Davis, Richard Hill (2015) A genre analysis of medical research articles. PhD thesis.

A Genre Analysis of Medical Research Articles

Richard Hill Davis

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in English Language

English Language
School of Critical Studies
University of Glasgow
April 2015

Abstract

Hospitals and other health institutions around the world have begun to tie staff promotion and careers to publication; accordingly, an increasing number of medical journal articles are being written by non-native English speakers and novice writers. This work aims to analyse medical journal articles as a genre, and follows Swales' (1990) framework for doing so, by interviewing a sample of the discourse community and finding the Rhetorical Moves that make up the genre, with additional investigation of stance, via selected reporting verbs, and cohesion, through selected discourse markers. I compiled one of the larger corpora of medical research articles (250), as well as one of the most recent (2001-2011). Previous studies reviewed 50 articles at most, drawn from earlier periods of time. As part of the examination of the genre, this study includes discussions with a sample of the discourse community, the users of the genre, with interviews from ten doctors and five editors from around the world who have a wide range of experience in writing, publishing and editing articles. In addition, I identified 17 Rhetorical Moves, with four considered optional, with the idea to identify a sequence that writers and educators can use to see how the medical article may be written. I also examined 13 reporting verbs to determine if it is possible to identify authorial stance regarding the information being reported, and were coded as being factive (the authors agreed with the information), non-factive (the authors conveyed no judgement on the information) and counter-factive (the authors disagreed with the information being reported). Finally, the study looked at how cohesion is maintained through examples of the five types of discourse markers. This study presents the most comprehensive examination of the genre to date, which, through the utilization of corpus analysis techniques, allows a more in-depth analysis than previous studies.

Table of Contents

List of tables

List of abbreviations

AMWA: American Medical Writers Association

BMC: Biomed Central

BNC: British National Corpus

CDM: Contrastive Discourse Marker

COCA: Corpus of Contemporary American English

DC: Discourse Community

DM: Discourse Marker

EDM: Elaborative Discourse Markers

EFL: English as a Foreign Language

ESP: English for Specific Purposes

GA: Genre Analysis

ICMJE: International Council of Medical Journal Editors

IDM: Inferential Discourse Marker

IMRD: Sections of an article, i.e. Introduction, Methods, Results and Discussion

KWIC: Key word in context

LSIOU: List of Serials Indexed for Online Users

LP: Lexical Priming

MDM: Meta-Discourse Marker

MRA: Medical Research Article

NIH: National Institute of Health (US)

NNS: Non-Native Speaker

NNW: Non-Native Writer

PLoS: Public Library of Science

PM: Pragmatic Marker

RA: Research Article

RRA: Research Review Article

SCI: Scientific Citation Index

SP: Semantic Prosody

TDM: Temporal Discourse Marker

UKWaC: UK Web Archiving Consortium

Acknowledgements:

I first wish to state my heartfelt appreciation and thanks to my supervisors, Dr. Wendy Anderson and Dr. Catherine Emmott: you have been amazing. I thank you for encouraging my research and for allowing me to grow as a scholar. Your advice on my writing and organization was always appreciated, though at first I may not have shown it. From your teaching me Harvard style and humility, I have grown not only as a writer, but as a person as well. As I have had to move around the world, it has been incredibly hard working with me, but your tireless efforts have been greatly appreciated. You were always patient and never cross, regardless of how many times you repeated yourselves or how often I forgot to do the checklist. If I can be just ten per cent the supervisor to my students as you were to me, then they will be very lucky, as I was truly blessed.

I would also like to thank the examiners: Professor Sheena Gardner, Professor Nicki Hedge, and the chair Professor Jane Stuart-Smith, for their help in this undertaking. Your hard questions and astute observations helped me immensely.

I thank the doctors and editors who were able to help me by giving up their time and sharing their expertise. They went above and beyond in agreeing to talk to a linguist about a topic that was outside of their comfort zone and for some, in a language that is not their native one.

I would especially like to thank Dr. Sara Campbell, Dr. Michael McEwan, Dr. Ashlie Darr, and Dr. Ryan Caldwell for answering my many questions. All of you supported me when I did not understand something medical or could not figure out how to do the stats while working on my Ph.D. thesis.

Bruce Fraser: You answered a strange student's desperate emails and helped him understand even more.

John Corbett: Thank you for starting me down the road and planting seeds in my subconscious.

Chapter 1: Introduction

1.1 Overview

This chapter will introduce the impetus behind undertaking this study, along with a description of genre analysis (GA) and corpus analysis, the two main frameworks used in the study. This will be followed by an outline of the four research questions this work will attempt to answer.

1.1.1 Rationale for the study

English is the global lingua franca for medical publication, which means that while medical researchers around the world publish in their own languages, for greater prestige, they must also publish in English, even if it is not their native or preferred language. Many hospitals and medical institutions around the world have begun to require publications for promotion in SCI (Science Citation Index) journals, the vast majority of which are written in English. Publication is more difficult for those who are not proficient at academic or scientific writing in English. In addition, many doctors now work in a 'publish or perish' environment; that is, unless academics or experts conduct research and publish in their field, they will not be hired or promoted and will therefore perish professionally. The ultimate aim of this study is an applied one: it is designed to help those whose native language is not English, as well as novice writers who may not be familiar with the expectations of the genre of medical research articles (MRAs). My primary objective is to apply Swales' (1990) framework of genre analysis (GA), with the assistance of corpus analysis, to identify certain linguistic and pragmatic language practices that may not be immediately apparent.

According to a number of doctors in China (e.g. Yuan, Xu & Hu 2013; Yu et al. 2013), there is immense pressure for young doctors there to publish in order to secure a position, but there is little internal support in the ways and means of publishing.

This pressure is not restricted to China alone, but rather occurs in many nations; in my own experience in Taiwan, doctors tell similar stories of institutional pressures.

Six years ago I was an EFL teacher in Taiwan. My wife and I owned an English 'cram school', which focused on students for the first eight grades. A parent of one of my students, a pharmaceutical representative, came to me one day and asked me to help his client, a doctor, with a paper she was working on. I tentatively agreed: I was not sure I could be of help, since I was not a medical professional. She had already submitted the article to a journal but it had been rejected for poor organization and issues with English usage. I read the rejection letter and looked at her paper: the science was sound, but her ideas were obstructed by poor grammar and the reviewers added that the article was not 'organized correctly'. While I could fix the English grammar problems, I was at a loss as to how to better organize the paper. When I read other articles that were published in the journal to which she had submitted, I noticed that her paper lacked some features that were common in other papers. A cursory Internet search for 'medical article templates' turned up surprisingly little to help novice or non-native speakers (NNS) with what belongs in a medical article. What did turn up, however, were discussions of research-based writing in fields such as business, explorations of different features of second-language writing, and discussions of different parts of research articles, such as abstracts or introductions. My growing interest in the characteristics of medical research articles led me to work on genre and GA: in order to publish her article, the young doctor needed to know more than proofreading tips. She needed to understand something about how a medical research article was organized and what its readers would expect to find. Beginning to identify her needs led me to the seminal work on GA by Swales (1990), which led in turn to my decision to attempt a genre analysis of medical research articles, as this chapter explains.

1.1.2   Swales' impact on other scholars

GA is an approach, with associated techniques, by which one analyses a text, based on parameters that help define the text as belonging to a certain genre. Techniques for GA have been used in literary criticism since Homer wrote the Odyssey (Todorov

1976, p.159; Corbett 2006). Early work in the language learning community included studies by Maher (1986) and Bhatia (1987), each of whom focused on genre in specific professional writing areas; however, it was not until Swales published his 1990 study, *Genre Analysis,* that this approach was applied to multiple areas of discourse and became more fully embraced.

Research into genre is an important component of materials development for language learners and language users, in that recognizing and understanding a genre helps solidify learner expectations about the context within which a text is written or spoken. By using GA, texts may be broken down into manageable chunks, which allow the learner to predict elements needed to fulfil the requirements of membership in a certain genre. Although this discussion will focus primarily on written text, the same techniques are used for the analysis of oral discourse. By using a genre-based approach, researchers can identify the parameters constraining the text and can see how authors comply with as well as break from those parameters.

In their reviews of Swales' work on GA, critics have noted how important the ideas incorporated in GA are for the analysis of text (e.g. Marius 1990, p.458; Allison 2006, p.244). Even though Marius (1990), in reviewing Swales' 1990 study, claims that the ideas are not clearly presented, he does acknowledge its importance. While I follow Allison (2006) in finding that the framework as explained in Swales' later work, *Research Genres* (2004), better fits specific examples of academic genres, I continue to assume Swales' initial definition of a genre: 'A genre is a class of communicative events, the members of which share some set of communicative purposes which are recognized by the expert members of the parent discourse community' (1990, p.58). This study will break down the analysis of a genre first by looking at its events and purposes and, next, by interacting with its discourse community.

In their work on genre, corpus linguists Biber and Conrad (2009) outline the differences between genre, register, and style. They assert that GA is concerned with complete texts containing specialised expressions and formatting based on rhetorical organization; that linguistic characteristics appear in text-specific locations (that is, introductory expressions occur naturally in the Introduction section and are not likely to happen in other sections); and that features are based on genre conventions but are

not often functional (2009, p.16). This can be summed up by saying that for Biber and Conrad, register refers to language used in a particular situation, and style is the chosen presentation of language, but that any features that are expressed because of expectations for the type of text is what makes something a genre. As Frow (2013, p.24) comments from his reading of Freadman (1988):

> A sermon is defined by its relation to other sermons, even though the kind of thing this is changes over time, but also by the fact that it is *not* a prayer or a eulogy or a political speech.

Biber and Conrad further argue that genre and register are both controlled by context; they develop a framework of defining situational characteristics of genre and register (2009, p.40). The following table will show their categories and my proposed mapping onto what I will argue is the genre of medical research articles (MRA).

Table 1.1: Mapping of MRAs onto Biber and Conrad's 2009 framework

Biber and Conrad's terms	MRAs
Participants (addressor and addressee)	The authors are experts in a medical field. The audience will be mostly other experts in the field.
Relations among participants	There is shared knowledge between the users and there is some interactivity in the form of letters to the editor, review articles, and any text in which the author agrees with or disagrees with what was reported in the MRA.
Channel	The mode is written, within a medium of submission and peer review.
Production circumstances	Articles are revised and edited.
Setting	The time is contemporary and usually placed within the realm of medicine.
Communicative purpose	The purpose for publishing an MRA is to highlight a specific medical problem or issue in the field and present a solution in a factual manner.
Topic	A specific problem exists that needs a solution.

While the seven characteristics are well reasoned, according to my proposed mapping, it would be hard to differentiate MRAs as a genre from that of a newspaper, for example. With the exception of participants, setting, and communicative purpose, everything else I suggested could be used to describe a newspaper article. Biber and Conrad (2009, p.53) further add *markers* of genre, register, and style, and it is in these that the difference between genres can be better realized. In my chapter on Moves (Chapter 4), I illustrate how almost every MRA begins with a statement of fact: this is

an example of a genre marker. Biber and Conrad describe genre markers as formulaic and as occurring only occasionally (2009, p.54). A genre marker for the opening of a newspaper article, according to Associated Press standards, is that the lead, or opening paragraph, contains answers to the five Wh-questions (*who, what, when, where, why*), whereas in an MRA there may only be one of the question-words addressed. Genre differentiation can be linked to small and subtle issues of usage. Table 1.2 will describe the difference between three types of genre: the generic research article, the genre of MRAs, and the genre of the newspaper article, using Biber and Conrad's framework.

Table 1.2: Comparison of three different genres applying Biber and Conrad's framework

Biber and Conrad	Research article	MRA	Newspaper
Participants A. Addresser B. Addressee	A. Experts in the area B. Experts	A. Medical experts B. Experts	A. Reporters B. Anyone
Relations among participants	Relative status like that of colleagues/peers	Relative status like that of colleagues	Strangers
Channel	Written, formal	Written, formal	Written, formal
Production circumstances	Revised and edited	Revised and edited	Revised and edited, but may also be in real time via social networks
Setting: generic topic	Area-specific	Medical	Any topic
Communicative purpose	To report on research in a particular field	To report on research in the medical field	To report on something
Topic	Specific to the area	Medical issue	An event that has occurred

The differences between the genres may be slight but, nonetheless, help to shape them. In this investigation, I follow others in the belief that research articles form a genre (Swales 1990, 2004; Biber & Conrad 2009), but with MRAs being their own genre within this broader genre. I will establish the framework of the MRA as a genre, based on ideas of Yates and Orlikowski (1992, p.303) of how there can be different purposes with different forms within a genre.

The framework for analyzing a genre as noted by Swales (1990, p.1) can be established through a three-stage process. The first stage involves the discourse community. That community is the group of people who interact with or use the genre, for example, authors and other readers. Any discussion of the analysis of the genre should solicit their commentary, whether by citation analysis, by opinion survey

or by some other means. This led to my seeking to interview members of the MRA discourse community: this is discussed in more detail in Chapter 3.

The second stage in Swales' process of defining a genre is that of identifying conventional expectations that have grown up around the genre. For example, broadly speaking, a novel tells a fabricated story while a newspaper article meets different criteria, including that of reporting accounts of current events. With regards to MRAs, this parameter is seen as being the identification, contextualization and reporting of a medical problem or hypothesis, the methods by which it was approached, the results, and its eventual solution. Thirdly, Swales (1990, p.58) states that a language-learning task is the third necessary component in GA. This third stage suggests possible pedagogical uses in a language classroom; however, when I write of the language learning classroom, I am not only referring to learning other languages but also to the language learning classroom in one's own (mother) tongue. Students need to know how to use a language within the framework of a situation as defined by its genre, and therefore this information is crucial whether it is in a primary or second language setting. It is Swales' third and highly salient point that grounds much of the present research into MRAs. By using GA as the lens for the study of the production and analysis of text, one can help language users to better understand what is expected of them when writing in a particular genre.

Other scholars define genre in different ways or emphasize different aspects of Swales' model. Biber (2010, p.241), stresses that a genre is defined by its major Moves, by which he means any change in focus in an area of text. According to him, these Moves are usually found only once in a text, and can be identified in an MRA by words like *Introduction*, *Methods*, and similar section headings. He further argues that the focus in identifying such patterning is limited to word level, and does not take into account the larger context of the article.

This assertion by Biber does not always fit research review articles (RRAs), which are articles in which the author reviews the work of other researchers on the given topic. One of the problems I faced when setting out to define MRAs was the fact that RRAs are very similar to MRAs at the word level. At the most basic level, review articles in medical fields share the headings or key words and much of the same technical jargon

and terminology present in MRAs. Accordingly, there must be a difference beyond similar headings and common vocabulary to distinguish MRAs from RRAs.

Michael Handford's (2010a, p.257) model of specialist genres, which builds on Swales' ideas of how a genre is formed (1990, p.46), may be a better fit for identifying MRAs as a genre. He claims that a genre is to a great extent defined by the specialists who use the genre, and its participants, both authors and readers, imbue it with certain criteria or expectations. With regards to MRAs, what sets them apart as a genre different from that of the broader research article is the fact that primarily medical professionals use them. It is unlikely that engineers, who would read research articles in their own field, would read an article in a medical journal to help with an engineering issue and vice versa. The major point here is that the specialist reader is part of the definition of what makes the genre. Handford defines these specialists as being members of the discourse community (DC). As described by Swales, discourse communities 'are sociorhetorical networks that form in order to work towards sets of common goals . . . and possess familiarity with the particular genre' (1990, p.9) (see Chapter 3 for a more in-depth description). In this study, I follow Swales and Handford in assuming that members of the DC create these networks and goals. An important difference between the two scholars is that Swales' definition is broader in allowing for the creation of genres, while Handford's work, built upon Swales' earlier findings, admits many specific genres, based on the specialists who interact with them, thereby helping form genres like MRAs.

GA is an established technique within the English for specific purposes (ESP) community, among scholars such as Boden (1994), Brett (1994), Bargiela-Chiappini and Harris (1997), Holmes (1997), Badger (2003), Poncini (2004), Swales (2004), Ding (2007), Handford (2010b), Koester (2010), and Milagros del Saz Rubio (2011). All of these scholars look at different genres, ranging from legal papers (Badger 2003) to personal statements in medical student applications (Ding 2007). A quick review of the issues since the inception (1995) of the journal *English for Specific Purposes* will turn up hundreds of articles with a focus on genre. However, even with the multiplicity of works within the field of ESP, there are some critics of GA and the issues they raise will be described in the next section.

1.1.3 Critics of genre analysis

While GA is a widely utilized method, it is associated with some problems and possible pitfalls. Handford outlines some of the criticism of GA (2010a, p.257). A common complaint is that the term is too often used, leading to a watering-down effect, wherein the term seems to be coloured by similar terms such as *register, style* or even *text.* Biber (1994, p.31) stresses that register is tied to the situation, which sounds very close to the idea of genre (as discussed in 1.1.2). While vocabulary is an essential part of the defining of a genre (for a more extended discussion of Swales' fifth criterion for a DC, see 3.1.3), it is only one part of the total perspective.

With regard to the use of GA in the field of language teaching, Bazerman (2000, p.14) warns that it is important to teach not only the expectations for the language of an article but its context as well. Students may then have a better grasp of what is necessary to produce a work in line with the conventions of the genre. Coutinho and Miranda (2009, p.36) assert that it is inherently difficult to be descriptive with genres and that any attempt to do so needs to be done within a framework. For that framework, Bawarshi and Rieff recommend using Bronckart's theory of socio-discursive interactionism (2003). Published originally in Portuguese, this is a theory that 'postulates that human actions should be treated in their social and discursive dimensions, considering language as the main characteristic of human social activity' (Bawarshi & Rieff 2010, p.75), even though they do note that this framework is for individual texts and not genres as a whole.

Approaches to GA differ by field. Lynne Flowerdew (2005, p.322) claims that there are three main schools of thought on using the genre approach to discuss writing: (1) English for Specific Purposes, (2) North American New Rhetoric studies, and (3) Australian systemic functional linguistics. Susan Hyon (1996, pp.695-697) offers a clear and concise comparison of the three schools. The following table encapsulates her summation.

Table 1.3: Schools of thought, adapted from Hyon (1996, pp.695-697)

School of Thought	Characteristics
ESP	Treats genre as a communicative event, with focus on purpose, style and audience. Work generally interested in structure and Moves within the text.
New Rhetoric	Focuses more on context than structure and tends to use ethnographic rather than linguistic tools for analysis.
(Australian) systemic-functional	Focuses more on primary and secondary education texts and less on professional work. Applies Halliday's schemes of: Field, what the activity is; Tenor, the relationship between participants; and Mode, the system of delivery.

