword list for *The Daily Record* newspaper are the words used to report the loss of a match for Andy Murray. The language used to do this is more descriptive than that seen in *The Guardian* in Section 4.1.5. For example, there is not just '**disappointment**' for Murray when he loses, but '**crushing**' disappointment according to the local tabloid authors. Along with '**suffered**' and '**heartbreak**', *The Daily Record* paints quite an emotionally striking picture when describing defeat for Murray.

> '*Andy Murray **suffered heartbreak** and **crushing disappointment** after he lost his first Wimbledon final to roger Federer.*'

4.4.3 - Criticism of Murray

The keyword list also suggests that compared to the other newspapers, *The Daily Record* is less sympathetic towards Murray in defeat. This semantic group shall look at some of these criticisms. For example, in Section 4.1.3 *The Guardian* uses the phrasing '*could not quite',*

which credits Murray for how close he came to victory. However, *The Daily Record* describes how Murray was,

> *'**Unable** to raise himself.'*

And,

> *'**Unable** to take advantage.'*

These fragments suggest that Murray was responsible for his own defeat and give no credit for his endeavours.

Another example is the word '**woefully**', which is used by *The Daily Record* authors to describe how Murray,

> '*Came up **woefully** short in the biggest match of his life.'*

Rather than choosing to praise his opponent, *The Daily Record* again blames Murray for the loss. This compares with Section 4.3.2 in which *The Mirror* is often seen to praise Murray despite his defeat.

A final example is the word '**crushing'**, which the concordance lines reveal was also used to describe one of Nadal's wins over Murray as a,

> '*Crushing 6-3 6-2 6-4 victory.'*

This adjective emphasizes the heaviness of the Murray loss and again offers him little sympathy.

Overall, it is perhaps surprising to find that the local Scottish tabloid, laden with Scottish imagery and rhetoric, is the most critical of the four newspapers towards Andy Murray (Meech and Kilborn, 1992).

4.4.4 - Tennis lexis

As found on each of the other keyword lists, there is a category of words that are related directly to the lexis of tennis matches in general. First of all is the word '**final**' which, as also seen in Section 4.1.6, appears due to Murray's 2012 final loss as well as his numerous semi-final defeats in which the articles reference the final. For example,

> *'Novak Djokovic was already waiting in the **final** and Murray's hopes of joining him were suddenly looking a lot brighter.'*

Also on the list is '**aces**', which is the plural form of an unreturned serve in tennis.

> *'Roddick began in typical fashion with two **aces** in the opening game.'*

'Aces' are achieved by players like Andy Roddick who have a very fast serve, and in his 2009 semi-final victory over Murray, the articles tend to focus on that part of Roddick's game. In fact, five out of the eight occurrences of the word 'aces', are found in the report of that match.

The word '**court**' refers to either Centre Court or Court One at Wimbledon or more generally to the playing surface of a tennis match. It is found to have a high keyness value in the successful match reports of *The Daily Record*.

> *'The Scot was scintillating on Centre **Court** as he coasted past his rival.'*

The appearance of 'court' may suggest that for successful matches, the authors deem in more necessary to state the location of the match. This may be because in the early stages of the tournament, when nearly all of Murray's matches are successful, the games take place on a number of different courts around Wimbledon. In the latter rounds however, most matches takes place on Centre Court and so the author may feel there is less need to remind the reader of the location.

As seen in Section 4.1.6, the set score line '**6-3**' has a high keyness value in successful match reports.

> *'Tsonga took the set **6-3** when Murray netted a backhand.'*

The concordance lines from *The Daily Record* also reveal that the authors tend to describe successful matches in more detail by reporting the score from each set.

4.4.5 - Other lexis

The word '**strength**' is used by the authors in matches when Murray lost to Rafael Nadal in 2008.

> *'Nadal's power play too much for Murray in test of **strength**.'*

It relates to the difference in the physical stature of the two players at that time and perhaps suggests that in unsuccessful articles, the national tabloid reports more in detail about Murray's opponent than in successful articles.