All three traditions acknowledge the fact that purpose and context are important, but differ in how to analyse the text. My work in this study will more closely approximate that of the ESP school, but there will be influences from all three. I mix two approaches in my analysis of MRAs, the first being GA and the second being corpus analysis as a tool to examine lexis and the possibility of stance. There is, however, some critical opposition to this mixed-methods approach. Handford (2010a, p.255) lists four criticisms of the combination of GA and corpus analysis, the first being that, according to Widdowson (2000), corpus analysis of a genre often presents the text in a decontextualized state. That is to say, when looking through the scope of a corpus analysis tool, one only sees a small piece of the text and may not be able to achieve a full perspective, with the result that some of the meaning is lost. Hunston (2002, p.23) further argues that one of the major drawbacks of corpus analysis is that the spatial context in which the original text is presented does not translate when using corpus analysis tools: she goes on to urge that, when examining language, a corpus analysis tool should be just one part of the analysis. Her criticism is counterbalanced by my creation of a sample of the discourse community, to provide the larger, genre-based context, and by my choice of a corpus tool (WordSmith Tools [Scott 2008]; see 2.2.3) that goes beyond collocation listings to allow examination of the word in the full text.

Handford makes the point that a bottom-up approach is used when employing a corpus tool to analyse a corpus of a single genre (2010a, p.255). That is to say that the focus may be too much on the lexis and therefore the context is not sufficiently considered. An additional point he makes is that by applying corpus analysis to the genre, the data analysis may be too quantitative by being heavily number-based. Treating the data exclusively as numbers might elicit results that are too narrow, leading researchers to make concrete claims based on their data which may in fact not

be a fair representation of the genre. His last criticism of combining the two approaches is that, for some, corpus analysis is focused on compiling the largest dataset, but that by working with a specialised genre, a researcher is able to look at a smaller data set than those who study the language as a whole (2010a, p.258). One of the issues facing corpus linguistics is that of the appropriate size for a corpus: it is not the case that the bigger the corpus, the better it is for the researcher. This line of thinking may lead some researchers to overgeneralize their findings, because they will feel that their corpus is sufficiently large to be representative of the genre, instead of understanding that genres, much like language itself, can vary and change, i.e. vary at any one type, synchronically, and also change over time, diachronically. Since size can be a difficult way to measure the value of a corpus, representation of the genre becomes of paramount importance. As Koester states, it is nearly impossible to build a perfectly representative corpus; however, a researcher should limit the number of variables nonetheless (2010, p.69).

The current research aims to take a snapshot of the genre and then, by applying other techniques such as the analysis of semantic prosody and lexical priming, to take both a bottom-up and then a top-down approach, through interviews and a study of the Rhetorical Moves, to understand the corpus (see 4.3 for a description of these). To understand what makes research articles their own genre, I work with lexis and with the construct of a discourse community, as well as with Rhetorical Moves and grammatical features.

1.2   Building a corpus

The benefit of building and then analysing a corpus is that it can allow researchers to see features of language that may not be otherwise readily apparent (Reppen 2010, p.30). As discussed in this chapter, this work undertakes a genre analysis of medical research articles (MRAs). After describing the methods to be used in compiling a corpus of MRAs (Chapter 2) and eliciting commentary from a sample of the MRA discourse community (Chapter 3), I will examine the Rhetorical Moves of the genre (Chapter 4), and specific traits in the language (Chapters 5 and 6). A corpus offers a highly convenient format through which to support the examination of texts and lexis,

since it will support the creation of standardized samples and allow machine-based textual analysis.

According to the National Library of Medicine, the '2014 edition of the LSIOU [List of Serials Indexed for Online Users] contains 14,772 serial titles, including 5,651 titles currently indexed for MEDLINE' with 5,642 of these indexed in both Index Medicus and MEDLINE. Most of these are in English, although we cannot be sure of the exact total. Between 2010 and 2014, the number of citations to foreign-language articles was 7% as opposed to the 93% of citations to English-language articles[1]. This figure is slightly misleading, as the 2014 census had not been completed at the time of writing. However, looking at the trend in the averages from 1970 is more instructive, when the average was 37%, through the 16% of 1990-1994 and the 9% of 2005-2009. Given the large number of journals, it is not unreasonable to expect them to have a number of common characteristics, and to constitute a genre in and of themselves, which offers preliminary justification for compiling a corpus.

1.2.1 Rationale for developing the medical research article corpus

Corpora have been used in many areas in linguistics from lexicography to stylistics, but one emerging area is in language teaching (Blecha 2012, p.29). The theory of GA was developed by scholars in ESP and holds the benefits of using corpora in language teaching as being, 'to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which addresses the specific needs of students in a particular domain more directly than those taken from more general language corpora' (McEnery & Wilson 2010, p.127).

1.2.2 Size

When designing a corpus, the question of its size is one of the most important features to identify before beginning; unfortunately, however, there is no definitive answer as

---

[1] http://www.nlm.nih.gov/pubs/factsheets/medline.html

to the best number of texts to include. In Stubbs' description of a corpus (2008, p.106), he uses the adjective *large* to describe the amount of text to be collected for a specific purpose; 'large: millions, or even hundreds of millions, of running words, usually sampled from hundreds or thousands of individual texts'. Sinclair (2004a, p.189) claims that larger corpora are better for certain things, for example, 'underlying regularities have a better chance of showing through the superficial variations'. Koester (2010, p.66), however, counters that larger corpora may overload the researcher with too many instances, and lead them to have to use smaller samples, thus negating the advantage size originally presented. She further adds, 'They [smaller corpora] allow a much closer link between the corpus and the contexts in which the texts in the corpus were produced' (2010, p.67). Since the current work emphasizes GA, context is very important.

According to Reppen (2010, p.31, cf Biber 1993, p.243), there are two factors needed to determine the size of the corpus. The first factor is representativeness, i.e. does the researcher have enough examples to make any comments of value? The second factor is practicality, that is, will the researcher have the time needed and storage available to build and analyse the corpus?

Handford asserts that one major benefit of using smaller, specialised corpora in conducting genre studies is that even though they are smaller, they can still be representative of the genre (2010a, p.258). He further adds that a specialised corpus of a million words would be considered a large one (2010a, p.258). Accordingly, when, constructing the corpus for this work, I used a million words as a target.

1.2.3 The use of specialised corpora in genre analysis studies

Specialised corpora, like the MRA corpus compiled for this current study, have been used in several studies such as Gledhill (2000), Marco (2000), Thompson (2000), Bondi (2001), Henry and Roseberry (2001), Upton and Connor (2001), Connor, Precht, and Upton (2002), Upton (2002), Lee and Swales (2006), Bruce (2009), and Chang and Kuo (2011). In these studies the researchers compiled specialised corpora to help analyse different aspects of genre. For example, Thompson (2000) examined

20 PhD theses to see how authors used citations, as either integral or non-integral in the text, while Upton and Connor (2001) built a learner corpus of texts by second-language students, to look at issues around tagging and to identify how politeness could be found in the students' writings. Henry and Roseberry (2001) built a corpus of application letters and introductions to speeches to study collocation of certain words, specific to those two genres. Upton (2002) explored a corpus of fund-raising letters to examine the Moves of the genre. Chang and Kuo (2011) built a corpus of 60 computer science articles to study key words and Moves. Therefore there is ample precedent for using specialised corpora to do work with GA. The following research questions reflect the confluence of corpus analysis with GA as the foundation for this study.

1.2.4   Potential sub-genres

Although this discussion is on medical research articles, they are not the only type of article found in medical writing. From the viewpoint of medical science, articles may be categorized clinically in terms of the levels of evidence presented, allowing the clinician to discriminate between, for example, individual studies and systematic reviews. As Harewood et al (2010) point out, training is needed for medical professionals to identify the rigor of randomized control trials (seen as the most rigorous), cohort studies, and case-control studies. The Oxford Centre for Evidence-Based Medicine maintains an online repository of documents tied to developing and evaluating evidence-based research (http://www.cebm.net/ocebm-levels-of-evidence/). Each of those three types of evidence-based medicine is tied to a problem-solution situation, either interventional or observational (Besen et al 2014), and is empirical in nature. As such, these types are included in the present corpus, but research reviews, which can have slightly different moves, are not (see Section 3.3.2).

As part of their series of articles on evaluation of scientific publication in *Deutsches Ärzteblatt International,* Röhrig, du Prel, and Blettner (2009) have published helpful discussions of study design (2009) and then Röhrig, du Prel, Wachtlin, and Blettner looked at types of study (2009) in medical research. Of particular interest is their division of medical research into primary and secondary, in which secondary research

includes meta-analysis, review, systematic review, and simple narrative (2009b: 263), each of which could present moves and steps that differ from the ones established here for primary research.

It is also possible that interventional and observational studies can show variation in their sub-moves or steps; for example, Step 3 in Move 3, *State the aim* (Section 4.3.1.1.3*),* will be incorporated in both types of studies, but the Step may be moved to Move 1, *Study situation* (4.3.1.1.1) for intervention studies. However, a lengthy and detailed comparison of the two types of studies has yet to be made. Another lacuna in the literature is a detailed analysis of Steps in a comparison of evidence-based studies.

However, these are by no means the only types of medical writing. The American Medical Writers Association provides a toolkit including over twenty kinds of medical writing (presumably for non-medical personnel) ranging from posters to regulatory documents.[2]

1.3   Moves, stance, and markers

The following are the four main research questions that this work will address. This work is primarily interested in the characteristics of MRAs as a genre in their own right, which leads to the first research question:

    i.   How can we define MRAs as a genre?

To better understand what might differentiate the genre of medical research articles from other genres, I elicited a sample from the discourse community (users of the genre) and interviewed the participants to identify the possible traits and trends of the genre. In Chapter 3, I outline how I identified members of the discourse community and the questions I asked them in order to clarify and define the genre. Through their

---

[2] http://www.amwa.org/toolkit_new_med_writers

responses and definitions, I was better able to determine how the structure of Moves and the language and style choices might help fulfil the expectations of the genre.

ii.   What are the Moves in this genre?

To understand how a medical research article is assembled, I followed Swales' (1990) model of identifying and studying the Rhetorical Moves within the genre. A Move is a shift in focus in a text, and, according to Biber (2010), Moves are key to analysing a genre which is why I need to study them (see 1.1.2). I identify the boundaries of each Move and define them in terms of purposes for and possible uses of the Moves, using examples taken from the corpus.

iii.   How are reporting verbs used to show stance, and how can a researcher identify the different ways in which the reporting verbs are being used?

One of the possible problems facing novice and non-native writers is that they may not know what words to choose when reporting their own or others' findings (Bloch 2010; Hyland 2005). Accordingly, I looked at a sample of reporting verbs by using corpus analysis. I then studied the sample using semantic prosody and lexical priming as ways to explain and define how selected reporting verbs are being used within this genre.

iv.   What are the different kinds of discourse markers used to maintain cohesion in MRAs, and how might they be used?

Swales (1990) notes that cohesion and how authors maintain it differ from genre to genre, with one aspect of how cohesion is maintained being the use of discourse markers such as *here,* and *therefore*. Therefore an examination of different types of discourse markers by close analysis of selected examples of each type, drawn from the corpus, is needed to answer this question.

1.4   Outline of thesis

The present work will follow Swales' approach to conducting GA; however, in an attempt to more fully describe particular facets of the genre that can only be explained through lexis, corpus analysis will also be employed. Chapter 2 (Building the medical research article corpus) will describe how the MRA corpus was developed, and this will be followed by Chapter 3 (The discourse community sample) which sets out how I drew on the expertise of members of the discourse community. Chapter 4 (Mapping of the Rhetorical Moves) will attempt to map the Rhetorical Moves of the genre, and this will be followed by Chapters 5 (An examination of thirteen reporting verbs in MRAs) and 6 (Selected discourse markers in medical research articles), each of which will employ corpus analysis to examine specific types of words and phrases that may be utilized by authors to show stance and cohesion, respectively. The ultimate goal for this work is to define the genre and its Moves, with an examination of lexical features that help show authorial stance, and of how features and stances are used to connect the text at sentence and word level.

Chapter 2: Building the medical research article corpus

2.1   Introduction

My initial impetus for developing the MRA corpus was to enable me to better understand the genre and its Moves in order to teach practicing medical personnel, especially novice writers or NNS, to understand the expectations of the genre and to write successfully within it. In this chapter, after highlighting some of the developments in the history of corpus use, I will characterize several well-known types of corpora in order to situate the particular kind of corpus I have developed: a specialised corpus of peer-reviewed medical research articles (MRAs). I will review details in developing this corpus, including those of size and representativeness, and explain from where, how and why I selected texts for analysis.

2.1.1   The medical research article corpus in its historical context

Before the mid-twentieth century, scholars manually collected full texts or selections from chosen texts, typically sacred or literary works, in order to create corpora and concordances, such as Baker's (1914) *A concordance to the poetical and dramatic works of Alfred, Lord Tennyson*. By the late 1960s, researchers were able to develop ways to use computers to more effectively gather and catalogue corpora, as reported by Francis and Kučera (1979) in their corpus manual for their 1961 corpus of contemporary American English, the *Standard Sample of Present Day American English*, or as it is usually called, the Brown Corpus, named for the university supporting its compilation. The Brown Corpus marked a departure from a nearly exclusive focus on compiling corpora of literary or sacred works; instead, it incorporated 500 samples from 15 genres of written texts chosen to provide a representative sample of written English in America in 1961. It was the first publicly available computer-readable corpus, and by its capacity to support machine analysis, it helped introduce technology-based advancements to corpus analysis. For example, with the generation of computerized corpora that were to follow, researchers were

able to tag words and thus more reliably and more quickly build statistical profiles of the language (Meyer 2002, p.99).

During the years following the development of the Brown Corpus, larger corpora have been developed, primarily for lexical analysis. The Brown Corpus was followed by the emergence of mega-corpora, with a significant example being the British National Corpus (BNC) compiled in the early 1990s with over 100 million words. The move to larger corpora continued with the increasing role of the Internet; as an example of large corpora, researchers at the UK Web Archiving Consortium[3] (UKWaC) are compiling a corpus with over a billion words: the UK Web Archive. This corpus is not a collection designed for linguistic analysis as such; instead, it has been created by six UK higher education centers to allow researchers to preserve articles and sites that may in the future disappear. Mark Davies at Brigham Young University has compiled very large corpora that allow the users to focus specifically on linguistic analysis.[4] As of 2014, he had created nine corpora, including the Time Magazine corpus (100 million words), the Corpus of Contemporary American English (450 million words), and the Corpus of Historical American English (400 million words), all of which support the analysis of language use over time, and of word and phrase variation. As more texts have become available to researchers through the Internet, corpora have grown to sizes previously unimaginable and the use of corpora in fields such as lexicology, syntax, phraseology, discourse, language change and language variation is growing.

The MRA corpus that has been compiled for use in the present work is certainly not big by the standards of general corpora today; however, without the Internet and the foundation laid by early developers of corpora, even the relatively small corpus used here would have been impossible. While large corpora are clearly valuable for the study of collocation, for example, small corpora continue to be developed by researchers to analyse specific questions that can go beyond the lexicon. Moore (2003, p.301) states that small corpora, 'have something unique to say on the nature of particular varieties of language', especially with regards to genre-based features.

---

3 http://www.webarchive.org.uk/ukwa/info/about
4 http://corpus.byu.edu

2.1.2 A sample of types of corpora

The nature of corpus analysis is a much-debated topic; according to Flowerdew (2013, p.174), analysts can be broken down into two camps: those who view corpus analysis as an approach or methodology, and those who view corpus analysis as having theoretical status. The current research belongs to the first camp, as it will apply corpus analysis approaches as the means of doing genre analysis.

There are many types of corpora, from mega-billion word sized corpora to smaller corpora designed for specific purposes, and frequently developed to support classroom teaching. The following table, adapted from the 15 types reviewed by Blecha (2012, pp.15-20), shows some of the many different types of corpora and their characteristics.

Table 2.1: Selected types of corpora, adapted from Blecha (2012, pp.15-20)

Corpus	Characteristics
General Corpus	'A collection of material which is broadly homogeneous, but which is gathered from a variety of sources so that the individuality of a source is obscured unless the researcher isolates a particular text' (Sinclair 1991, p.17). These corpora are used when looking at trends across all forms of a language and across time periods.
Monitor Corpus	This is a corpus in which researchers, studying how languages undergo change, continually update and add to the corpus; they may also remove older texts in order to provide users with a constant 'rate of flow' (Pearson 1998, p.45).
Specialized Corpus	This is a corpus that is designed to look at one 'special' facet or a genre, as is the case with the current work, and 'which do[es] not contribute to a description of a language, either because they contain a high proportion of unusual features, or their origins are not reliable as records of people behaving normally' (Sinclair 1991, p.24). They can, however, be used in comparison with a general corpus 'to identify those features of a language that differ from general language' (Bowker & Pearson 2002, p.12).
Reference Corpus	A 'reference corpus is any corpus chosen as a standard of comparison with your corpus' (Smith & Hardie 2009, np). The Brown Corpus is an example of a reference corpus, as could be one of the sections of the British National Corpus (BNC), depending on the nature of the research being undertaken.
Bi- or multi-lingual Corpus	This is a corpus that contains two or more languages and

	is designed to allow researchers to investigate differences between the languages, if the texts from the different languages are collected in the same manner (comparable corpora) or when the texts are translations of each other (parallel corpora).
Learner Corpus	This corpus typically collects output in language A produced by students who speak language B, to allow comparisons with language produced by students who are native users of the target language A.
Diachronic Corpus	This is a corpus that is designed to look at a collection of texts that span a specific time period, thus enabling scholars to follow changes and trends in language over that period. This stands in contrast to a synchronic corpus.

The entries in this table represent but a few of the many types of corpora. For a more robust listing, see the website, *Bookmarks for Corpus Based Linguists*,[5] the list of corpora maintained by the Linguistic Society of America[6], or refer to Corpora4Learning (Sabine Braun's website).[7]

The present work uses a corpus composed of only written texts, with all the texts taken from a ten-year span: since I am looking at only one genre, and hope to highlight characteristics of that genre, this corpus can most appropriately be considered a broadly synchronic specialised corpus.

2.2   Building the medical research article corpus

An overview of previous studies with similar backgrounds is important here to establish the extent to which my findings will be comparable. Four of the major studies on the genre and Moves of MRAs are detailed in the following table by showing the author, years covered by the articles collected, number of articles within their corpus, field of medicine examined, and total number of words in each.

Table 2.2: Previous MRA corpora

Author(s) and year	Number of articles in corpus	Area examined	Publication of articles	Number of words in corpus
Skelton (1994)	50	General	1989-1993	Did not state

[5] http://www.uow.edu.au/~dlee/CBLLinks.htm
[6] http://linguistlist.org
[7] http://www.corpora4learning.net/

		Practice		
Nwogu (1997)	30	Medicine (no specified area)	1985-1987	Did not state
Li and Ge (2009)	25 (Corpus A) 25 (Corpus B)	Medicine (no specified area)	1985-1989 (A) 2000-2004 (B)	68,515 (A) 78,570 (B)
Fryer (2012)	16	Obesity	2004-2006	Did not state

For a more complete discussion of each of the studies, see Chapter 4, on Rhetorical Moves, as that is the primary focus of these studies. However, it is important to note that only one of the studies specified the size of the corpus in terms of word tokens, and all of the corpora contained fewer than 50 articles. Two of the studies (Nwogu 1997; Li & Ge 2009) used three criteria in their selection of articles: 'representativity, reputation and accessibility' (Li & Ge 2009, p.95). Li and Ge (2009) argued that articles selected should represent those commonly found in the genre, and should be drawn from a reputable journal. There is a potential contradiction in these qualifications, because articles selected from some of the highest ranked and leading journals are not necessarily representative of most MRAs. Therefore, to be truly representative, articles should be extracted from a wide range of journals. Li and Ge (2009, p.95) further state that they selected articles written only by native speakers: to do this they used the proxy of the authors' surnames being 'native to the country concerned'. This criterion will not be used in the present study, as it would be extremely difficult to try and determine if an author is a native speaker based solely on their surname. In addition, in the current work, I am more interested in examining language produced in articles in medical journals, some of which are written by NNS. Their criterion of accessibility is one that this work will follow: it is important to select articles to which all members of the DC could have access, and all the articles in the MRA corpus developed here are open-access, as discussed in 2.2.

Skelton's (1994) primary criterion for selection was that the articles be original works, all published in one journal. He chose the *British Journal of General Practice*; however, he did not provide a rationale as to why he decided on a total of 50 articles, stating only that 50 were randomly selected. For his discussion, all articles from 1993, the most recent year of his collection, were used. Fryer (2012) selected articles from four journals*: the British Medical Journal*, the *Journal of the American Medical Association*, *The Lancet*, and the *New England Journal of Medicine*, with articles

drawn only from the subject area of weight and obesity. Fryer (2007, 2012) noted that this narrowing of both journals and subject was done to limit the size of the corpus.

2.2.1   Criteria for article selection

For this work only MRAs written in English, and published between 2001 and 2011, were selected. They were chosen from that specific time span because it reflects a recent period in which non-native English medical personnel wrote about coming under increased pressure to publish. This type of pressure can be seen in the 2013 discussion by doctors in China, highlighted in 1.1.1 and 3.3.7. In addition, this is the period in which the open-access publishing model became widespread, via the free online archives, PLoS (Public Library of Science) and BMC (Biomed Central). Their advent greatly expanded access to peer-reviewed articles, through both increased availability as well as growth in the number of journals. PLoS was founded in 2000, and became a publisher in 2003.[8] BMC became a publisher in 2000.

Previous genre studies of MRA (see 2.2) used at most 50 articles and were primarily interested in looking at Rhetorical Moves; this work will look in addition at specific lexis. Hoping to ensure representativeness of the examples of language use within the genre, I collected 250 articles and ended up with 1,051,368 words in total, to represent examples of authorial stance and persuasive writing in the various sections.

2.2.2   Obtaining and selecting articles

To locate MRAs, I contacted the National Institutes of Health (NIH) in Washington, DC, which maintains the PubMed Central online repository of medical articles. PubMed, sponsored by the NIH, offers free searches using MEDLINE, the bibliographic database of the National Library of Medicine. MEDLINE includes more

---

[8] http://www.plos.org/about/plos/history/

than 21 million references from 1946 to the present, linked to more than 5,600 journals in more than 40 languages.[9]

Anyone may search PubMed for references; however, PubMed Central specifically forbids 'any kind of automated process to download articles in bulk from the main PMC site' without permission.[10] I was granted permission to conduct my research on articles held in PubMed Central, and to download two terabytes of articles from the database. Using a random number generator, I first downloaded 325 articles and then selected 250 articles from the database, within these parameters:

- were peer-reviewed
- were published within the date parameters of my study (2001-2011)
- were not research review articles. While these genres are similar, as I discuss in 3.3.2, they are not identical.
- were not articles written in any language other than English,
- were not articles lacking the section-heading key words *Introduction, Methods, Results/ Findings, Discussion* or their equivalent,
- were not short articles, including letters providing medical findings, of less than 2000 words, editorials, and popularizations

I deliberately did not limit the articles to any one field of medicine, any particular author, any single journal, any impact ranking, or any specific topic. As stated earlier (see 2.2) when discussing the issues with Li and Ge's (2009) work, such limits removed a certain degree of representativeness. Because of copyright restrictions, many of the peer-reviewed articles selected were published by PLoS or in the multiple collections of BMC: on copyright, see McEnery, Xiao & Tono (2006). However, articles from a number of journals that were not open access were also available because the NIH requires that publications from research it funds be immediately accessible on PubMed Central. The collection includes articles with topics as diverse as the effects of plastic on oceanic birds and their subsequent impact on environmental disease, to a study of Maple Syrup Urine Disease. I wanted the topics to be as broad as possible in order to have the corpus be representative of all types of

---

[9] http://www.nlm.nih.gov/pubs/factsheets/medline.html
[10] http://www.ncbi.nlm.nih.gov/pmc/about/faq/#q5

medical research articles. (see Appendix 1 for a list of all 250 articles and Appendix 2 for a list of all journals). Since all of the articles in the database are peer-reviewed as well as edited, I was confident of their being fairly similar in grammatical and stylistic accuracy.

The files in the NIH database are stored in PDF format. As noted above, corpora of texts are machine-readable collections of discourse, in written form, though the discourse may have been written or oral in its original form. In the latter case, it is transcribed in order to perform analyses. The MRA corpus consists of published articles, so the text was already in written form, although not in the format needed for analysis. Once the format is selected, researchers need to be aware that conversion programs may cause errors and that they should go through each text to ensure any 'bugs' are located and fixed (Blecha 2012).

After downloading and selecting the 250 articles, I converted them from PDF to text format in order to input them into WordSmith Tools (Scott 2008), the corpus analysis program I had selected to use (see 2.2.3 for a further discussion of WordSmith Tools). Rather than 'resaving' the files as text files, since not all PDF files support that capability, and the optical character recognition built into word processing programs is not always reliable, I copied and pasted the different sections of the PDF into Word document format, and then re-saved them as text format, thus allowing me to ensure that any errors on transfer could be caught before being examined mechanically with WordSmith (Scott 2008). Each article was split into the four sections by using their section titles: Introduction, Methods, Results and Discussion. In order to focus solely on the narrative text, I then removed all titles, abstracts, headwords, images, captions, figures, tables, references and appendices. This left me with 250 samples in each of the four text divisions, i.e. Introduction, Methods, Results, and Discussion. The linguistic differences between each section will be further elaborated in the following chapters.

Table 2.3: Number of words in the four IMRD sections of 250 articles in the MRA corpus

Introduction	Methods	Results	Discussion	Total words
153,257	286,251	311,394	300,466	1,051,368

2.2.3 The choice of WordSmith Tools

Anthony (2013) states that there are many different tools that allow the researcher to conduct corpus analysis, but that the analyst needs to be sure of what feature(s) they are planning to investigate when selecting the appropriate tool. Some of the most common features of corpus analysis tools are described in the following table:

Table 2.4: Typical features of corpus analysis tools

Feature	Description
Lemmatizer	Allows researchers to examine all forms of a word as 'a set of lexical forms having the same stem and belonging to the same major word class differing only in inflection and/or spelling' (Francis & Kučera 1982, p.1).
Word lister	The lister creates a list of words, and gives researchers the ability to do basic statistical analysis on the corpus, 'for instance, it will calculate the total number of "tokens", count how many times each individual word form appears. […] The words in the list can be sorted in different ways (e.g. in alphabetical order, in order of frequency) to help you find information more easily' (Bowker & Pearson 2002, p.13).
Concordancer	'A concordancer allows the user to see all occurrences of a particular word in its immediate contexts. This information is typically displayed using a format known as key word in context (KWIC). In a KWIC display, all the occurrences of the search pattern are lined up in the centre of the screen with a certain amount of context showing on either side. As with word lists, it is possible to sort concordances so that it becomes easier to identify patterns'. (Bowker & Pearson 2002, p.13)

These are not the only features in a corpus program, but they are the most basic; many programs offer options for parsing and aligners, to support the comparison of a text across two or more languages. These are also the three features that I most frequently utilized in my research on MRAs and as such were a necessity in the program that I chose to use. All support analysis of lexical components of text such as lemma, node, and collocation, which may be defined briefly as follows:

- lemma: a head word; for example the lemma TAKE includes its different word forms, *take*, *takes*, *took*, *taking*, and *taken*;
- node: the word being examined, within a specific number of words to the left and the right of the word;
- collocation: two or more words that co-occur more frequently than chance, such as *cotton candy, best practice,* or *rich and powerful*; they do not have to be adjacent, but they should be close to each other (typically within 3 to 4 words on either side)

An issue seldom mentioned in the literature is the problem of costs. While many

scholars may have institutional backing and or support (i.e. they may already have institutional membership allowing access to certain corpus management programs such as WMatrix®), self-funded researchers may not have that luxury and some of the programs are quite expensive.

I selected WordSmith Tools version 5 (Scott 2008) for several reasons. First, it is well known in the field of corpus linguistics, and used previously in genre studies by Barbara and Scott (1999), Bondi (2001), Tribble (2002) and Hoey (2005). Tribble (2012) asserts WordSmith Tools and AntConc were the second most used software packages for text analysis and that the most popular is a still-growing set of corpora made available online by Mark Davies at Brigham Young University (2.1.1). However, the latter are web-based corpora with integrated tools, not stand-alone tools to use on one's own corpus. Speed and ease of use were other considerations: I was already familiar with the user interface for WordSmith Tools, and it performed analyses with considerable speed. Anthony, the developer of AntConc, acknowledges that WordSmith was at that time speedier (Anthony 2013), but the difference was insignificant for my purposes.

WordSmith Tools version 5 (Scott 2008) has three main functions, all three of which were exploited in this research: a *Concordancer*, a *KeyWord* builder and a *WordList* maker. The collocate feature within the *Concordancer* function allows researchers to see what words appear with the node being examined. The *WordList* function makes a list of the words forms in the corpus and can be used to find different forms of the same word. The *KeyWord* function takes the list of words made in the *WordList* and shows all the variations in form of the node, which allows a researcher to compare one corpus to another.

2.3   Characteristics of the Medical Reearch Article corpus

To summarize, the full MRA corpus is comprised of 250 primary research articles from different fields of medicine, e.g. oncology, viral studies, environmental health, dermatology, etc. (see Appendix 1 for a list of the corpus contents). Each article has been assigned a Text number, in order to keep the example anonymous. The corpus

has 1,051,922 word tokens, and 29,417 word types. Since a one million-word corpus was targeted as being of sufficient size to ensure representativeness, the current work has exceeded that amount. The articles were all published in English and were peer-reviewed before publication. The corpus is specialised and is representative of the MRAs written in English from the PubMed database and the PubMed Central repository between the years 2001 and 2011. Because the corpus includes articles from all areas of medicine instead of just one focus, this inclusion could lead to problems comparing the findings to those from other, more specific, corpora. WordSmith Tools version 5 (Scott 2008) was selected as the primary corpus analysis tool, for its ease of use and familiarity.

This corpus, although wide ranging in types of medical issues, can offer researchers and teachers a representative sample of relevant texts, and will lead to findings that can be applied to materials and instructional designs for NNS and novice writers. This corpus is larger than previous corpora designed to look at the genre of MRAs, but it is nonetheless a specialised corpus. As it incorporates a recent range in years of MRAs (2001-2011), it thereby provides a relatively up-to-date sample of the genre.

Chapter 3: The discourse community sample

3.1   Introduction

Writers of medical research articles (MRAs) must ensure that they are meeting the expectations of their readership, so identifying those expectations can be important. As I am not a medical professional and this work treats MRAs as a genre, in accordance with Swales' (1990) framework on GA, I needed to consult with the normal users of the genre, its discourse community (DC), whose responses help index expectations for a genre (Swales 1990, p.22). The present study uses a sample of members of the discourse community to help determine the exact definition for the genre of MRAs. After a brief description of GA and DCs in general, I will set out the interview procedure and, finally, I will summarise the responses obtained.

3.1.1   Genre analysis: an overview

In 1.1.2, Swales' impact on this work was discussed; this chapter will extend his work on GA to medical research articles (MRAs). Genre analysis, in brief, is an approach with associated techniques by which one analyses a text to determine the parameters that help identify the text as belonging to a certain genre.

Research utilizing GA is an important component of materials development for language learners and language users, as knowing the characteristics and contexts of a genre may help learners and users understand the potential impact on a text from its context (Hyon 1996, p.693). Through studying a text from the perspective of GA, it is possible to delineate a number of the expectations that the users of that genre have of it. For example, different genres have differing expectations about the length of a section or a paragraph of text. Using GA, texts may be seen as having identifiable chunks to allow the learner to predict elements needed in a text to fulfil the requirements of membership in a certain genre (Hyland 2012). Although this discussion will focus primarily on written text, the same techniques are used for the

analysis of oral discourse. As noted, by using a genre-based approach, researchers can identify the parameters constraining the text and see how authors comply with, as well as break from, those criteria.

3.1.2   What makes a genre: purposes and contexts

Swales (1990, p.58) sets out the following definition of what a genre is: 'a class of communicative events, the members of which share some set of communicative purposes which are recognized by the expert members of the parent discourse community'.

Defining a genre means differentiating this concept from a number of related constructs such as register (Biber & Conrad 2009) with linguistic characteristics that appear in text-specific locations (that is, introductory expressions occur naturally in the Introduction section and are not likely to occur in other sections); and finally, that language features are based on genre conventions but are not often functional (2009, p.16). For Biber and Conrad (2009), 'register' is the language that is used in a particular situation, and 'style' is the chosen presentation of language. However, particular stylistic features are often what help a researcher identify a genre. With MRAs, style will include a high amount of medical terminology and phrases that are specific to MRAs, such as:

> (1) The study was approved by the Institutional Review Board of the Medical and Health Sciences Colleges at the University of Sharjah and was implemented in compliance with the Helsinki Declaration. (Text 101)

In the above example, words like *Medical* and *Health Sciences* help inform the reader that this text belongs in the realm of medical sciences. Technical terminology is abundant, such as the undefined *Helsinki Declaration*: this text expects the reader to identify the *Declaration* as a specific protocol regarding the handling of animals and human test subjects in research within the medical field. Such wording typifies what Biber and Conrad would call 'register'. 'Style' in Example (1) is the use of formal language ('was implemented in compliance with') and rhetorical devices that help identify a text as belonging to the genre of MRAs.

Probably the most familiar definition of a DC for writing specialists is that offered by Swales (1990), as he first differentiates a discourse community from a speech community by showing that, among other things, they have different purposes: the DC is sociorhetorical in nature with its communicative goals, while the speech community fulfils social functions (1990, p.24). He identifies six characteristics that identify a DC; these are summarized in 3.1.3 in which MRAs are compared to those characteristics. To examine the DC, I draw on Swales (1990) and Bhatia (1999) and on a set of interviews I conducted personally between April and June 2012, with 5 editors of medical journals, and with 10 doctors who have published or use MRAs in their own work.

3.1.3   What makes a discourse community

In his seminal work on GA, Swales (1990, pp.24-27) developed requirements for the 'defining characteristics of a DC'. He begins by explaining that the DC must have a broadly agreed-upon set of goals, which are open to the public. Swales' second condition for the DC is that there must be a built-in communication channel for the members. The third concept underlying the formation of a DC is that the members can offer input to the field. The fourth characteristic is that the DC can use one or more genres in its repertoire; while the fifth is that the DC has its own vernacular and technical jargon. The final component is that there should be a 'threshold level of members with a suitable degree of relevant content and discoursal expertise' (Swales 1990, p.27). The following table maps Swales' characteristics onto the DC of writers and readers of MRAs.

Table 3.1: Mapping Swales' characteristics of a DC (1990, pp.24-27) onto MRAs

Characteristics of Swales' Discourse Community	Characteristics of the Discourse Community of readers and writers of MRAs
1. Broadly agreed set of common public goals.	Anyone can look at the journal's requirements; the public understands that MRAs are about medical problems.
2. Mechanism of inter-communication among its members.	In addition to MRAs (which by peer review and by citing other articles are themselves a form of inter-communication), this is achieved primarily

	by online Internet list-servers or by postings on the Internet message boards of the journals themselves. Such boards can be found listed on journal websites, or through organizations such as the American Medical Writers Association (AMWA), a gathering of professional writers and editors.
3. Uses its participatory mechanism primarily to provide information and feedback.	Since the citation of another article in an author's work is a form of feedback, this mechanism is built into MRAs.
4. Has one or more genres in the communicative furtherance of its aims.	Aside from MRAs, the members have research review articles, conference papers and presentations, along with letters to the editors of journals.
5. Has acquired some specific lexis.	MRAs contain specific technical terminology.
6. Has a threshold level of members with a suitable degree of relevant content and discoursal expertise.	Associations like AMWA help doctors and researchers to publish, along with senior fellows who help novices learn how to write and publish articles.

The table illustrates that the DC comprising readers and writers of MRAs meets Swales' requirements.

One way to define a genre is through the eyes of the DC that utilizes it. Bhatia (2002, p.14), in his work on how to implement genre studies, said:

> Interviews and case studies have become increasingly important means of collecting data in academic and educational contexts, particularly for the investigation of developmental aspects of learner language, learning style preferences and writing practices of expert writers.

The answers to genre-related questions such as the one above can perhaps best be given by experts and members of the DC that use MRAs as an intercommunication mechanism. The following section will outline how I established a sample of members of the DC.

3.2   The discourse community sample: an overview

As previously mentioned, in this work I have undertaken the task of building a sample of members as a representation of the DC that I could interview about their perception of characteristics of the genre. This section contains a description of issues around building the sample, then a description of the members who comprised the sample of

the DC, and finally, an outline of the questions asked and how the interviews were conducted.