However, much of the remaining lexis on this list is targeted at Murray. For example, '**anything**' seems to be used to highlight the desperation of Murray's situation in unsuccessful match reports.

> '*I couldn't do **anything** about it.*'

And as seen in Section 4.4.3, the word '**short'** when used in collocation with 'woefully', describes how far Murray was from winning the match. It is also used to describe a more technical failure of Murray's game, detailing the poor length of his shots.

> '*His length dropped too **short** and Nadal made him pay.*'

With this word being key to unsuccessful matches, it shows that the authors tend to detail the failures of Murray's game to a greater extent than the failures of his opponents when Murray wins. This, as in Section 4.1.5, highlights how Andy Murray remains the main focus of each article, whether he wins or loses.

At first glance the word '**fashion**' suggests that the local tabloid paper may be concentrating on the attire worn by the players or the crowd during unsuccessful match reports. However, analysis of the concordance lines show that it is the manner in which something is done, that 'fashion' is describing.

> '*Roddick began in typical **fashion** with two aces.*'

The word '**seven'** is repeatedly used in unsuccessful match reports in order to inform the reader of certain contextual significances. One example was to mention that by defeating Murray, Federer had won a record equalling seventh Wimbledon title.

> '*Federer, who tied with Pete Sampras with **seven** Wimbledon titles.*'

Another example which informs the reader contextually, states that by 2011,

> '*Murray had now been in **seven** grand slam semi-finals without going on from one of them.*'

These types of statements help summarise the tournament and so they are more likely to be used in unsuccessful reports written when Andy Murray has been knocked out of the competition.

Finally, the word '**through**' is found to be key in *The Daily Record's* successful match reports. For example,

> '*Murray recovered from a shaky start to cruise **through** his first-round match.*'

This term would unlikely be used in matches that Murray did not progress from and so appears far less in unsuccessful articles.

This completes the keyword analysis for *The Daily Record* newspaper.

4.5 - Summary

This concludes the results and discussion chapter of this dissertation. A summary of the four newspaper's findings shall be made in Section 5.1 of the conclusion chapter.

Chapter 5 - Conclusion

This chapter shall discuss the conclusions of the investigation. First of all, the findings shall be summarised in regards to the research questions. Next, the investigation shall be discussed in terms of its limitations, before finally looking at what further research could be made in this area.

5.1 - Summary of findings

The analytical comprehensiveness of corpus linguistics meant that one of the main features of this investigation was that it produced findings in areas that were not initially anticipated or planned to be researched. The bottom-up nature of this type of research allowed for the production and analysis of findings related to Murray's national identity as well as results that covered areas far beyond this. These shall now be summarized in relation to the research questions.

5.1.1 - Murray's national identity

The first research question considered how Andy Murray's nationality is referred to in matches when he loses compared to matches when he wins. Across the four newspapers, the analysis showed no evidence that the outcome of Murray's matches influenced the media's use of language concerning his national identity. No keywords were seen to refer in any way to Murray's nationality in either unsuccessful reports or successful ones. These findings suggest that the intuition of Murray being British when he wins and Scottish when he loses, has no basis in terms of language use by the UK media.

5.1.2 - The language of the local, national, broadsheet and tabloid media

The second research question looked at the language of the local, national, broadsheet and tabloid media in relation to whether Murray loses or wins. This section shall be divided into the semantic groups related to their use of language.

5.1.2.1 - Identity

There was general consistency in the semantic field of identity as in all four newspaper, the names of Murray's opponents were key to the language of unsuccessful reports. The keyness of these identities was most likely explained by their relatively high frequency among loss articles, given the size of that corpora. Names like 'Nadal', who defeated Murray three times at Wimbledon, would be mentioned proportionately more often in the smaller losses corpus than in the much larger wins corpus for each newspaper.

This semantic field of identity also highlighted some of the differences in language use between the UK media. Firstly, there were references to players' national identities in some newspapers, but not in others. For example, both local titles, *The Scotsman* and *The Daily Record*, referenced the national identities of some of Murray's opponents. Literature by Rosie et al. (2004) suggested the language of local papers was often explicitly Scottish, however this evidence suggests that nationalistic language referencing other countries is also prevalent. *The Mirror* also referenced these national identities, which, along with *The Daily Record*, meant that both tabloid papers displayed nationalistic language. This concurred with the research of Conboy (2007). Overall, all three of these papers support the literature by Bairner (2015) who stated that the idea of nation versus nation was at the heart of every global sporting event.

A second difference was the use of first-name terms to identify players, which was displayed by tabloid titles but not by the broadsheets. While 'Rafa was a keyword in *The Mirror* and 'Rafael' was a keyword in *The Daily Record*, neither broadsheet used this type of identity marker, which perhaps suggests that personalisation is a more common feature of tabloid language. This corresponded with the literature by Uribe and Gunter (2004).

The semantic field of identity also highlighted key language only found in the broadsheet newspapers. The pronoun 'I' had the highest negative keyness value for both *The Guardian* and *The Scotsman*, but did not appear on either tabloid keyword list. This suggested that the broadsheet newspapers preferred to give a voice to the tennis players, only when Murray was successful in matches. The tabloid newspapers perhaps did not make this adjustment.

5.1.2.2 - The crowd

Another difference between the language of the broadsheet and tabloid newspapers concerned the semantic group, 'The crowd'. While there were no keywords related to this field in either tabloid paper, the broadsheets either spoke of the 'crowd' (*The Guardian*) or the 'fans' (*The Scotsman*). Referencing the crowd like this may be an attempt by the broadsheet newspapers to connect to their readership, given that both tennis and broadsheets likely attract a similar type of audience.

5.1.2.3 - Praise of Murray, Sympathy for Murray and Criticism of Murray

There were also differences between the positivity shown towards Andy Murray in unsuccessful match reports. The semantic group 'Praise of Murray' was only located in *The Scotsman*, while the 'Sympathy for Murray' group was only part of *The Guardian's* keyword list. This shows that both broadsheet papers tended to display more encouraging language towards Murray, compared with the tabloid papers. On the opposite side of this spectrum was *The Daily Record*, which often used heavily critical language of Murray, exemplified by the adverb 'woefully'. This type of language displays further the differences between each type of discourse.

5.1.2.4 - Tennis lexis

Within the other semantic fields the language of the local, national broadsheet and tabloid media tended to differ considerably. While tennis lexis was key to the language of all four newspapers, the keywords displayed by each list showed a great variation of terms. For

example, while *The Mirror* and *The Daily Record* displayed particularly niche tennis vocabulary like 'volley' and 'aces' respectively, the key words of *The Guardian* used more general tennis terms like 'final' and 'round'. This was quite surprising given that the research by had suggested that tabloid papers tended to simplify their discourse (Connell, as cited in Uribe and Gunter, 2004; Djupsund and Carlson, as cited in Uribe and Gunter, 2004; Rooney, as cited in Uribe and Gunter, 2004).

<u>5.1.2.5 - Loss of a match</u>

Generally the language contained in this semantic field remained fairly consistent. For example, the keyword 'disappointment' was found in both *The Guardian* and *The Daily Record* newspapers, while 'tears' appeared in both *The Guardian* and *The Scotsman*. These examples show that occasionally, the language used by the UK media can sometimes be consistent across local, national, broadsheet and tabloid papers. However, the other vocabulary in this semantic group did highlight more differences between the papers. The tabloids tended to use more descriptive and powerful language like 'crushing' compared to the more general broadsheet language of 'defeat'.

<u>5.1.2.6 - Overall</u>

These findings have shown that the newspapers differ in terms of how they portray unsuccessful and successful tennis matches. While there are some differences that seem to be influenced by whether the newspaper is a local or national publication, the major language differences occur between the broadsheet and tabloid media.