3.2.1 Challenges to building a discourse community sample

Identifying a discourse community as a part of analysing its written or oral communication is now a reasonably common way to begin the study of writing in different environments such as the workplace. A good example of writing in the workplace from an ethnographic perspective is a study of office language by Beaufort (1997). She interviewed the participants, four office workers, numerous times over a one-year period and then followed up by asking the participants' supervisors for comments regarding what the four had said. She did this to confirm that what the subjects had reported made sense within their particular context (and as part of her ethnographic practice). Beaufort (1997) made sure to include members who were involved in all the facets of the genre: writing, community communications, and readers of texts. To emulate that involvement, members of the MRA discourse community sample needed to have written and read MRAs or to serve as editors, as they are integral to the production of the genre. Doctors who conduct and write up their experiments may perhaps read articles differently from editors, and when doctors write, they may write with different assumptions from editors with regards to what the genre requires.

There is no one-size-fits-all answer to the question of what constitutes a representative number of members in a sample, as there are many factors to consider when designing the research, including 'time given to complete a research project, finding and keeping in contact with participants, and the institutional demands of ethics committees' (Baker & Edwards 2012, p.6). They also recommended that there should be more than 12 but fewer than 60 people (2012, p.6). Their recommendation was for the DC to serve as the entire piece of research; however, as the DC is one part of my study, I decided on a total of 15, from novice to experienced writers, and from editors who work at major journals to freelance editors whom doctors can hire through an association of medical writers.

3.2.2   The discourse community sample: composition

I interviewed ten doctors from around the world, and five editors of medical journals, tying the interviews to a standardized questionnaire (see Appendix 3), which was sent to the doctors and editors prior to the recorded interview. This was a surprisingly difficult endeavour: both groups of experts are very busy and were reluctant to spare the time to participate in the interviews. Due to their busy schedule and logistical issues including differences in time zones, some of the interviews were face-to-face, others were conducted over Skype, and for four of them I accepted written replies to the questionnaire. This mix of methods for interviewing can lead to some problems, of course: on the one hand, people have greater time to think through their replies in written responses, and have more opportunity to be prompted to expand their answers if their interaction is oral. The following table shows how each interview was conducted, the country of origin of each participant and the means of introduction.

Table 3.2: Doctor and editor interviewees

Doctor	Means of conducting the interview	Country	Means of introduction
Doctor 1	Skype	US	Family friend
Doctor 2	Skype	UK	Message board
Doctor 3	Written	Brazil	Message board/personal introduction
Doctor 4	Skype and written	Holland	Message board/personal introduction
Doctor 5	Written	US	Personal introduction
Doctor 6	Face-to-face	UK	Family friend
Doctor 7	Face-to-face	UK	Message board
Doctor 8	Written	Taiwan	Family friend
Doctor 9	Skype	US	Family friend
Doctor 10	Skype	US	Family friend
Editor 1	Skype	US	Email solicitation to the journal
Editor 2	Skype	US	AMWA*
Editor 3	Skype	US	AMWA
Editor 4	Skype	US	AMWA
Editor 5	Skype	US	AMWA

*Note: AMWA stands for the American Medical Writers Association, an organization that allows editors and doctors to connect with each other about publishing articles

To locate editors, I contacted some of the leading journals in the field of medicine, such as the *New England Journal of Medicine*, which required me to apply for

permission through their media relations office. By joining the American Medical Writers Association as a scholar, I was granted access to their directory of freelance writers and editors. This allowed me to contact over 20 editors who either are freelance medical editors or attached to specific medical journals, of which, I was able to conduct four interviews in this way with editors. The following table shows the alias given to the doctor or editor (the number of the interview is their assigned number or alias), their area of expertise, their number of years practising, the total number of articles they have published (doctors) or edited (editors) in peer-reviewed journals, and their first language.

Table 3.3: Bio-data of doctors and editors

Alias	Areas of expertise	Years practising	Number of articles written or edited	First language
Doctor 1	General Practice	16	1	English
Doctor 2	Anesthetics	3.5	1	Scots/English
Doctor 3	Andrology	12	3	Portuguese
Doctor 4	Sexology	39	15	Dutch
Doctor 5	Pulmonary & Critical Care	3	2	English
Doctor 6	Paediatrics	38	Approximately two hundred	English
Doctor 7	Occupational Medicine	25	0	English
Doctor 8	Anesthesia	20	45	Mandarin
Doctor 9	Physical/Rehab	3	2	English
Doctor 10	Paediatrics	19.5	38	English
Editor 1	Colon and Rectum Disease	25	Over one hundred	English
Editor 2	Cardiology	23	Over one hundred	English
Editor 3	Hematology/Oncology	20	Over one hundred	English
Editor 4	Gynecology/ Oncology	15	Nearly a thousand	English
Editor 5	Molecular cancer/genetics	15	Over one hundred	English

3.2.3   The interviews

The oral interviews lasted between 20 minutes and an hour and a half. Three of the doctors declined an oral interview: two were concerned that their spoken English would not be good enough to conduct the interview so they submitted lengthy written responses to the questionnaire instead, and a third had to submit a lengthy written response to the questionnaire due to serious time constraints. A fourth doctor did complete an oral interview, but the quality of the recording was so poor that he also supplied a written response when requested.

The doctors' questionnaire was composed of three parts: a biodata section, a response section with 20 open-ended questions, and a chart of specific features typifying MRAs, such as materials used in the experiment or limitations of the study (see Appendix 3). Participants were asked to place each feature into one of the four sections of a typical article: Introduction, Methods, Results, Discussion, and two other sections entitled Abstract and Elsewhere for the components that the respondents did not feel would occur in the main IMRD sections. The same format was followed for the editors' questionnaire with one difference: the editors were asked an additional five questions on language usage and linguistic concepts, such as the use of hedging devices or signposting in developing the framework for an article, as they would have had more exposure to these concepts.

In both questionnaires, the most pertinent questions concerning issues of genre were the following (words in parentheses indicate the words added or changed for the editors):

1 What is your own definition of a research article?
2 Is there a difference between a research article and a review article? If so, please explain.
3 How do you feel about (editing for) journals that are online-only vis-a-vis printed journals?
4 When you read (edit) an article, do you read straight through or jump around, and why?
5 What is the hardest part of reading research articles?
6 Have articles changed in your time as a doctor (an editor) and if so, has the change been for the better or the worse?
7 Are articles too data-driven or not enough?
8 Have you ever noticed any difference between American doctors' papers and non-American or foreign doctors' work?

The questions aimed to elicit a range of types of information, i.e. on genre, lexis, and linguistic features. Participants were asked to explain what research articles were in order to define the genre clearly. This question was posed to explore whether there might be variation in the responses from doctors and editors. Participants were asked

in Question 2 to distinguish between research articles and research review articles. Question 3 investigated the potential for other forms of the genre: since online publishing is becoming more frequent, could online-only journals be perceived as something distinct from journals available in both print and online formats. Question 4 dealt with how the reader interacts with the text, possibly signalling patterns that would help the genre stand alone through their use. Question 5 again examined audience reception, to see if differences were perceived between the author and the intended readership. Question 6 was used to elicit users' views on possible diachronic shifts within the genre. With Question 7, I hoped to discover whether there was a collective idea shared by members of the DC as to possible criticism of the genre. Question 8 explored whether there were variations that occurred due to language or disciplinary practices in different countries.

3.3   Responses of the discourse community sample

The responses to questions posed in oral interviews and written questionnaires by the doctors and editors allowed the identification of a number of themes, including their definitions of MRA. A few participants mentioned that original research is the distinguishing feature defining the MRA; others commented that the MRA is problem-based. Several framed their explanations of differences as being based on the impact the MRA had on the research community. For example, Doctor 9 stated that he considered an MRA to be a write-up of an experiment done to benefit the field and to advance science. Editor 2 stressed that the MRA was the writing up of an experiment based on a plan and not done 'off the cuff'.

3.3.1   The discourse community sample's definition of the genre

Four major themes emerged from the analysis of the sample's responses; these are originality, problem-centeredness, premeditation, and the effect on the field. This supports the formation of the following definition of the genre, which will be drawn upon in the linguistic analysis that follows in later chapters:

A piece of original written research on or into a specific medical problem, either in previous treatments or as a new issue, carried out according to a design whose findings are measured and reported along with a discussion of the implications for the community.

The MRA needs to be an original work; often an author will replicate an experiment, but change one or more variables; as long as the authors are doing the work themselves, it will constitute a fresh look at the problem. Next, a problem needs to be identified, either within the field or in relation to a patient. That is to say, the author may be arguing in favour of a new technique to advance the field, or presenting a new treatment to improve patient outcomes. A third aspect of the definition is that the experiment must be done with forethought and planning. In the interview, Editor 4 raised the issue of grant writing now being at the beginning of all research, so there is already a focus on relevance and procedure that may be where ideas for the presentation of the research originate. Finally, the article needs to be immediately relevant for helping the advancement of the field or sub-field in general as well as presenting outcomes of the problem and suggestions for the restricted or specialised discourse community interested in the particular topic.

3.3.2   Difference between medical research articles and medical research review articles

When differentiating between a research article and a review article, almost all of the editors and doctors in the sample explained that since a review article was being used to review primary research, it should be considered as secondary research. A few of the clinical doctors added that they preferred review articles. As Doctor 7 stated, since review articles cover a great deal of the literature, it is easier for him to apply the findings to his practice. Reviewing these interviews helped me to determine that while review articles and MRAs differ in their impact and their scope, they are still similar enough that research review articles could be considered a sub-genre of MRAs, particularly given the distinction made by the DC sample between primary and secondary research, which demands a different stance from the authors. Myers (1991, p.56) comments that:

> The persuasive power of a review arises not just from the apparent coherence of its story, but from its ability to enlist readers and to make them see their own work as part of this ongoing project.

In MRAs, the authors wish to persuade readers that their particular version of empirical truth in a specific experiment is the most plausible and credible, while in medical research review articles, the authors need to persuade readers of their trustworthy stance in summarizing a series of experiments and in identifying those versions of empirical truth which seem 'best' according to the criteria for the medical research review article.

From comments made across my sample of the DC, MRAs and medical research review articles apparently have similar purposes in providing information and results, which can be used to advance the practices of doctors. Linked to the distinction made by practitioners between primary and secondary research, research review articles that review medical studies may be a sub-genre of MRAs. While it is true that they have similar language and structures, even including the IMRD (Introduction, Methods, Results and Discussion) structure (although review articles usually use Methods to describe the process of selection of articles to review), the difference is in how the research is conducted. For the MRA the experiment is conducted by the author(s) who place that experiment in the context of a review of previous research, while for a research review article, the experiments of others are reviewed by the author(s) as the primary and/or exclusive focus of the article.

3.3.3   The use of statistics in the medical research articles

Talking about differences between MRAs and research review articles often led the doctors to discuss data analysis. Given the expressed concern by several of the interviewees about the quality of writing in MRAs (see 3.3.4), I asked the members of the community if they felt that MRAs were overly reliant on statistical analysis and therefore too data-driven, or if there was not enough focus on data. Two members had no opinion regarding this issue; four felt there was too great a focus on data and statistics, while two felt there should be an increased focus on data and statistics. Four doctors and three editors, the largest group of the sample, believed that that the focus

on quantitative data was appropriate. A few of the doctors admitted that they did not like statistics, as they were not comfortable with the practices surrounding different statistical tests. As Doctor 2 argued, the MRA is reporting the results of an experiment and therefore needs to have data. While the DC sample is small, this seems to suggest that while some would like there to be less statistical information, most are comfortable with how the genre utilizes and presents the data.

3.3.4   Reading habits in the discourse community

I then examined whether self-disclosed reading practices for MRAs could help define and clarify the genre. One set of questions focused on members' reading practices, i.e. did they read an article consecutively, from beginning to end, or jump around among identified sections, and could they identify the most difficult feature in reading MRAs. All but a single person in the sample reported not reading an article straight through. Some read the Abstract first to help determine if they wanted to read the article at all, while others start with the Discussion. Doctor 2 added that the reason he reads only parts of sections and does not always follow the IMRD structure when reading is that articles can be too tedious to read completely, a point others mentioned with regard to what they felt was the hardest part of reading MRAs. Doctor 5 explained his rationale as follows: 'I read the abstract first to see if the article is something I care to invest time in reading. I then look at the figures first because they should tell the 'story' of the article concisely. I then read start to finish'.

In terms of the difficulties that this sample of the DC identified with reading an MRA, there was a consensus among the editors that their largest challenge when reading newly submitted articles was poor writing. Several editors also commented that a number of published articles were weak. While interviewing them, I probed for whether they saw the problem as poor scientific writing or weak grammar: three editors claimed that the problem lay in the fact that the writers could not express their ideas with clarity and purpose, while the other two editors responded that the mixture of weak grammar and poor scientific writing was troublesome. For the doctors, four identified statistics and their uses and explanations as problematic, asserting they did not always have enough statistical knowledge themselves to understand what the

author was claiming in published articles. For example, Doctor 8 answered the question, 'What is the hardest part of reading a research article?' as follows: 'Results, because sometimes you have to think of how they obtain these results. Do the statistical analyses make sense? Questions like this would keep coming'. Three doctors (Doctors 2, 3 and 4) complained that the language and tone of published MRAs are too monotonous and boring.

There seems to be an inherent 'testing' quality accompanying reading MRAs, meaning that doctors and editors are wary of committing to reading the whole article as it could turn out to be a waste of their time. Some of them skip one section while others skip another; overall, the most skipped section was the Methods, unless 'a question arises, then I will read methodology to go deeper' (Doctor 9). All but two of the doctors start with the Abstract and, depending on their needs, such as a problem in an area of clinical practice, will move on to the section that holds the pertinent information. This signals that each section in the IMRD structure has a specific role, something I will be discussing at length in the next chapter (see 4.3).

With articles being boring and poorly written cited as the two main problems facing readers of MRAs, it is difficult to pinpoint what needs to be addressed. The doctors complained that the articles were uninteresting and the editors claimed the problem was bad writing. There was, however, consensus that there was a problem in the realm of writing even though no one was able to be specific beyond comments about a need for greater clarity linking Results to Methods, perhaps some expansion in explaining the use of particular statistics, and what several called the 'monotonous' tone. This is relevant because it indicates one of the possible causes for the genre to change. In Chapter 4, I will describe how genres change, but the fact that members of the DC are dissatisfied with the quality of writing may indicate one of the internal causes for change in the genre.

3.3.5   Changes in published articles

The next genre-related question was, 'Have articles changed in your time as a doctor (an editor) and, if so, has the change been for the better or the worse'? I designed this

question to obtain perceptions of any changes in the field over the span of their collective experience and to elicit possible reasons for their perception of evolution within the genre. Six of the doctors felt there has been no change within their time practising, although three of them have been practising less than five years. One of the doctors mentioned, however, that the biggest change for him was the introduction of what he calls a 'league table' or a ranking mechanism that has influenced journals greatly (Doctor 7; such tables rank journals by impact factors). Three of the editors cited the growth of the Internet as the biggest change, while two editors were divided, with one saying the greatest change was the decline in the quality of science writing, while the other thought it was getting better. One editor and three doctors mentioned the fact that in their time, with an average of over 17 years of experience, journals have become more focused and specialised. Doctor 4 voiced his or her opinion about the most negatively seen change:

> What I said right now about boredom: that is definitely a development that seems to be unstoppable. Yet another annoyance is the pressure on quotations: as a reviewer myself, I often comment that I find the number of quotations unnecessary and sometimes ridiculous (even in an introductory paragraph every sentence ending with at least one source). But I will sometimes be asked for my own articles by reviewers to give more sources, for details that I find self-evident.

3.3.6   The impact of technology on defining the genre

One of the issues facing the genre is the emergence of the Internet as a channel for publication. One question asked if any members in my sample of the DC saw a difference between print journals and journals published exclusively online. Three members of the DC commented that they feel that the Internet is the eventual medium for the genre, and that within ten years (in the 2020s) almost all journals will be online. When asked to identify the biggest change in their time as editors, two replied that the growth of online publishing was the largest change. Editor 5 went on to state that the Internet allows journals to keep the size of the print article down since the author may post an appendix and supplementary findings on the Internet. A difference between the two groups of respondents that comprise this sample is that only one out of the five editors felt that online-only journals were not as reliable as print, while six

of the ten doctors commented on the fact they felt that online-only journals were not as reliable. There was a difference in the level of credibility between the two media, print and online, in that online-only journals were seen as being less authoritative. However, two of the doctors remarked that with recent advances in technology, such as tablets and smart phones, they preferred online to print for its ease of access (Doctors 6 and 7).

Conversely, one of the editors who disparaged online publishing wondered whether it was from his/her own distrust of the form and commented that he or she may be 'too old' (Editor 3). From the standpoint of the DC sample, it seems that the transition is inevitable, though there is some concern over the possibility of MRAs losing credibility and the potential for weak peer reviews or for an unreliable peer review system (Doctor 6).

While concern was expressed about online publishing, there were some other benefits besides ease of access identified by some of the members of the DC. Editor 5 talked about how hosting articles online allowed for more supplementary information to be displayed and one doctor noted the ecological benefit of reducing paper that could be found in a shift to online publishing (Doctor 9).

The implications seem to be that the MRA is a genre in flux, between the channel in which it has been traditionally delivered, and a new cyber-channel that will allow greater access and ease of use, but members of the DC feel there need to be mechanisms in place to ensure that the credibility and integrity of peer review is maintained.

3.3.7   Differences across the globe and the impact of house styles on the genre

Given the variety of experience and nationalities represented in this sample of the DC, small as it was, it was appropriate to ask whether doctors and editors felt that national borders had any effect on the genre: that is, would a journal in the US or the UK be seen as significantly different from a journal published in another country and potentially in another language? Two of my respondents reported that they only read

articles from the US or UK as they feel that the studies conducted there would best match their clinical needs. Three mentioned that the only difference they saw in English-language journals in the US or UK was in spelling, while five claimed there was no real difference, and two claimed not to have noticed any differences. However, Doctor 4 (Northern European) asserted that there were major differences between US/UK journals and continental European journals from Germany and France; Doctor 4 further added that 'In my field, the Journal *Sexologies* (French/English) is much more intellectual than most American journals'. When pressed to explain 'more intellectual', Doctor 4 responded that the French journal is not as concerned with statistics but rather allows a narrative to develop. Doctor 8, perhaps more concerned with pressure to publish and journal rankings than with house style (Asian) explains that journals in his/her country are selected by the doctor based on their impact factor as there are regulations tied to job advancement based on it: 'In addition, here in Taiwan, we always ask the impact factor (IF) of the journals before submitting our paper because the Department for Education asks IF for your lectorate/professorship'. Conversely, Editor 4 noted that he/she feels that Asian journals are not as prestigious, due in part to their lacking robust data as a requirement.