5.2 - Limitations

There are two possible limitations in this dissertation. The first of which, concerns the relatively small number of files that reported on matches lost by Andy Murray. It had been decided that this study should only investigate matches which took place at the Wimbledon Championships in London, as this would guarantee that the Andy Murray matches were well reported on by the UK media. However, it also meant that a maximum of only one loss file per paper could be added to the corpus from each year, and in the case of 2013, when Murray won Wimbledon, no loss file could be added at all. With Murray progressing to the latter stages of the tournament in most years, the files reporting his wins greatly outnumbered his loss reports. In fact, out of the 182 files used for corpus analysis, only 30 of them were on matches where Andy Murray lost. However, given Murray's similar rate

of success at most other tournaments, this disproportionate scale is simply the nature of a losses versus wins investigation and so difficult to avoid.

A second limitation was the slightly narrow scope of sources analysed. Given the number of different newspaper that are sold in the UK, and the variety of titles that make up the local, national, tabloid and broadsheet markets, this study may have benefited from analysing a wider representation of the UK press. *The Guardian* for example, is just one of several national broadsheets, which each have their own target audience, readership and political leaning, while the tabloid market is even larger. However, the four newspaper analysed in the investigation were each selected on the basis that together, they covered the market in the most representative manner for this size of project.

5.3 - Further research

The UK's daily newspapers are just one way in which the UK public learn of their sporting news. The biggest rivals to the online editions of these newspapers come from the broadcast media's online news websites like the BBC News website. Thus, a possible area of further research would be to create a corpus of online match reports published by national broadcasters like the BBC, SKY and ITN, and compare them with local Scottish broadcasters like STV. That research along with the results of this dissertation would together cover a significantly larger proportion of the written sports news media.

A second area of further research could look into how Murray is portrayed within the spoken TV news media. In this case the corpora would be made up of spoken dialogue taken from the news reports of the various UK broadcasters TV. For example, the BBC and ITV both produce separate local and national daily news programs, and the language of these shows could be compared. Furthermore, with the BBC nationally owned, publically funded and with a remit of impartiality, corpus linguistic analysis in areas such as local and national sports coverage could be used to statistically calculate whether such remits are being followed.

5.4 - Conclusion

This dissertation has shown that while people may have beliefs about language, these intuitions are not always reliable. In the case of the perceptions regarding the language of the UK media and Andy Murray's national identity, this unreliability has been shown to be apparent based on a significant amount of empirical evidence. Thus the main impact of this

corpus investigation is that it has helped reconsider a belief that does not seem to have any basis as far as language is used in the UK newspapers.

References

Bairner, A. (2015). Assessing the sociology of sport: On national identity and nationalism. *International Review for the Sociology of Sport, 50*(4-5), 375-379.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum International Publishing Group.

Bale, J. (1986). Sport and national identity: A geographical view. *The British Journal of Sports History, 3*(1), 18-41.

Biber, D. and Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.

Billig, M. (1995). *Banal nationalism*. London: Sage.

Bloyce, D., Liston, K., Platts, C. and Smith, A. (2010). Pride of the lions: A sociological analysis of media coverage of the 2005 tour from the perspective of the four home nations. *Sport in Society, 13*(3), 448-469.

Bowker, L. and Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora.* London: Routledge.

Brigham Young University, (n.d.). Retrieved from http://www.corpus.byu.edu/coca.

Clarke, D. (2013). Retrieved from http://www.irishtimes.com/blogs/screenwriter/2013/07/07/when-murray-wins-hes-british-when-he-loses-hes-scottish.

Conboy, M. (2007). A tale of two battles: History in the popular press. *Media History, 13*(2-3), 257-271.

Conrad, S. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics, 22*, 75-79.

Delin, J. (2000). *The language of everyday life*. London: Sage.

Harris, P. (2012). Retrieved from http://www.dailymail.co.uk/news/article-2183997/Andy-Murray-wins-Wimbledon-Schoolboys-embrace-Olympic-champion-captures-buoyant-mood.html.