Therefore, while there is some difference across land borders, it does not seem to be a crucial element in the use and production of journals. Instead, as noted by Editor 5, the most important cause of difference, it seems, is the 'house style', referring to the style and format that each journal requires of manuscripts submitted to them for publication. According to O'Connor, Cooter, and Ufnalska (2013, p.2):

> Journals and publishers usually prepare style sheets: lists of their preferences on what is to be abbreviated, capitalized, italicized, or hyphenated, and on what spelling or punctuation should be used when acceptable alternatives exist. This is house style, built up from a mixture of experience and editorial idiosyncrasies. As well as encouraging consistency, house style saves time that would otherwise be spent making decisions about the same matters over and over again.

From this quotation, it would seem that a large part of house style is based on copyediting, for example the *British Medical Journal* does not want a comma before

the word *and* in a list (the 'Oxford comma'), while *The Lancet* does.[11] However, there is another level of impact that does affect the structure of the papers and hence the genre, in which house styles recommend different things that need to be included in the paper. For example, as of January 2015, articles submitted to the Lancet family of journals now must include a 'panel putting their research into context with previous work', as well as a section in the Discussion of how the current work fits that context (see also Kleiniert et al. 2014, p.2176). Neither the *New England Journal of Medicine*[12] nor *BMJ* require such a panel, which is basically a single-column table summarizing available evidence. The panel is a new addition to the house style of *The Lancet* but it will cause a minor shift in how papers are written for that journal.

3.4   Conclusion

As Swales notes (2004, p.63, citing Fishelov 1993), genres are similar to a biological species inasmuch as they will adapt, and/or possibly splinter over time due to evolution. Similar patterns emerged amongst the members of the DC in their response to the question regarding if and how articles have changed 'in your time'. Some of the doctors who have only been in practice a short while found no change; however, those in the sample who had practiced longer than 10 years did find some changes (as described above), as did all of the editors. In general, the editors remarked on the previously discussed topic of the impact of the Internet on articles. Given the current emphasis on publication for doctors and researchers, many may need to publish in order to meet the requirements of their institution, with the potential to cause a watering-down effect in the science, since articles could be written for the benefit of the author as opposed to the field.

In summary, the aim of this chapter was to elicit and analyse responses from a sample of the members of the discourse community who read, write and edit MRAs, in order to better understand and define the genre. The analysis of the responses from those interviewed suggests that it is possible to claim the MRA as its own genre, and to

---

[11] http://www.bmj.com/about-bmj/resources-authors/house-style
http://www.thelancet.com/pb/assets/raw/Lancet/authors/lancet-information-for-authors.pdf
[12] http://www.nejm.org/page/author-center/manuscript-submission

define it (see 3.3.1), even with the proviso that as a genre it has changed and will continue to change.

In the following chapter, I will apply discourse analysis to identify the linguistic features of the genre, especially as they bear on rhetorical structure. This should help instructors to describe what is entailed in the writing of an MRA.

Chapter 4: Mapping of Rhetorical Moves

4.1   Introduction

The previous chapter defined the genre of medical research articles (MRAs) through responses to interview questions by a sample of members of the discourse community. This chapter utilizes another aspect of Swales' (1990, p.140) framework for genre analysis, the study of Rhetorical Moves in a genre. The reasoning behind conducting a study of the Moves throughout the full text of an MRA is to understand how the genre is structured. According to Biber and Conrad (2009, p.131), research articles are one of the clearest examples of how a genre can be 'conventionalized'. However, MRAs are not as well understood or studied. For example, most research articles, medical or otherwise, follow an organizational framework or meta-structure of Introduction, Methods, Results and Discussion (IMRD). This overall structure produces a scaffold within which authors may work, and since each section has its own purpose, the delineation between the sections would seem to be obvious. However, as suggested by an analysis of the Moves in the MRAs in this corpus, there can be some overlap within or repetition of information from different sections of the MRA. This repetition can be seen as the rhetorical device of amplification. A number of rhetorical devices are present within an MRA: the present study attempts to address how the Moves within a genre have rhetorical function and are themselves rhetorical devices. This repetition may not be seen as such by the audience, depending on how a person chooses to read an article. In personal interviews with my sample of the discourse community, both doctors and editors commented that they as readers typically jump around between sections and rarely read an entire article from start to finish (see 3.3.4). If reading an article piecemeal in this manner, they may not view repeated or duplicated information as repetition but rather as the first time they read the information.

This chapter begins by describing what comprises a Move, both in general and in MRAs specifically; it presents a brief overview of studies of individual sections of MRAs and then focuses on how other researchers have identified and located Moves across all sections in the full texts of RAs and of MRAs. In 4.2, a comparison with their findings will allow me to differentiate my own framework for deciding what

constitutes a Move and its components, and to describe my process of locating the Moves. A comparative tabulation of Moves heading the discussion of the main IMRD sections of each article will be the basis for my explanation of how my framework differs from those proposed by other researchers. In addition, I will identify grammatical patterns of voice and tense, and characteristic lexical patterns, following Gledhill's study of colligation and its contribution to cohesion (Gledhill 2009). In a number of instances, some of these lexical patterns can also be analysed as formulaic: that is, as frequent multiword utterances or as part of conventionalized routines such as stating ethical consent, study limitations or characteristics of studies by other researchers, as well as reiterated discourse markers, hedges or specific verb choices (Wray 2002, p.8; Oakey 2002; Durrant & Mathews-Aydınlı 2011, pp.62-64). As part of this analysis, I will attempt to identify one or more rhetorical functions underlying a particular Move as part of mapping out the Moves and their Steps, as they arise throughout a full article. Finally, in 4.3, I will present an overview comparing the present work with studies by other researchers who also focused on identifying Moves as discourse organization across a full research article. As Upton and Cohen (2009, p.3) comment in their discussion of corpus-based analysis, 'we know little at present about the general patterns of discourse organization across a large representative sample of texts from a genre'. Given this lack, the current work will try to fill the niche by looking at MRAs.

4.1.1   What is a Move?

Part of Swales' framework for doing Genre Analysis is finding the rhetorical organization within texts from that genre (1990, p.140). A way to study their organization is to look at what Swales calls 'Rhetorical Moves'. A Move can be broadly defined as a shift in the information flow: that is to say, a change in topic, or in information being covered. In his 1990 work, Swales outlined what he felt were the three Moves in an Introduction section of a research paper: *Establishing a territory; Establishing a niche*; and *Occupying the niche* (p.141). Swales further added 'Steps', or 'sub-Moves', as Nwogu terms them (1997, p.124), as a way to better define the requirements of the main Move.

Swales argues (1990, p.213) that when discussing Moves and how to apply them pedagogically, it is important to alert learners to the rhetorical aspect of the Move. Accordingly, I will attempt to build on the work of scholars such as Teufel and Moens (2002), Kanoksilapatham (2005), and Fryer (2012) in identifying the uses of rhetoric in MRAs by tying the different Moves to one or more possible rhetorical devices. There have been several attempts to develop a more encompassing definition of a Move, such as that by Holmes (1997, p.325), who defines a Move as: 'a segment of text that is shaped and constrained by a specific communicative function'. However, using this definition introduces an issue in studies of Rhetorical Moves, in determining what has constrained the Move, that is, whether it is possible to identify the Move's boundaries in an objective manner, or is the identification more subjective. This issue will be further explored in 4.1.3.

4.1.2   Importance of studying Moves

Since the present research applies genre analysis to a specialised genre, it must address the question of why it is necessary to identify Moves. To write an MRA, which is a highly specialised text, writers need more than just the ability to write clearly on a complex topic: they need to conform to the expectations of the genre, one of which is producing a work that covers themes that the audience thinks should appear (De Groot 2011, p.129). Upton and Connor (2001, p.314) stated:

> Attention to writing for specific purposes is crucial since particular tasks require additional strategies beyond general writing ability. Furthermore, knowing the situation, context, and stimulus is important since these may elicit different types of language based on cultural differences in interpreting purpose and genre by native and non-native speakers of the language.

The goal of the present research is to provide authors, especially second language writers, with a description of the Moves in an MRA, tied to the rhetorical principles that they are being used to fulfil, so that writers may better understand and therefore more easily produce texts in the genre.

4.1.3   Defining a boundary

Moves organize the flow of information within a text. While acknowledging that textual organization is important to examine, Paltridge (1994) questions how limits can be established for the definition of each Move. His concern is whether a way exists to make the establishment of boundaries more precise and less subjective, which in turn supports Bhatia's (1993, p.87) claim that when doing this type of analysis, researchers often have to make up their own methodology, which he finds troublesome and not always reliable.

Some textual boundaries can be seen as physical: for example, headwords and phrases such as *Limitations of the study* clearly demarcate a Move for discussing the limitations of that study. Other types of boundaries can be seen from spacing or indentations for new paragraphs, although some Moves go beyond the paragraph boundaries. Clause boundaries, which are often marked by coordinating or subordinating adverbs, are sometimes taken as a model for discourse relations in text. For example, Webber (2004, p.1) comments that Mann and Thompson (1988) extended their observations on clausal relationships using a systemic functional approach to what they have called Rhetorical Structure Theory (RST), which attempts to structure discourse relations and is used to identify different text sections in computer-based textual summarization. RST is what its authors call a 'framework for describing rhetorical relations among parts of a text' (Mann, Matthiessen & Thompson 1989, p.5), used to 'identify particular relationships that can hold between two text spans' such as the relationships marked by conjunctions (1989, p.11).

However, the problem remains of how to define the limits of a Move that cannot be immediately signalled by overt boundaries such as headwords or clearly demarcated clauses. In this study, I follow Kanoksilapatham's (2005, p.272) framework of setting boundaries based on 'content and linguistic criteria'. Each Move has a purpose or what Holmes (1997, p.325) called *communicative function*, which is visible in the content and is defined by linguistic cues (see 4.2). Linguistic criteria refer to key words or phrases that act like discourse markers, which will also be discussed in further detail in Chapter 6.

4.1.4   Previous studies of the analysis of Moves in full-text scientific articles

Many researchers have looked at Rhetorical Moves in individual sections of research articles, across numerous fields (e.g. Sciences, Law, Economics). For example, the *Introduction* section has been looked at by Swales (1981), Cooper (1985), Hopkins (1985), Crookes (1986), Samraj (2002), and Ozturk (2007), while the *Methods* section has been examined by Lim (2006), Swales (1990), and Bruce (2008). Brett (1994) studied the *Results* section in sociology research articles; Yang and Allison (2004a) developed a Move-structure framework for the *Results* section in applied linguistics articles. The *Discussion* section was analysed first by Adams-Smith in 1984, and other studies include Hopkins & Dudley-Evans (1988), Swales (1990), Lindeberg (1994), Holmes (1997), and Peacock (2002). According to Swales (1990) and Yang and Allison (2004b), neighbouring sections in the IMRD framework have an impact on each other; this suggests that examining only a single section might not allow a researcher or a reader to see the entire text as an interconnected piece of work, and could thereby cause the researcher not to see the impact of one section on another. However, while several researchers, such as Sheldon (2013) analyse more than one section in the same discussion, very few have tried to describe all the Moves for an entire genre. The most notable attempts are by Skelton (1994), Nwogu (1997), Posteguillo (1999), Kanoksilapatham (2005) and, more recently, Li and Ge (2009) and Fryer (2012). All of them, like this current work, start with Swales' framework as a basis.

For the remainder of this study, I will primarily focus on the works of Skelton (1994), Nwogu (1997), Posteguillo (1999), Kanoksilapatham (2005), Li and Ge (2009), and Fryer (2012), as they also focus explicitly on Moves. The following table compares the different works, focusing on size, area of focus (e.g. biochemistry), and dates of articles in their corpora.

Table 4.1: Comparison of six previous studies of all sections of RAs

Name	Number of articles	Area examined	Publication dates of articles
Skelton (1994)	50	Medical	1989-1993
Nwogu (1997)	30	Medical	1985-1987
Posteguillo (1999)	40	Computer Science	1992-1993

Kanoksilapatham (2005)	60	Biochemistry	2000 only
Li and Ge (2009)	25 (Corpus A) 25 (Corpus B)	Medical	1985-1989 (A) 2000-2004 (B)
Fryer (2012)	16	Medical	2004-2006

4.1.5   Description of six previous studies

At first glance, the connection between the six works set out above seems slight. However, all six studies influence the current study, as they all draw on genre theory, offer Move frameworks, and analyse the entire article.

4.1.5.1 Skelton

Skelton (1994) is the first to look at entire MRAs, using 50 articles published between 1989 and 1993. He develops his corpus using only articles that had been original works. He includes submissions from a single journal, the *British Journal of General Practice*: while a well-respected journal, it would be difficult for this corpus to be seen as representative of the field as a whole. As he mentions (1994, p.455), most journals state their own requirements or suggestions to the author, which can have an impact on the structuring of a paper. For example, some journals such as the *British Journal of Anaesthesia*[13] want the aims of the study to be explicitly stated in the Introduction section, while others do not mention this requirement, as exemplified by the *British Journal of General Practice*[14] (see 3.3.7 for a discussion of impact of house styles).

Skelton (1994) found 15 Moves utilizing the CARS framework and noted that for the most part (with the exception of three Moves) the Moves he found followed a certain pattern and could be considered, based on frequency scores, to be in a set order. In doing so, he makes the claim that MRAs may be studied to develop a template for others to follow.

---

[13] http://www.oxfordjournals.org/our_journals/bjaint/for_authors/general.html
[14] http://bjgp.org/authors

4.1.5.2 Nwogu

Nwogu (1997) examines entire articles but chooses MRAs only from top journals such as *The Lancet* and *New England Journal of Medicine*. While working with selections from the leading journals should logically allow researchers to claim that their findings can be applicable across all MRAs, the interviews with a sample of the discourse community (Chapter 3) reinforced what Skelton (1994) learned about the importance of a journal's in-house style guide. Nwogu's corpus contains articles published between 1985 and 1987, and they may not have the same Moves as articles written today, according to Li and Ge (2009).

Nwogu (1997, p.22) describes his process for identifying Moves: 'Moves and their constituent elements were determined partly by inferencing from context, but also by reference to linguistic clues in the discourse' (1997, p.122). Among his clues are temporal and locative adverbials (1990, p.126). He describes 11 Moves, with three of them classified as 'optional', which he defines as those appearing in fewer than 50% of articles in the corpus. While the texts he examines were older (pre-1990) and from only five journals, he establishes a reasoned set of Moves that compares well with my own findings, as discussed in 4.3 of this chapter.

4.1.5.3 Posteguillo

Posteguillo's (1999) work focuses on RAs in computer science. Instead of compiling a single corpus that would support his examination of individual sections in the Introduction, Methods, Results and Discussion (IMRD) framework, he develops three separate corpora of varying size for three of the sections: Introduction (40); Results (22) and Discussion/Conclusion (34) respectively. He does not examine the Methods sections, as he comments (1999, p.154) that computer science RAs typically do not have a separate Methods section. In addition, he notes that it is very hard to study the Moves in this type of RA since there seem to be no standards set forth in the genre, at least for the time period examined.

Posteguillo further argues that since computer science RAs do not fit neatly within the normalized IMRD structure, describing their Moves becomes even more difficult. I will draw on his insights that Moves can occasionally occur in unexpected locations within a text, and that not all texts will fit smoothly into categories, but will not discuss his study in detail, or include his work in tables or other comparisons of Moves later in the chapter as the structure of computer science articles (i.e. lacking a Methods section) and the topic (computer science) are too different from MRAs to allow a fair comparison.

4.1.5.4 Kanoksilapatham

Kanoksilapatham's (2005) corpus is comprised of articles from the field of biochemistry. This study was included in order to help differentiate possible differences from genres in similar disciplines. She selects 12 articles from five different journals published in 2000, thus taking the issue of time out of the discussion when she describes the Moves found in her corpus. Her study is the first in this group of researchers to apply inter-coder reliability tests in validating her findings. She and a second coder, a graduate student in biochemistry, located the Moves separately and measured how much agreement was found between the two sets of identifications, helping to reduce subjectivity after re-coding based on the findings of the inter-coder reliability tests. She was able to create a taxonomy of Moves with less subjectivity than previous studies and thereby offers a collection of Moves with which any collection of research articles can be compared.

4.1.5.5 Li and Ge

Li and Ge (2009) examine two corpora of MRAs in order to study the effect of time on Moves. They use the 11 Moves identified by Nwogu (1997). Their first corpus is made of articles appearing between 1985 and 1989 and the second comprises articles published between 2000 and 2004. As Paltridge comments (2013, p.348, drawing on Swales 2004), the 'communicative purpose of a genre … may evolve over time, it

may change, it may expand, or it may shrink'. Accordingly, the study by Li and Ge is highly useful, as their data allows us to see that Moves and their component Steps can change over time, and may offer an explanation as to why some of my findings differ from those in earlier studies. In tables throughout this study, their corpus using articles from 1985 to 1989 will be referred to as Corpus A and the corpus of articles published from 2000 to 2004 will be Corpus B.

4.1.5.6 Fryer

Fryer (2012) applies a Systemic-Functional Linguistic framework to identify Moves in MRAs. He narrows his focus to 16 articles selected from his larger corpus to comprise only MRAs published between 2004 and 2006 in five of the top medical journals (2012, p.7). He further restricts his scope to articles dealing with obesity. What differentiates his framework from that of the previous researchers is that he then uses his adaptation of Halliday and Matthiessen's (2004) constructs of theme and rheme in his categorization of the Moves. In his discussion, theme is controlled by the writer/speaker as being known or given information and the rheme is new information that the author wishes to present (Halliday & Matthiessen 2013, p.93). In choosing the theme, the author affects the presentation of the rheme.

4.1.6   Differences among Move-assignments

This section compares studies of texts collected before 2005, and focuses on how their authors classify Moves as optional as opposed to obligatory. The studies by Nwogu (1997), Li and Ge (2009) and Kanoksilapatham (2005) identify three Moves in the Introduction section, similar to Swales' (1990) CARS framework. Li and Ge (2009) adopt Nwogu's (1997) Moves, but they note several changes concerning Moves that Nwogu had previously marked as optional. Nwogu (1997) claims that Moves One (*Present background information*), Six (*Describe data-analysis procedures*) and Eight (*Indicate non-consistent observations*) are optional, as they appear in fewer than 50% of the articles in his corpus. In their corpus of texts published between 1985 and 1989, Li and Ge (2009), using the same benchmark of 50% for optionality, found the same

Moves and that they were still not mandatory. However, in their corpus of texts published between 2000 and 2004, using the same benchmarks, they found that Moves One and Six were no longer optional but obligatory, and, while Move Eight remained optional, they added a ninth optional Move (*Highlight overall research outcome*). This shift in optionality of Move-assignment may reinforce the commonly-held idea that genres undergo changes over time, and suggests that the time period in which the data are published or that the data represent needs to be included in the discussion of possible reasons for differences among studies.