Higgins, M. (2004). The articulation of nation and politics in the Scottish press. *Journal of Language and Politics, 3*(3), 463-483.

Kiely, R., McCrone, D. and Bechhofer, F. (2006). Reading between the lines: National identity and attitudes to the media in Scotland. *Nations and Nationalism, 12*(3), 473-492.

Koester, A. (2010) Building small specialised corpora. In A. O'Keefe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). London. Routledge.

MacInnes, J., Rosie, M., Petersoo, P., Condor, S. and Kennedy, J. (2007). Where is the British national press? *The British Journal of Sociology, 58*(2), 185-206.

MacRury, I. and Poynter, G. (2010). 'Team GB' and London 2012: The paradox of national and global identities. *The International Journal of the History of Sport, 27*(16-18), 2958-2975.

Mail Online, (2012). Retrieved from http://www.dailymail.co.uk/news/article-393095/Andy-Murray-faces-backlash-England-comments.html.

McCarthy, M. and O'Keefe, A. (2010). What corpora are available? In A. O'Keefe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3-13). London. Routledge.

Meech, P. and Kilborn, R. (1992). Media and identity in a stateless nation: The case of Scotland. *Media, Culture & Society, 14*(2), 245-259.

Mesure, S. (2012). Retrieved from http://www.independent.co.uk/sport/tennis/my-regrets-over-england-joke-by-andy-murray-8142475.html.

National Readership Survey, (2015). Retrieved from http://www.nrs.co.uk/latest-results/nrs-padd-results/newspapers-nrspaddresults.

Petersoo, P. (2007). What does 'we' mean? National deixis in the media. *Journal of Language and Politics, 6*(3), 419-436.

Rosie, M., MacInnes, J., Petersoo, P., Condor, S. and Kennedy, J. (2004). Nation speaking unto nation? Newspapers and national identity in the devolved UK. *The Sociological Review, 52*(4), 437-458.

Scott, M. (2010). What can corpus software do? In A. O'Keefe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 136-151). London. Routledge.

Scott, M. (2012). *WordSmith Tools* (Version 6) [Computer Software]. Liverpool: Lexical Analysis Software. Retrieved from http://www.lexically.net/downloads/version6/HTML/index.html?getting_started.htm.

Scott, M. and Tribble, C. (2006). *Textual Patterns; Key words and corpus analysis in language education.* Amsterdam: John Benjamins.

Sinclair, J. (2005). Corpus and text: Basic principles. In M. Wayne (Ed.), *Developing Linguistic Corpora: A guide to good practice* (pp. 1-16). Oxford: Oxbow Books.

Sparks, C. (2000). Introduction: The panic over tabloid news. *Tabloid Tales: Global Debates Over Media Standards,* 1-40.

Thomas, R. J. and Antony, M. G. (2015). Competing constructions of British national identity: British newspaper comment on the 2012 Olympics opening ceremony. *Media, Culture & Society,* 0163443715574671.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.

Uribe, R. and Gunter, B. (2004). Research note: The tabloidization of British tabloids. *European Journal of Communication, 19*(3), 387-402.

Van Dijk, T. A. (1998). Opinions and Ideologies in the Press. In Bell, A., & Garrett, P. (Eds.). *Approaches to media discourse*. Wiley-Blackwell.

Viner, B. (2010). Retrieved from http://www.independent.co.uk/voices/columnists/brian-viner/brian-viner-why-murray-is-a-brit-when-he-wins-and-a-scot-when-he-loses-2016381.html

Appendix

Appendix 1

Example of the text file layout used to create each corpus

<header>
<paper>guardian</paper>
<title>Andy Murray beats Mikhail Youzhny to reach Wimbledon quarter-finals</title>
<author>Kevin Mitchell</author>
<update>Monday 1 July 2013</update>
<source>http://www.theguardian.com/sport/2013/jul/01/andy-murray-mikhail-youzhny-wimbledon</source>
</header>

<hl>Andy Murray beats Mikhail Youzhny to reach Wimbledon quarter-finals.</hl>

<bp>British No1 overcomes minor crisis to win 6-4, 7-6, 6-1.

Scot reaches Wimbledon's last eight for sixth year running.</bp>

<c>Andy Murray waves to spectators after defeating Mikhail Youzhny of Russia at Wimbledon.</c>

<l>Some people have mid-life crises. Andy Murray has his in the middle of tennis matches, some times at Wimbledon, as in the middle of the fourth round on Centre Court on Monday against Mikhail Youzhny, before sailing by the Russian in three sets at the end without a care in the world.</l>

He won 6-4, 7-6, 6-1, a result not quite as straightforward as the numbers would suggest. The second set took nearly an hour and a quarter of the 157 minutes in total and Murray was within an ace or two – he served 15 in the match, four in that set – of dropping a set for the first time in this tournament. His injured back at one point looked to trouble him but he came through that and it was Youzhny, who needed attention to his right shoulder before the start of the third, who was suffering more in the end.

Murray plays Fernando Verdasco in the quarter-finals on Wednesday and, although he has his number with eight wins in nine encounters, there is one issue he has to attend to first. "I haven't played a leftie all year, believe it or not," he said, referring to the left-handed Spaniard, "so I have to do some work taking leftie serves before then."

His overriding emotion after seeing off Youzhny, the grass-court specialist from Moscow and seeded 20th here, was relief. "Today there were some tight moments but I felt a little calmer before the match, even though Serena [Williams] had just lost on there. Serena, Rafa [Nadal], Roger [Federer] all have lost here – and they're better players than me – so, if they can lose, so can I. All I can do is just concentrate on the next match."

Sometimes it is as if the Scot invites grief (which is an absurd notion) to test himself. But he is through to the quarter-finals, for the ninth time running in grand slam tournaments, although his expectations exceed that admirable landmark. The US Open champion is reasonably well placed to add Wimbledon silverware to his collection, with either of the Poles Lukasz Kubot or Jerzy Janowicz waiting in the semi-finals. He would have taken that sequence at the start of the fortnight.

Murray had a quirky start, even by his standards: three aces to get through four deuce points, then a quick break to go 2-1 up inside 20 minutes.

The net quickly became his friend, as he forced the Russian lower in the shot, unable to get the ball back consistently from deep. Having failed to press his early chances, Youzhny, who lost their two previous encounters, went into "opponent mode", as they say in the fight game. He was hanging on at 2-3 after 25 minutes.

Murray and Youzhny were the two highest seeds left in the bottom of the draw, at No2 and 20, but the Scot was more than 18 places clear in quality. The 2012 finalist was reaching for his sixth straight quarter-final at Wimbledon and it was only the second time he had got this far without dropping a set.

Youzhny got to the quarter-finals for the first time last year, when he lost to Federer, and here was coming off a decent run of form on grass (he has won 53 times on the surface), taking a set off Federer in the final in Halle two weeks ago.

He held to trail 3-4 after 35 minutes. A quarter of an hour later he pushed a tired forehand into the tramlines and the first set was Murray's, 6-4.

Youzhny, though, is an old-fashioned classicist, with a single-handed backhand and a single-minded attitude, so he kept playing his shots and waited for the breaks. Murray, meanwhile, was still looking for his best game. Some of his groundstrokes were on the money, a few were not and he had to rely on his solid serve to get him through an ordinary patch in the second set, as the level dipped on both sides of the net.

In the third game two shots might have changed the tone of the contest: a half-volley drop shot of exquisite touch that Youzhny could only admire from a distance, and Murray's cross-court forehand, skimming the net, to break at the same point as in the first set.

The crowd sensed inevitability. Murray would not have allowed such complacency to enter his thoughts. He knew only Federer, Hewitt and Tommy Haas in this draw had been to Wimbledon more often – respect justified when the Russian broke back to level at 2-2. There was a good bit of tennis left in this match after an hour and a quarter.