While each of the authors follows Swales' model of identifying Moves, they do not all identify the same Moves. Skelton (1994), Nwogu (1997), Li and Ge (2009), and Fryer (2012) look at corpora of MRAs, while Kanoksilapatham (2005) examines RAs in a different field or discipline. A possible reason for differences among Move-assignments and whether the researchers consider the Moves to be optional or obligatory may be the different sizes of the corpora used or sampled by different researchers. In effect, the researchers have a threshold percentage to indicate whether a Move is obligatory or optional. The smaller the corpus, the more impact a Move has on the percentage, which may impact the findings of the smaller corpora when compared to the larger corpus (250 articles) that has been compiled for the current research.

The following table lists the Moves, but not the Steps that were found by the different researchers, and separates the Moves by their appearance in the different sections of IMRD. Such tabulation may be useful in showing similarities in frameworks for Moves as well as differences. It is important to note that Li and Ge (2009) apply their framework to two different corpora within a single paper. Moves held to be optional by the different authors are preceded by an asterisk. Skelton (1994) set the threshold of obligatory to optional at a frequency of 65% for a Move, while Nwogu (1997), Li and Ge (2009) saw the Move as optional if it occurred in less than 50% of the articles in their corpora. Kanoksilapatham (2005, p.272) selected 60%, which is the number this work will follow; she argued that in order to 'establish which Rhetorical Moves are more conventional than the others, the rather high cut-off frequency would be beneficial in enhancing the distinction between the two categories of Moves'.

Table 4.2: Moves frameworks adapted from Swales (1990) by authors examining full RAs/MRAs

Skelton (1994)	Nwogu (1997)	Kanoksilapatham (2005)	Li and Ge A (2009)	Li and Ge B (2009)	Fryer (2012)
INTRODUCTION					
Assert importance of the field of study Discuss previous literature Identify gap in the literature The writer's aim	*Present Background information Review related research Present new research	Announce importance of the study Prepare for the present study Introducing the present study	*Present background information Review related research Present new research	Present Background information Review related research Present new research	Present study background Identify gap(s) in existing research Statement of research purpose
METHODS					
Identify the population The procedures used Statistical tests to be used	Describe data-collection procedures Describe experimental procedures *Describe data-analysis procedures	Describe materials Describe experimental procedures *Detail equipment *Describe statistical procedures	Describe data-collection procedures Describe experimental procedures Describe data-analysis procedures	Describe data-collection procedures Describe experimental procedures Describe data-analysis procedures	Describe material/ participants Describe experimental procedures Describe data-analysis procedures
RESULTS					
*Adjust-ment and exclusions from original population Represent some part of results in tables Discuss data in words Assess the data	Indicate consistent observations *Indicate non-consistent observations	State procedures Justify procedures or methodology State results State comments on the results	Indicate consistent observations *Indicate non-consistent observations	Indicate consistent observations *Indicate non-consistent observations	Report observations
DISCUSSION					
State limitations and defend successes	Highlight overall research outcome	Contextualize the study Consolidate results	Highlight overall research outcome	*Highlight overall research outcome	Discuss main findings State study limitations

*The central achieve-ment of the study Contextual-ize researcher procedures/ findings Recom-mendations about what should or must happen	Explain specific research outcomes State research conclusions	State study limitations *Suggest further research	Explain specific research outcomes State research conclusions	Explain specific research outcomes State research conclusions	State conclusions

NB: *denotes Moves or Steps that the authors coded as being optional; Li and Ge A and B are parts of the same study

The studies vary in the number of Moves or Steps their authors identify as well as in how they see and name them. For example, while both Skelton and Kanoksilapatham each identify 15 Moves, their Moves are named quite differently, perhaps because they were using very different data sets. Kanoksilapatham (2005) identified 15 Moves in her sample of 60 biochemistry texts, 12 each from five top biochemistry journals ranked by impact factor and categorized three of the Moves as optional; Skelton found two Moves optional in his study of articles from a single journal. In the Methods and Discussion sections, Kanoksilapatham (2005) described four Moves while Nwogu (1997) and Li and Ge (2009) found three; in the Results section, Nwogu identified only two Moves while Kanoksilapatham had four. However, Nwogu and Kanoksilapatham share other components of their Moves framework. Nwogu's fourth Move, *Describe data-collection procedures*, includes what he calls 'constituent elements' or sub-Moves, what I call 'Steps' in this work, that include '(A) *Indicate source of data*, (B*) Indicate sample size* and (C) *Indicate criteria for data selection'.* (1997, p.128) These are similar to Kanoksilapatham's (2005, p.277) fourth Move, *Describe materials*, which is comprised of three Steps:

> Move 4 can be realized as Step 1: Listing materials explicitly itemizing materials or substances used in the study, Step 2: Detailing the source of the materials identifying how these items are obtained, such as, by purchase, as a gift, etc., and Step 3: Providing the background of the materials including the description, properties, or characteristics of the materials.

In biochemistry articles as reported by Kanoksilapatham (2005), materials seem to refer to subjects or data being tested in MRAs and not to materials that are being used in an experiment. Both of the Moves, however, revolve around what is being tested and where the subjects were procured or found.

Nwogu's fifth Move, *Describe experimental procedures,* has the same name as Kanoksilapatham's fifth Move, but there are subtle differences in the constituent elements. Both frameworks offer evidence as to what constitutes a Move and how its Steps need to be fulfilled. Nwogu (1997, pp.129-130) states that his fifth Move may be comprised of the following: '(A) *Identification of main research apparatus,* (B) *Recounting experimental process,* and (C) *Indicating criteria for success*', while the fifth Move in Kanoksilapatham (2005, p.278) comprises these Steps: '*Documenting established procedures, Detailing procedures* and *Providing the background of the procedures*'. Their first Steps are similar; however, Nwogu's second Step maps onto Kanoksilapatham's second and third Steps. Nwogu's third Step is not included in Kanoksilapatham's fifth Move. This is in part due to the different fields being studied, which could suggest that members of the discourse community of biochemists in 2000 may not have needed to express how they hoped to measure success, while the community of medical professionals needs to set targets for others to follow.

Examining these studies highlights five issues with describing Rhetorical Moves: time, field, size, subjectivity and the possibility of outliers, or anomalies, in a corpus (Guthrie, Guthrie & Wilks 2008). Li and Ge (2009) show that *time* does seem to affect Moves, insomuch as what was once considered mandatory can become optional or unnecessary. The differences between Nwogu and Kanoksilapatham's studies suggest that different *fields* or areas of study may have different organizational needs and that not all RAs are the same, which is further supported by Posteguillo (1999). As for the issue of corpus *size,* it is increasingly difficult to make any sort of representative claims about such small corpora, because if the threshold of determining if a Move is optional versus obligatory is set at 60%, a small corpus could be skewed. *Subjectivity* is nearly impossible to avoid in a study such as these. Kanoksilapatham (2005) and Sheldon (2013) both attempt to mitigate the effects of subjectivity, through using inter-coder reliability measurement on their samples. *Outliers* (a term borrowed from statistical analysis) occur when an author puts a Move

that normally appears in one section into another section. For example, I found in Text 17 in my corpus an example of a Move that I would later describe as being part of the Methods section (*A statement of ethical approval*) but it occurred in the Introduction section. Posteguillo (1999) finds numerous instances where a Move that occurred in one section of the text might appear in another section of a different text, and suggests that there cannot be a one-size-fits-all approach to identifying the Moves in RAs.

4.1.7   The rhetoric in Rhetorical Moves

The previous five studies all list Swales' (1990) CARS framework as a major influence on their work. When Swales described the Moves, he called them *Rhetorical Moves*. He also describes how the CARS framework could be seen in rhetorical terms as being a problem-solution model (1990, p.138). None of the five researchers attempts to define their Moves explicitly by the possible rhetorical purpose for each Move, a lack I address.

Aristotle, the father of rhetoric, describes the three main types of rhetorical appeal as: Pathos (emotions), Ethos (credibility) and Logos (logic). As explained by computational linguists, Teufel and Moens (2002, p.412), most scientific work will not have Pathos in the strictest sense (although it will appeal to various social purposes including persuasion), so that in their study of the types of appeals found in scientific writing, they separated the rhetorical components of scientific writing into three other categories. In their study, which is widely cited by scholars in rhetoric (such as Taboada & Mann 2006, pp.572-573) and applied linguistics (White 2004) as well as in computational linguistics, they define the three categories as follows:

> Rhetorical status in terms of problem solving: What is the goal and contribution of the paper? This type of information is often marked by metadiscourse and by conventional patterns of presentation
> Rhetorical status in terms of intellectual attribution: What information is claimed to be new, and which statements describe other work? This type of information can be recognized by following the 'agent structure' of text, that is, by looking at all grammatical subjects occurring in sequence

Relatedness among articles: What articles is this work similar to, and in what respect? This type of information can be found by examining fixed indicator phrases like *in contrast to . . .* section headers, and citations (p.412).

In this study, I have chosen to apply a number of Teufel and Moens' terms to different structures to help classify the Moves and Steps in the information flow of the MRA: they include 'problem-solution, intellectual attributions, scientific argumentation and attitude towards other people's work' (2002, p.413). I will give brief definitions for each in turn.

Problem-solution can appear, according to Teufel and Moens (2002, p.413), in three ways in a research paper: through problems, i.e. the gap or goals; through solutions, which are covered by the methods; and through results, which act as the final answer to the problem. Van Dijk (1980) discusses the structure of scientific papers as comprising what he calls 'metastructures', which is the basis for Teufel and Moens' schema. Van Dijk's framework incorporates:

(a) the necessity of ordering, organizing information units (propositions), by functional categories;
(b) the semantic-pragmatic constraints on information distribution (e.g. presupposition-assertion articulation);
(c) the pragmatic aspects of the discourse: what is the general point or function of the discourse as an act for the hearer/reader;
(d) the interactional aspects of speaking, writing, and reading/hearing: how to start; how to stop; relations between speech participants; etc.;
(e) cognitive aspects: as previously for information acquisition, plus needs for expression, evaluation, news, problem solving, suspense, etc. (van Dijk 1980, p.112).

Authors need to persuade their audience that their work has merit, in that the work can benefit the field. The structure of *intellectual attribution* is met by the fact that authors need do two things to make their work stand out. First, authors must show how they have used existing works by others as a foundation from which they build and the second is their comparison to the work of others, cited or discussed, either negatively or positively, so that authors are then able to help their research stand apart, including when their findings support previous work.

Scientific argumentation aligns the most closely to Aristotle's idea of Ethos, as it is

the structure that scientific authors use to convince the reader that their work is valid and beyond reproach. Authors present a claim and proceed, in logical steps, to justify it by presenting evidence or other findings. An author needs to convince the audience that their science is solid and that their work is of quality, which will suggest that their results must also be correct. In a personal interview Doctor 6 said that when he is asked to be a reviewer, the first section he looks at is the Methods: he claims that by looking at their Methods section he can quickly determine whether the authors had conducted a valid experiment worthy of publication.

Attitude toward other people's work is a structure that authors use when passing judgment on previous studies and can be closely aligned with intellectual attribution.

> In the flow of the argument, each piece of other work is mentioned for a specific reason: it is portrayed as a rival approach, as a prior approach with a fault, or as an approach contributing parts of the authors' own solution. (Teufel and Moens 2002, p.413)

It is this idea, that every bit of information reported from other sources has a specific purpose, that lays the groundwork for my hypothesis that in MRAs, reporting verbs (see 5.1.2) may not occur in a neutral state, and are used either to support or negate the current research being discussed.

Originally, the framework established by Teufel and Moens was developed to examine every sentence in an article; however, a Move (or Step) may be composed of a series of sentences linked by various cohesive devices. In this study, I will act as if sentences forming all or a part of a specific Move will in some way contribute to the rhetorical function of that Move or Step. Indeed, the rhetorical principle underlying a Move or a Step may act as a primer for word choice, as discussed in more detail in Chapter 5.

4.2   Methodology

A methodology for identifying Moves and their components needs to incorporate the five issues of time, field, size, subjectivity and outliers as defined in 4.1.6, as any of

these could influence the validity of researchers' findings. To address the issue of *time,* my corpus, as described in 2.2.1, is comprised of 250 articles published from 2000 to 2010. Li and Ge (2009, p.102) make the argument that 'genre has an evolutionary nature and medical RAs have undergone some significant structural and linguistic changes over the past two decades'. While articles studied by both Li and Ge (2009) and Fryer (2007; 2012) overlap in time with a portion of the articles in my corpus, my collection is the most current of all the works being compared in the studies drawn on here, and as such may offer the most current description of Moves found in MRAs.

This work has utilized the specialised corpus of 250 MRAs, thus allowing for a focus on one *field.* However, as will be discussed in 4.3, I have broken down my defining Moves and their Steps according to whether they are obligatory or optional, based on the subjects being tested upon in each individual article (see 4.3.2).

As for the issue of *size* and the impact of offering findings as percentages to show representativeness, my corpus is the largest in comparison to the other studies reviewed in this chapter, as I have included and analysed each of the 250 texts in my corpus.

Subjectivity cannot be avoided: to make my research as valid as possible, I follow Nwogu's (1997) study and offer clear linguistic and content-based examples for each of the Moves I present. As Crookes (1986, p.66) warned, a non-specialist in the hard sciences may miss words or clues from lacking a background in those fields. For unfamiliar medical terminology or tests, I have consulted standard medical dictionaries such as Mosby's *Dictionary of medicine, nursing, and health professions* (2012) and a number of practitioners.[15] For clarifying advanced statistical tests or terminology, I also consulted a specialist.[16]

Outliers, as defined in 4.1.6, are impossible to allow for, as their appearance is

[15] Ashlie Darr, MD (ENT, Clinical Instructor, Department of Otology and Laryngology, Harvard Medical School, USA), Ryan Caldwell, MD (Internist, Carolinas Medical Center, USA) and Sara Campbell, PhD (Virologist, Centre for Drug Research and Development, Vancouver, Canada)
[16] Michael McEwan, PhD (Mathematics and Statistics, University of Glasgow, UK)

random. Posteguillo's study (1999) was instrumental in helping me to understand that some Moves may appear in different places in different articles, and that my framework should not try to impose a top-down, pre-formed rigid structuring. While the Moves are sequentially numbered within sections for convenience of presentation, they do not have to appear in that order in a text and sometimes may not appear at all. Since this study looks at the whole article as opposed to any one section, my framework is flexible in order to respond to the possibility of outliers.

The following is a brief overview of the process of Move identification I used in this study. After reviewing Swales (1990), I began by using Nwogu's (1997) list of 11 Moves as the basis from which to identify Moves, adjusting the list as more Moves or Steps became apparent. Following the precedent set by Kanoksilapatham (2005), I established criteria for a Move to be deemed obligatory and therefore for this study an obligatory Move needs to occur in over 60% of the articles in my corpus. After an initial read-through of the articles, I compiled all of the Moves and Steps that were different from Nwogu's and then re-read the articles to ensure that no Move had been omitted. However, in analysing my initial tabulation, I noted that certain Moves were obligatory for some types of texts and optional for others, based on the characteristics of the types of subjects (human, animal, cell, etc.) in the study. For example all articles that had human subjects had a Move of announcing that an ethical protocol was followed, whilst articles that looked at past data did not have this Move. I then performed Chi-square tests on the results of Moves based on types of subjects to validate my findings; this is described in more detail in 4.3.2.

4.3   Moves within Medical Research Articles

There are three nested layers of organization within an MRA, the first being the IMRD (Introduction Methods Results Discussion) structure. This is followed by a layer of Moves, which can be comprised of a third layer, and may include the Steps. After reading the 250 texts and using Nwogu's (1997) 11 Moves as a preliminary basis, I have identified a total of 17 Moves with four of them being deemed optional. There are 3 Moves that typically appear in the Introduction section; 3 in the Methods section; 4 Moves in the Results section; and 7 Moves in the Discussion section. Steps

are defined as units of information that help fulfil the Move. For example, in my first Move, *Study situation*, two Steps can fulfil the Move: (A) a *Framing statement* and (B) *background on the problem/field*. A separate table illustrating Moves and Steps identified by each of the researchers previously discussed heads each section below: Table 4.3 presents Moves and Steps in the Introduction sections; Table 4.5 displays those for the Methods sections; Table 4.9 presents Moves and Steps for Results sections; and Table 4.11 illustrates them for the Discussion sections.

4.3.1   Examples of each Move

In this section I will describe the Moves and their possible Steps as identified throughout the 250 articles in my corpus. I will attempt to explain the primary rhetorical purpose as well as to present linguistic characteristics of each Move. I will also identify the percentage of times the Moves and Steps occur in the entirety of my corpus (i.e. their frequency). Then, in 4.3.2, I will investigate the correlation between Moves and subject matter, as I have found that the nature of the subject has an impact on whether or not the Move or Step appears.

4.3.1.1 Moves and Steps in the Introduction section

Table 4.3: Moves and Steps in the Introduction section

Introduction section						
Skelton (1994)	**Nwogu (1997)**	**Kanoksilapatham (2005)**	**Li and Ge A (2009)**	**Li and Ge B (2009)**	**Fryer (2012)**	**Present study**
1. Importance of the field of study 2. Discussion of previous literature 3. Gap in the literature 4. The writer's aim	*1. Present background information 2. Review related research A. Refer to previous research B. Refer to limitations of previous research 3. Present new	1. Announce study importance A. Claim centrality B. Topic generalizations C. Review previous research 2. Prepare for the present study 3. Introduce the present study	*1. Present Background information 2. Review related research A. Refer to previous research B. Refer to limitations of previous research 3. Present new	1. Present Background Information 2. Review related research A. Refer to previous research B. Refer to limitations of previous research 3. Present new	1. Present study back-ground A. Establish-ed knowledge 2. Identify gap(s) in existing research A. Lack of data in specific area related to established field;	1. Study situation A. Framing statement B. Back-ground work by others 2. Describe the problem A. Identify gap B. Present hypothesis 3. Way to

	research		research	research	B. Reason for need to fill gap 3. Statement of research purpose A. Hypothesis or objective; B. Describe material/ Methodology	the solution A. Outline the procedures B. Describe subjects C. State aims of the work D. Preview the findings

NB: *denotes Moves that the authors coded as optional; Li and Ge A and B are parts of the same study

4.3.1.1.1 Move One: Study situation

In the current research study, Move One comprises two Steps: (A) a framing statement, and (B) background work by others. This Move is used to establish the primary subject area of the problem that the article will discuss. Taking the topic of cancer as an example, it would be very difficult for authors to present research on all types of cancer in a single research article. There are numerous types of cancer as well as multiple causes and symptoms, which would be far too much for an individual or even a team to present and research all aspects; instead, they may focus on one specific area under the umbrella of cancer. It is in their first Move that they establish the overarching theme or focus of the article.