And then a twist nobody saw coming – except, perhaps, Youzhny: Murray double-faulted to drop serve for the second time in 15 minutes and go 4-2 down. He needed quickly to refocus and it was plainly a struggle for him, as Youzhny continued to play his waiting game. So, after 10 near flawless sets in this tournament, he was serving at 3-5 to stay in the set.

Forehands long and netted, with an over-cooked backhand to finish, had Youzhny rubbing his right shoulder and Murray took heart to hold, then break back with a sublime backhand. The crowd – perhaps drained after the shock of Serena Williams's exit against Sabine Lisicki on the same grass – finally came to life, urged on by an impatient Murray.

He forced the tie-break and finished it with probably the backhand of the match. The third was a celebration set, as Youzhny's resistance collapsed.

Appendix 2

Unsuccessful versus successful match reports in the national broadsheet (*The Guardian*)

	Key word	Freq.	%	Texts	RC.Freq.	RC. %	Keyness	P
1	DIMITROV	20	0.27	1	4	0.01	47.39	0
2	NADAL	40	0.53	4	43	0.13	37.49	0
3	FEDERER	26	0.35	6	37	0.11	17.37	3.08E-05
4	APPLAUSE	5	0.07	2	0		16.84	4.07E-05
5	SIMPLY	5	0.07	3	0		16.84	4.07E-05
6	NALBANDIAN	9	0.12	1	4	0.01	15.9	6.67E-05
7	RODDICK	17	0.23	2	19	0.06	15.27	9.34E-05
8	DEFEAT	10	0.13	7	6	0.02	14.97	0.000109
9	DISAPPOINTMENT	4	0.05	2	0		13.47	0.000243
10	AGE	4	0.05	2	0		13.47	0.000243
11	CROWD	15	0.2	7	18	0.05	12.44	0.000421
12	WANTED	6	0.08	5	2		12.03	0.000524
13	COULD	25	0.33	7	45	0.14	11.44	0.000719
14	ANOTHER	12	0.16	7	14	0.04	10.27	0.00135
15	CRACKS	3	0.04	3	0		10.1	0.001482
16	NADAL'S	3	0.04	2	0		10.1	0.001482
17	YESTERDAY'S	3	0.04	3	0		10.1	0.001482
18	TEARS	3	0.04	1	0		10.1	0.001482
19	HE	188	2.51	8	641	1.95	8.95	0.002771
20	RARELY	4	0.05	4	1		8.88	0.00289
21	FEDERER'S	4	0.05	1	1		8.88	0.00289
22	AGGRESSION	4	0.05	3	1		8.88	0.00289
23	FINAL	25	0.33	7	52	0.16	8.48	0.003581
24	6-3..	4	0.05	3	65	0.2	-9.65	0.001897
25	ROUND	3	0.04	2	58	0.18	-10.02	0.001545
26	BEFORE	2	0.03	1	75	0.23	-19.03	1.29E-05
27	I	15	0.2	4	193	0.59	-22.13	2.54E-06

Appendix 3

Unsuccessful versus successful match reports in the national tabloid

(*The Mirror*)

	Key word	Freq.	%	Texts	RC.Freq.	RC. %	Keyness	P
1	RODDICK	19	0.54	2	8	0.04	45.51	0
2	FEDERER	20	0.56	3	18	0.08	32.66	8.1E-09
3	NADAL	22	0.62	3	31	0.14	25.08	5.48E-07
4	SWISS	6	0.17	1	1		19	1.67E-05
5	RAFAEL	9	0.25	3	7	0.03	16.09	6.05E-05
6	PERCENTAGE	4	0.11	2	0		15.99	6.38E-05
7	ENDED	6	0.17	3	2		15.57	7.96E-05
8	DREAM	7	0.2	4	4	0.02	14.73	0.000124
9	BIDDING	3	0.08	3	0		11.99	0.000535
10	STUNNED	3	0.08	3	0		11.99	0.000535
11	FIRST-SERVE	3	0.08	2	0		11.99	0.000535
12	VOLLEY	6	0.17	3	4	0.02	11.69	0.000629
13	MURRAY'S	11	0.31	5	17	0.08	11.4	0.000733
14	HOPES	4	0.11	2	1		11.27	0.000786
15	THAT	7	0.2	3	135	0.6	-11.58	0.000665

Appendix 4

Unsuccessful versus successful match reports in the local broadsheet (*The Scotsman*)

	Key word	Freq.	%	Texts	RC.Freq.	RC. %	Keyness	P
1	NADAL	38	0.58	4	22	0.07	66.26	0
2	DIMITROV	12	0.18	1	1		36.61	0
3	RODDICK	21	0.32	2	21	0.06	25.11	5.37E-07
4	ARGENTINE	6	0.09	1	0		22	0
5	AMAZING	6	0.09	3	1		16.26	5.51E-05
6	FEDERER	18	0.27	4	24	0.07	16.22	5.64E-05
7	NADAL'S	4	0.06	1	0		14.43	0.000146
8	NALBANDIAN	6	0.09	1	2		13.37	0.000256
9	TIEBREAK	10	0.15	4	9	0.03	13.03	0.000307
10	NO3	3	0.05	2	0		10.82	0.001003
11	TEARS	3	0.05	2	0		10.82	0.001003
12	TOP	15	0.23	5	24	0.07	10.79	0.001023
13	FANS	5	0.08	2	3		8.53	0.003491
14	PLAYED	19	0.29	6	42	0.13	7.99	0.004701
15	THEY	2	0.03	1	50	0.15	-8.27	0.004035
16	ROUND	6	0.09	3	88	0.26	-8.71	0.003161
17	FEW	1	0.02	1	44	0.13	-9.86	0.001685
18	I	31	0.47	3	296	0.89	-13.44	0.000247

Appendix 5

Unsuccessful versus successful match reports in the local tabloid

(*The Daily Record*)

	Key word	Freq.	%	Texts	RC.Freq.	RC. %	Keyness	P
1	NADAL	44	0.77	3	23	0.09	75.86	0
2	RODDICK	17	0.3	1	3	0.01	43.38	0
3	DIMITROV	14	0.24	1	2		37.4	0
4	FEDERER	25	0.43	4	25	0	27	1.71E-07
5	DISAPPOINTMENT	6	0.1	4	0		20.86	4.94E-06
6	FINAL	23	0.4	5	35	0.13	15.61	7.76E-05
7	UNABLE	4	0.07	4	0		13.91	0.000192
8	SUFFERED	5	0.09	2	1		12.36	0.000438
9	NADAL'S	6	0.1	2	3	0.01	10.56	0.001154
10	SWISS	6	0.1	1	3	0.01	10.56	0.001154
11	WOEFULLY	3	0.05	1	0		10.43	0.001241
12	STRENGTH	3	0.05	1	0		10.43	0.001241
13	BULGARIAN	3	0.05	1	0		10.43	0.001241
14	CRUSHING	3	0.05	2	0		10.43	0.001241
15	HEARTBREAK	3	0.05	1	0		10.43	0.001241
16	ACES	8	0.14	3	7	0.03	9.79	0.001752
17	ANYTHING	5	0.09	3	2		9.78	0.001764
18	SHORT	5	0.09	3	3	0.01	7.96	0.00479
19	RAFA	5	0.09	3	3	0.01	7.96	0.00479
20	FASHION	5	0.09	3	3	0.01	7.96	0.00479
21	SEVEN	5	0.09	3	3	0.01	7.96	0.00479
22	COURT	13	0.23	4	128	0.48	-8	0.004686
23	6-3..	2	0.03	2	46	0.17	-8.13	0.004353
24	TSONGA	1	0.02	1	36	0.13	-8.22	0.004151
25	THROUGH	1	0.02	1	54	0.2	-14.39	0.000149