Both Steps, an opening statement and a background overview of previous research or of the history of the theme being written about, are used to realize this initial Move. The opening statement usually presents a claim that is fairly neutral and informative, while the background overview highlights previous research or the history of the issue. This can be seen as an appeal to the readers' sense of logic (Logos), as before a topic can be discussed, it must first be defined. Following work on research articles by Vološinov (1973), White (2003), Martin and White (2005), and Fryer (2012), I agree that academic publication, of which MRAs are a part, is actually a discussion or conversation between the author and the audience. In this discussion, the authors must attempt to answer questions that they predict the reader may have. The author also has to set the stage for the conversation and this is done with the first Move. The following extracts exemplify the first Step of Move One in the opening statement:

(1) Conception of health has been changed with the development of medicine and medical sciences since the 1970's. Health is defined as a dynamic state of human wellbeing characterized by a physical, mental, and social potential which satisfies the demands of a life corresponding to age, culture, and personal responsibility, and not merely the absence of disease or infirmity. (Text 17)

(2) BMPFD is a progressive autosomal dominant disorder, characterized by the adult onset of muscle degeneration, abnormal bone metabolism, and drastic behavior changes. (Text 180)

This Step seems to be used as a framing agent to help set the parameters of the article: as in a debate, if the audience can accept this initial claim, then they might accept a greater amount of the paper as being true. The tenses used are the present simple or the present perfect, as if this is a truth that has been recognized. It is also usually written in active voice, but can appear in the passive (see Example [1]). The Step uses verbs like *define* or *characterize*, which are verbs used to describe and explain specific nouns that carry the content load.

Skelton's (1994, p.456) first Move was *Importance of the field of study,* which corresponds to the first Step in my first Move, which in its framing must identify such importance. Skelton's second Move *Discussion of previous literature* maps onto *Background work by others,* the second Step in my first Move. I consider his two separate Moves as being parts of a single Move in my framework, because they share a common rhetorical function combining *intellectual attribution* and *scientific argumentation*. Additionally, Skelton comments that in his study, his Moves can be hard to separate, adding 'A minor variation here was to assert significance by reference to another writer, thus combining the first and second Move in the same sentence' (Skelton 1994, p.457). In relation to characteristic lexis in this Move, Nwogu adds that he finds usage of temporal and locative adverbials (1997, p.126), which also appear frequently, though not exclusively, in the opening statements in my corpus.

The second Step in Move One (see 4.1.3 for a definition of Steps vis-à-vis Moves) in my framework is *Background work by others*. In this Step, the author may be appealing to the reader's sense of Ethos and Logos. In order for a writer to gain

credibility as an expert he/she must demonstrate knowledge of the topic and should logically allude to the specific topic or problem before addressing it fully. An example is the allusion to recent evidence in the example below:

> (3) Recent evidence demonstrates that the 14-3-3-gene promoter region is un-methylated in normal epithelial cells while inactivated via hypermethylation of its 5' CpG islands in epithelial malignancies. (Text 9)

This Step provides information on how others have previously examined the overall subject and, like the first Step, is typically written in active voice and the present simple. It usually contains lexis such as *previous studies/research*, as well as footnotes, and citation of named authors. However, possibly due to the subjective nature of Move coding, Nwogu (1997) does not state that Move One is covered in two Steps; instead, he sees the opening Move as one that is completely focused on background, which could diminish the persuasive qualities in an opening framing statement, as a niche for the current author(s) is not suggested. Such a two-Step sequence is not uncommon at the outset of expository writing. For example, in her work with second-language writers, Reid (1996) reports that many writers have trouble completing the rhetorical requirements of an opening sentence because they cannot manipulate the sentence that follows. It is in the second sentence that the author must develop initial support of the topic sentence. Although Reid (1996) focused on shorter writings that are not at the level of professional MRAs, the need to focus on the second sentence for development can be seen as mirroring my first two Steps in Move One.

In her study, Kanoksilapatham (2005) separates her division of Move One into three Steps, the first being an opening statement that claims centrality; the second, a Step of making topic generalizations; and a third, a review of previous research. Step A in my construct of Move One combines her first two Steps. In the MRA corpus, I find that topic generalizations can appear in either of the two Steps that form Move One. In the opening statement, generalizations can become a way to set forth a proposition for the reader to consider or with previous research to generalize for the author, as exemplified above in Excerpt 3. It is in their statement of generalization that the authors initiate an effort to persuade readers to believe them and to see their work as credible, not overly controversial and based appropriately on work done by others.

Fryer (2012, p.9) also finds that the first Move is one of relaying the background of the situation to the audience, and adds that it is this Move that helps the author state the setting of the paper. He explains that this Move covers generally held knowledge that the author assumes is readily available to any reader. However, I would add that this Move, identified in both of our studies, may be affected in the future as the Internet is providing lay readers with greater access to MRAs, as mentioned by interviewees in my sample of the discourse community (see 3.3.6)

Applying Teufel and Moens' (2002) framework, I argue that the authors seek to persuade the readers by using two of the qualities of scientific rhetoric: *intellectual attribution* and *scientific argumentation*. The authors frame the argument of the paper and introduce the reader to their authorial bias or orientation, which helps to make their work appear both valid and different from that of others. It is important to note that both of these Steps appear 100% of the time in my corpus, suggesting these Steps are linked together and appear regardless of subject matter for the research.

4.3.1.1.2 Move Two: Describe the problem

Move Two in my framework mirrors Swales' second Move, of *Establishing a niche* (1990, p.142); however, in Swales' work, this Move is comprised of four Steps, while my construct has only two: the gap and the hypothesis. The principle behind this Move is embedded in its title: *Describe the problem*, which is typically either the existence of a deficiency in the field or a specific medical problem. Since, as Swales (1990, p.138) argued, most articles seem to follow a problem-solution pattern, this Move initiates that pattern. The first Step in this Move is *Stating a gap that needs to be filled:* the terminological phrase comes directly from Swales and is apt for this Step. An example of locating the gap in the field is:

> (4) However, there were significant shortcomings to these studies: they were conducted in healthy middle aged patients; they used a before and after study design that did not allow comparison of the video to the standard advance care planning approach of a verbal narrative; they did not measure knowledge of the disease to test whether understanding of the disease improved; and they

did not follow patients' preferences over time. (Text 76)

The gap noted in this example is located in previous research, and the implication in the statement is that those limitations, or 'significant shortcomings', listed in the Move will be addressed in the present research. The following example illustrates a different kind of problem, one that needs to be fixed in the larger society, to which the findings may be beneficially applied:

> (5) Because minority populations and patients facing socioeconomic barriers to health care access have been shown to have the worst adherence to medications and poor glycemic control [14], determining the association between ethnicity and adherence to oral hypoglycemic agents among Japanese, Chinese, Filipino, Hawaiian, and white patients in Hawaii will help reveal the ethnic disparities that exist in Hawaii and identify those groups who most need to be targeted for intervention. (Text 176)

Here, the gap that needs to be filled is one of knowledge that is not presently available to doctors; with that knowledge, they may better address a negative situation in society. The following example presents a problem of medical technique that needs to be addressed:

> (6) Knowledge of vascular reactivity in patients undergoing cardiological evaluation for a preventive purpose (check-up) relative to demographic and laboratory variables may be a question of everyday practice for preventive s and poor glycemic control (14), deform this investigation. (Text 228)

In Text 228, the problem lies in a specific procedure that needs to occur in order to help surgeons perform surgery. The following exemplifies a problem that is both a lack of a satisfactory technique and a shortcoming in the field:

> (7) A major problem to date has been an ecologically meaningful measure of the response to stressors and how often they occur in nature. Blood levels of glucocorticoids, by virtue of their instantaneous sample and logistical limitations of capturing birds, cannot encapsulate effects of a temporally dynamic environment. (Text 201)

The implication here is that the authors are proposing a solution to this problem, which will be covered in the following Move. This Step occurs 97% of the time in the MRA corpus, meaning that it is apparently an obligatory part of a research article. As

to why the Step does not occur at a level of 100%, it may be that, according to Crookes (1986), Kanoksilapatham (2005), Nwogu (1997), Posteguillo (1999), and Swales (1990), it is very difficult to find a Move one hundred per cent of the time. The seven articles out of the 250 in my corpus that did not have this Move treated the opening statement as the problem to be addressed, suggesting that the authors were in that sense describing the *Study situation*.

This Step is generally written in the present simple tense when the problem to be examined is a current one, or in the present perfect tense when discussing the failings in previous studies. Key words that may alert the reader are: *needs, shortcomings, however, determining, discovering, reason*, and *problem*, all of which indicate a need for an answer to a problem or a problem with previous work in the field, or with the field in general.

As seen in Table 4.3 above, Skelton's (1994) Move Three, *Gap in the literature*, matches Step A in my construction of Move Two while both Nwogu (1997) and Li and Ge (2009) divide what I call Step B (*Present hypothesis*) across two separate Moves in their categorization. This is because they consider disagreement with the field to be part of the review of previous work, and when the stated problem needs a new technique for its solution, they consider that need to be a Move for new research. Kanoksilapatham's second Move, *Prepare for the present study*, has two Steps. Step A is *Indicate a gap,* which is matched by the first Step in my construct of Move Two. However, her Step B is *Raising a question*, while my Step B states a hypothesis. My Step B appears in 66 of the articles (26.4%) when there is an explicit statement of a hypothesized result, as seen in:

> (8) We predicted that imagery with re-scripting would reduce the strength of patients' negative self-beliefs and their anxiety about feared social situations. We also anticipated a reduction in the frequency, vividness, and distress of patients' recurrent images. By contrast, we predicted that simply exploring the memories would not be beneficial. (Text 164)

Example (8) shows the authors stating what they originally had predicted because of their experiment or intervention. Interestingly, it is written in the past simple tense, to signal what they had presumed earlier but now know to be different. Key words are:

hypothesized, predicted, and *anticipated*, all indicating a proposed result. None of the other authors noted this Step in any of their Moves in the Introduction section, although Kanoksilapatham identified such a statement as a Step in one of her Moves in the Results section and Posteguillo located it in the Discussion section. Possible reasons for the locating of similar Steps/Moves by researchers in different sections in their studies may be in part due to the effect of field on the genre, meaning they were looking at research articles from different disciplines. Perhaps this Step did not appear in Nwogu (1997) and Li and Ge's (2009) works because their corpora may have been too small or the Step has become prevalent in the genre more recently than their corpora covered, as 42 of the 66 articles in which such a statement appeared in my corpus were published after 2007.

This Move and its two Steps serve the rhetorical purpose of introducing the problem, and as such I would categorize this Move as incorporating all four of Teufel and Moens' (2002, p.413) different rhetorical structures: *problem-solution*, as it is the problem; *intellectual attribution* as authors sometimes find fault with previous works that they will strive to rectify; *scientific argumentation*, as they are presenting a problem and may hypothesize what the results may be; and *attitude toward other people's work*, which is seen when the authors state what the previous studies lacked. The authors have set the stage and we are now focused on a specific problem; the writers can logically follow with the solution, which becomes the third Move of the Introduction section.

4.3.1.1.3 Move Three: Way to the solution

If authors describe a situation and note that it has a problem, they are announcing their intent to find a solution, which is the rationale for my establishment of a third Move in the Introduction section: *Way to the solution*. This Move is comprised of four possible Steps: *Outline the procedures; Describe the subjects; State the aims* (typically, of the experiment); and *Preview the findings*. In Swales' CARS model (1990, p.142), this Move is represented by what he calls *occupying the niche* which, he argues, is the way authors attempt to fill the niche created in the previous Step. Nwogu claims (1997, p.128) that the main purpose behind his third Move is

introducing the reader to the underlying purpose for the current research and accordingly names his third Move, *Present new research*. The Steps in Nwogu's third Move are similar to mine, but I disagree with his primary rhetorical purpose for this Move. I have previously argued in Move One that MRAs follow a problem-solution model, with the problem identified in Move Two. I have found that its solution is Move Three, *Way to the solution*. This Move in my framework has four Steps. The first Step is *Outline the procedures*; an example of this is:

> (9) In this research, we evaluated the biocompatibility and stability of self-assembled Fe3O4-MNPs/DNR by hemolysis testing, micronucleus assay, and detection of median lethal dose (LD50). (Text 154)

In Example (9), the authors state their procedures, which, on the surface, is a Step that would seem to belong exclusively to a part of the Methods section. However, according to a personal interview (Doctor 2, March 20th, 2012) from the discourse community sample, many readers jump around in their reading, thus suggesting the usefulness of the repetition of similar material in different sections, which suggests that writers, from their own experience in reading articles, may consider repetition to be helpful. It is interesting that the handout for Sciences at the University of North Carolina (Chapel Hill) Writing Center (n.d.) suggests that repetition may be needed to avoid ambiguity and to increase precision.[17] This Step is written in the simple past tense and is presented in the active voice, which differentiates it from a description of procedures in the Methods section (see Move Six). It is written in the simple past tense because it describes what the authors have already done, and its use of active instead of passive voice may be in part due to the fact that the authors wish to stress that they are experts, which helps bolster their credibility, or Ethos. Typical lexical features are the use of *current* in reference to research and verbs often found in describing the process of an experiment, such as *evaluated, conducted, examined*, and *analysed*. This Step appears in 199 of the 250 articles in the corpus, or 79.6%, thus indicating that the Step is seen as obligatory, by this definition.

Describe the subjects is a Step in which the authors state who or what will receive treatment. All of the other studies of full texts, with the exception of Skelton (1994),

---

[17] http://writingcenter.unc.edu/handouts/sciences/

mention a similar Step. Skelton does not mention such a Step but in the example he gives for his fourth Move (*Writer's aim*) he does describe *who* is being studied, in the context of what I call my third Step, *State the aims*. While there is an entire Move in the Methods section devoted to describing the subjects and bio-data, a brief highlight or preview of the information, nevertheless, does often appear in the Introduction section as a Step, as follows:

> (10) In this study, we used a multi-disciplinary approach to investigate a select group of highly conserved lncRNAs that are expressed within the embryonic and early postnatal mouse brain. (Text 240)

From Example (10) it can be seen that this Step, *Describe the subjects,* can be combined with the previous Step of *Outline the procedures*. Since the Step is describing who or what was treated, it is written in the past tense, and can be written in either the active or the passive voice. An example illustrating it as a stand-alone Step is:

> (11) The C3(1) SV40 T-antigen transgenic mouse was selected as the model for this study. This mouse is a well characterized model developed in the laboratory of Dr. Jeffrey E. Green. (Text 233)

Example (11) is written in the passive voice, a stylistic choice discussed below in 4.3.1.2. Key words that signal the Step are: *select, this study, group, examined*, and *analysed.* Move Three includes Steps describing the procedures and the treatment, and the next Step presents the rationale for performing the experiment in the chosen way on specific subjects.

In only 71 of the 250 articles in the present MRA corpus (28.4%) is there a Step that explicitly states the Aims of the study. This Step is where the authors report their reasoning behind and purpose(s) for the research. All five of the other studies of Rhetorical Moves mention this Step, though Skelton is the only one who treats it as a full and separate stand-alone Move. I have chosen to treat it as a Step because its function is to report why the authors are attempting to find the solution, as illustrated by:

> (12) The purpose of this study was to determine the rate of malignancy

following microdochectomy or subareolar exploration. (Text 105)

The Step illustrated in Example (12) above can also use the present simple tense, as italicized:

> (13) This study *aims* to compare maternal PlGF and sFlt-1 levels in GH mothers periodically (antepartum, intrapartum and postpartum) and later correlate these biomarker levels with placental morphometry. (Text 005)

Since this Step is an explicit statement of the aims, it characteristically presents specific lexemes, such as: *aim, purpose, reason,* and *objective*. This Step is always written in the active voice throughout the whole corpus, as the focus is on the agent. This focus is similar to the last and least frequent Step observed in the Move, one of stating or summarizing the findings.

The optional Step, *Preview the findings,* is peculiar for a number of reasons. Only Kanoksilapatham (2005) mentions occasionally finding a similar Step in her corpus, and it appears in my corpus only 66 times out of the 250 articles (26.4%). In addition to its occurring sparingly, this Step has the added drawback of seeming to be in the wrong place as the Introduction section is primarily focused on familiarizing the audience with the field, the problem, and the experiment designed to find a solution. However, when examined through a rhetorical lens, that of setting the work apart from studies by other researchers, this Step may be used by authors to show a difference between their and others' work. An example of this Step is:

> (14) Our results show that, although the changes in the vaccination program have reduced pertussis morbidity in childhood, they have not affected the increased infection rate in adolescent and adult pertussis. (Text 250)

It is written in active voice in the present simple as if it were a fact not to be disputed. This structuring could also be helpful when presenting particularly controversial findings, as a way of alerting the reader that first, these are the results and next, the authors will show just how they drew their conclusion. Once again, the Step is fairly explicit and has certain key words and phrases to alert the reader to its presence, such as *findings* and *results*; the Step is connected to the authors with the use of possessive adjectives and pronouns like *our* and *we*.

The following table shows the percentages of each Step in my framework as they occur in the texts in the corpus. The three Moves that appear in almost every article are assumed obligatory, and expected by the discourse community, although some of the Steps seem optional. A Move does not have to include every Step to be fulfilled, although it must have at least one of the Steps. For example, Move Two, *Describe the problem,* can comprise two Steps, one for describing the gap and one for expressing the authors' hypothesis. The Step regarding the gap appears in 244 of the articles in the corpus of MRAs, while the statement of the author's hypothesis only occurs in 66 texts. Move Two happened in 100% (250) of the articles, which means that the six MRAs presenting no description of a gap presented instead a statement of hypothesis.

Table 4.4: Steps in the three different Moves in the Introduction section

	Study situation (100%)		Describe the problem (100%)		Way to the solution (100%)			
Steps Observed	Statement	Background	Gap	Hypothesis	Outline procedures	Subjects	Aims	Findings
Number observed	250	250	244	66	199	239	71	66
Total percentage	100	100	97.6	26.4	79.6	95.6	28.4	26.4

The table illustrates that both Steps in Move One, *Study situation*, are obligatory, while only the Step of *Describe the gap* in Move Two is obligatory. In the third Move of the Introduction, the only Steps that are mandatory are those describing the subjects being treated and, largely, the brief outline of the procedure. It is worth reiterating that a possible reason why some of my Steps were not found in Nwogu's or Skelton's works is because their corpora are much older: since the time when their corpora were amassed, the expectations of the genre have shifted. An additional reason for the differences in the identification of Moves and Steps is posited in Posteguillo's (1999) study: since there is a lack of a central governing body to control all aspects of publication, writers are faced with a situation in which both authors and journals can apply a multiplicity of techniques and requirements.

I predict that the Step of explicitly stating the *Aims of the work* will become more common in the future, as it helps the reader understand the intended purpose more explicitly and should be useful in guiding NNS in following discourse patterns that may differ from the ones they have learned in other contexts. The Introduction section

has two primary rhetorical goals: that of convincing the reader of the credibility of the authors, and of helping the authors of a particular work to stand apart from others writing on a similar topic. Explicit statements of aims can address both goals.

4.3.1.2 Moves and Steps in the Methods section

Table 4.5: Moves and Steps in the Methods sections

Methods section						
Skelton (1994)	**Nwogu (1997)**	**Kanoksilapatham (2005)**	**Li and Ge A (2009)**	**Li and Ge B (2009)**	**Fryer (2012)**	**Present study**
5. Describe population to be studied

6. Describe procedures used

7. Statistical tests to be used | 4. Describe data-collection procedures: A. Indicates source of data B. Indicates data size. C. Indicates criteria for data collection

5. Describe experimental procedures: A. Identify main research apparatus B. Recount experimental process C. Indicate criteria for success

*6. Describe data analysis procedures: A. Defines terminology B. Indicate process of data classification C. Identify analytical instrument/pr ocedures D. Indicate modification to instrument/ procedures | 4. Describe materials A. Listing materials B. Detailing the source of the materials C. Providing background of the materials

5. Describe experimental procedures A. Document established procedures B. Detail the procedures C. Provide background of the procedures

*6. Detail equipment

7. Describe statistical procedures | 4. Describe data-collection procedure: A. Indicates source of data B. Indicates data size C. Indicates criteria for data collection

5. Describe experimental procedures: A. Identify main research apparatus B. Recount experimental process C. Indicate criteria for success

*6. Describe data-analysis procedures A. Define terminologies B. Indicate process of data classification C. Identify analytical instrument/pr ocedure D. Indicates modification to instrument/ procedures | 4. Describe Data-collection procedure: A. Indicates source of data B. Indicates data size C. Indicates criteria for data collection

5. Describe experimental procedures: A. Identify main research apparatus B. Recount experimental process C. Indicate criteria for success

6. Describe data-analysis procedures: A. Define terminologies B. Indicate process of data classification C. Identify analytical instrument/ procedure D. Indicates modification to instrument/ procedures | 4. Describe material/ participants A. Size of study sample B. Study period C. Selection criteria D. Type of data collection E. Frequency of data collection F. Approval /informed consent

5. Describe experimental procedure A. Take measurements B. Define terms C. Sample categorization D. Endpoints/ outcomes

6. Describe data-analysis procedure A. Statistical test techniques B. Software | *4. Ethics statement

5. Describe what is treated A. Location B. Bio-data C. Ration-ale

6. Describe experiment A. Materials B. Proced-ures C. Tests D. Statist-ical tests |

* NB: denotes Moves or Steps that the authors see as being optional; Li and Ge A and B are parts of the same study.

The Methods section of a text is where the authors report the materials and tests used in order to answer their research questions. The priority of the section is replicability, the assumption that any other person could copy the Steps an author took, using the same materials, and derive the same results. Azevedo et al. (2011) claim this section to be the most important as it is the section where readers can judge how worthy the paper or work is by the excellence of the science. Interestingly, while the idea of the Methods section being the most important may be true, according to my interviews with members of the discourse community about reading habits, (see 3.3.4), this section may be the least read. However, Azevedo et al. (2011) also underscore an important aspect of the Methods section, which is its emphasis on validity. In short, they focus on the Ethos of the study, seeing the Methods section as fulfilling the rhetorical principle of convincing the audience that the speaker/author is reliable and trustworthy. Ethos, in Aristotle's sense, goes beyond providing an Ethics statement; it is the backbone associated with the Moves in this section, because the reader is dependent on the Ethos of the authors to provide true and trustworthy details of how the data (as materials, experimental design or subjects) are located, acquired, or manipulated.

As the Methods section of an MRA describes what the authors either did or used, it is usually written in the past tense. Another grammatical issue of importance to this section is that it generally is written in the passive voice. This continues the trend noted by Thompson (1993) for biochemistry articles and further described by Swales and Feak (2004, p.158), and adapted here:

Table 4.6: Use of tense and voice in IMRD sections of research articles (adapted from Swales and Feak (2004, p.158)

	Introduction	Methods	Results	Discussion
Present tense	High frequency	Low frequency	Low frequency	High frequency
Past tense	Mid frequency	High frequency	High frequency	Mid frequency
Passive voice	Low frequency	High frequency	Variable frequency	Variable frequency

Writing centres at American and British universities usually offer short courses or online recommendations about style. According to a handout and a demo from the University of North Carolina's Online Writing Center, in scientific papers, the passive

voice is used to promote objectivity (UNC Handout: Passive Voice 2010-2013).[18] In his seminal work on how to write scientific papers, Kirkman (1975) argues that authors primarily utilize the passive voice when they feel that the focus of the sentence should be on the subject and not the agent, for a variety of reasons, such as when the agent is well known and understood, or because they want to de-personalize the information in the hope of appearing more objective. However, Sheen (1982, p.21) notes that the passive voice is overly used in medical writing: 'Writers sometimes use passive voice in an attempt to sound scholarly and scientific, when actually they are perpetuating a writing tradition that is fraught with ponderous and obscure language.' In their work on the impact of style guides on medical writing, Millar et al. (2013) describe a backlash against the use of passive voice, asserting that many journals now recommend the use of active voice in order to make the article easier for the audience to read. While the shift away from passive voice may be seen more in other sections of RAs, the Methods sections in the corpus compiled by Millar et al. (2013) as well as in my corpus are usually found to be written in passive voice, as can be seen in Move Four (see 4.3.1.2.1), in which the work has been approved as being ethical by an institutional body or by following international standards.

The following table shows the occurrence and percentage of each Move and Step in this section as they occur in my corpus.

Table 4.7: Moves and Steps in the Methods section of articles in the total corpus (n=250)

	Ethics statement (48%)	**Describe what is being treated (100%)**			**Describe the experiment (100%)**			
Steps Observed	--	Area	Biodata	Rationale	Materials	Procedures	Statistics	Tests
Number observed	120	22	243	153	250	250	220	247
Total percentage	48	8.8	97	61	100	100	88	98

Note: --- indicates no Steps observed

4.3.1.2.1 Move Four: Ethics statement

---

[18] http://writingcenter.unc.edu/handouts/passive-voice/

The first Move, *Ethics statement*, is one I consider optional, as it only occurs in 120 (48%) of the articles that appear in my corpus. However, as I describe later in this chapter (see 4.3.1.2.1), it may be seen as optional for certain types of articles but mandatory for others. I define this Move as an explicit statement of the authors' having followed an approved ethical protocol, required either institutionally or field-wide, as governed by the World Medical Association's Helsinki Guidelines.[19] An example of it is:

> (15) The study was conducted under the guidelines of the local ethics committee, and in accordance to the tenets of the Declaration of Helsinki. (Text 105)

The International Council of Medical Journal Editors (ICMJE 2014a) adds that 'The requirement for informed consent should be included in the journal's instructions for authors.[20] When informed consent has been obtained, it should be indicated in the published article' (ICMJE 2014a).

The ICMJE also refers researchers to the International Association of Veterinary Editors for reporting experiments on animals if their institution does not have a board or committee; an example where such a committee exists is:

> (16) All studies involving animals were reviewed and approved by the University of Pittsburgh's Institutional Animal Care and Use Committee. (Text 2)

This Move is a standalone Move and as such has no Steps to expand it. The Move includes explicit lexemes such as: *ethics, committee, approved, observed, testing*, and *guidelines*. The Move is written in the simple past tense and is in the passive voice. This study and that by Fryer (2012) are the only studies that mention this particular Move, although Fryer (2012) treats it as a Step in his fourth Move of *Material/participants*. The reason that I considered it an independent Move is that it presents a different rhetorical function than that of describing who or what was

---

[19] http://www.wma.net/en/30publications/10policies/b3/
[20] http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html

studied and implies that the authors had to have their work reviewed or approved before undertaking the research. This is an appeal that goes beyond a statement of compliance to satisfy Teufel and Moens' (2002) principle of *scientific argument*. With this Move, the authors of an article are arguing that their work has been approved by others as following specific ethical principles and guidelines. This Move is stronger in explicitly showing the validity of the work than the following Move, *Describe what is being treated.*

4.3.1.2.2 Move Five: Describe what is being treated

The fifth Move identified, *Describe what is being treated*, is similarly described in all of the other studies that include discussion of Rhetorical Moves. This Move is mandatory as it occurs in every article in my corpus. It can be realized through any or all of the following Steps: Describe (A) the study location, (B) the biodata of the subjects, and/or (C) the rationale for inclusion or exclusion of subjects in the experiment. This Move is fairly straightforward: it alerts the reader to any relevant information that the authors feel is necessary, about who or what is the subject matter of the experiment. Of the three possible Steps in Move Five, the first one was by far the least common as it occurred only 22 times in the corpus, when the author described in detail the geographical region in which the experiment was being done. An example of this Move is:

> (17) Shanghai is the biggest city in eastern China. It consists of 18 districts and 1 county that are geographically divided into 3 strata as urban, suburban, and rural regions. Using a stratified multiple-stage sampling method, 4 residential areas in the urban region, 2 villages in the rural region, and 3 residential areas in the suburban region were selected following the sequence of district-block-residential area. (Text 17)

Of the 22 instances of this Step, all involved the description of areas outside of North America or Western Europe. Sometimes the authors give details about the area where the subjects were pooled; other times the authors provide detailed accounts of the facilities in which the experiment was being undertaken, as in:

> (18) The study was conducted in the Histopathology Unit of the Department of

>Laboratory Services at the Muhimbili National Hospital (MNH) in Dar es Salaam, which is also the teaching hospital for the Muhimbili University of Health and Allied Sciences (MUHAS). MNH is the national referral health care facility in Tanzania with a bed capacity of over 1000 and receives biopsies from most of the country except the lake zone which has its own manned pathology department. (Text 12)

As the Step reports on location, and not data or procedure, it is generally presented in the simple present tense as the information could be considered factual. The only explicit lexemes, i.e. words that clearly demarcate the Move or Step as they are part of the title, are country or region names. It is worth noting that authors of studies based outside either North America or Western Europe may feel it necessary to convey information about either the area or facilities used in their research, although a predominantly Western audience may not consider this to contribute to the strength of the research. Interestingly, during my personal interviews with members of the discourse community, one of the doctors (Doctor 7, personal interview, March 11th 2012) mentioned that he/she typically would not read articles from 'outside' as the work or subjects generally did not reflect the author's own practices. Given that doctors from all over the world publish MRAs, this viewpoint calls for future investigation.

The next Step, *Describe biodata*, appeared 97% of the time in articles within my corpus, to describe the pertinent information regarding the subjects. This Step appears widely throughout all of the other studies and it is considered by some to be a Move in itself, but since it is sharing the same rhetorical principles of the other Steps in this Move, that being intellectual attribution, I have coded it as belonging to the same Move. This Step is an explicit recount of numbers, sizes, origins, ages and occasionally, when appropriate, the genders of the subjects being treated, as seen in these examples:

>(19) Eight patients with asthma and eight normal volunteers were recruited for the study. (Text 202)

>(20) Yeast strains were selected from a larger collection of strains surveyed for variation in sensitivity to copper sulfate and those used are listed in Table 1. Of the nine strains, seven were isolated from vineyards in Italy between 1993 and 1994 by R. Mortimer [17]. The diploid, sequenced lab strain, S288C, was obtained from the Botstein lab (DBY8268). The lab strain S288C is mostly

derived from EM93, which was isolated from a rotting fig in California in 1938 [55]. The woodland strain, YPS163, and the S. paradoxus strain, YPS125, were isolated from oak tree exudates in Lima, Pennsylvania in 1999 [56]. (Text 139)

Both examples are written in the simple past tense and passive voice. The specific words that commonly occur are numbers (as in the number of subjects), words that indicate location from which the samples were derived or obtained (*procured, came from, isolated from, recruited, derived,* and so on…) and nouns describing subjects like *patients, strains, subjects, tissues*, and *samples*. The next logical Step is the Rationale for why these materials or subjects were selected.

Step C of Move Five happens in only 61% of the articles in the corpus. The previous Step described what was being tested; this Move presents reasons why those subjects were or were not included. Of the three Steps, it is apparently in this third one where the persuasive rhetoric used to convince the reader that the author is practising good science is most obvious. An example of this Step is:

(21) Patients (age 18 years+) were eligible for the study if they were admitted to the emergency ward for symptoms of psychosis as determined by a score of 4 on one or more of the items Delusions, Hallucinatory behavior, Grandiosity, Suspiciousness/persecution, or Unusual thought content in the Positive and Negative Syndrome Scale (PANSS) [13], and were candidates for oral antipsychotic drug therapy. (Text 89)

In this example, we can see those selected for the study had to meet certain criteria to be 'eligible'; the following example shows the reasons for excluding subjects:

(22) Four young and two older participants were excluded because they either generated unusable data (n = 3 young), showed outlying scores of more than three standard deviations away from the mean on the primary behavioral task (n = 1 young, n = 1 older), or did not perform the task properly (n = 1 older). Those who recalled more than 8 items only showed off-line improvements over sleep, whereas those recalled 4-items or less demonstrated off-line improvements over both wake and sleep [9]. To remove this possible impediment to off-line skill improvements, participants who recalled more than 4 items of the sequence were excluded from analysis (n = 2). (Text 204)

The Step typically uses simple past tense and its characteristic lexemes are as follows: for nouns, *criteria, eligibility, rationale*, and *reasons*; as for verbs, *included* (113 out

of the 153 texts that had this Step) and *excluded* (53 out of the 153) were the most common. Of the other studies, only Skelton (1994) and Fryer (2012) mention this Step as an individual Step, Nwogu (1997, p.129) provides an example in his description of his fourth Move, 'Criteria for inclusion in the trial were . . . ' but he does not mention a declaration of criteria as a Step, possibly because his corpus was smaller. Something that appears only 60% of the time in my larger corpus of 250 articles may occur only a few times in a smaller corpus of 30 articles, like Nwogu's. Kanoksilapatham (2005) does not include or mention this Step in her work.

The fifth Move, *Describe what is treated,* is mandatory and appears in all 250 articles; its rhetorical purpose is to demonstrate to the readership that the authors practise good science. This Move fulfils two of Teufel and Moens' (2002) rhetorical principles, those of *scientific argumentation* and *intellectual attribution*. The authors are detailing not only who or what they studied but why the subjects were chosen, which fulfils *scientific argumentation* and since they may be excluding or including subjects in a way that differs from other researchers, they are differentiating their work from that of others, which fulfils *intellectual attribution*. This can be seen in Example (22) above.

4.3.1.2.3 Move Six: Describe the experiment

The sixth Move found in my corpus of MRAs is *Describe the experiment*. This Move, identified as such by all of the other studies, describes the procedures and tests, both medical and statistical, being conducted in the experiment. However, the other researchers usually separate what I have identified as one Move into two Moves. While I consider describing the Steps taken and the measurements used in the test or experiment as sharing similar features and similar rhetorical intent, and thereby being parts of the same Move, other researchers separate these two Steps into two distinct Moves. For example, both Nwogu's (1997) fifth Move, *Describing experimental procedures* and his sixth Move, *Describing statistical procedure*s, share similar features, such as being written in the passive voice, defining or explaining reasoning for particular Steps or statistical tests. These similarities do not warrant separation of the Steps into full Moves (see 4.3). Fryer (2012) follows Nwogu's pattern in

separating the two. Skelton is the only author to note that the Moves might be separate because there typically is a heading or headword clue that statistical and experimental procedure and tests make up different sections of the text. I would argue that the similarity in the rhetorical intent of both Steps in the Move makes it more difficult to separate them. In this Move the authors attempt to persuade the reader that they have followed the best possible science: that is, they have chosen the best and most scientifically appropriate way to design and conduct their experiment and measure the results. I separated the Move into four Steps: Steps that describe the materials used, the Steps taken in the experiment, any medical tests, and the choice of statistical tests. An example of Step (A) is:

> (23) Serial sections were stained with H&E and CD31 (Clone JC70A, Dako Denmark, dilution 1: 30) following routine immunohistological methods using Aethylcarbazole as chromogen and hematoxyline as counterstain. (Text 162)

This Step describes what instruments (tools) or chemicals were used on the subjects, from surveys to chemical staining, while the next Step outlines the Steps and or precautions taken while doing the experiment. An example from the corpus is:

> (24) As described previously [20,58], we accounted for possible receipt of anti-retroviral therapy (ART) by using the calendar year of membership in the cohort as a proxy for receipt of mono/dual therapy (1990-1996) and HAART (1996-onwards). All subjects who were recruited in the therapy eras were pooled into a single group [20]. The therapy eras and the number of subjects with cohort membership during each era are as described previously [20]. Predictably, membership of a subject to the era in which HAART was available was associated with a significantly reduced risk of progressing to AIDS [20]. (Text 196)

The next Step describes any medical tests run on the subjects, as exemplified by the following:

> (25) The relative equilibrium binding constants, Krelative, of 26 individual clones, lac P1, ICAP with I-D and three mutant of G8.05 were measured by EMSA in a competition assay as previously described (30). All experiments were performed at least in triplicates. In this assay, a mixture of two different sized DNAs (5-20 pM), both containing a binding site for CRP, competes for a limited amount of CRP protein simultaneously. (Text 148)

The last Step is a reference either to the statistical test(s) to be run or to how statistics

will be employed in the research, such as:

> (26) Statistical analysis was performed using SPSS (version 11, Chicago, IL) statistical software. Categorical variables were compared with Chi Squared Test, or Fishers' Exact Test (2-tailed), where appropriate. Continuous variables were analyzed with the Mann Whitney U test (Text 105)

Table 4.8 describes each Step in terms of tense, voice, explicit lexemes and other stylistic constituents of the four Steps in this Move.

Table 4.8: Characteristics of Steps in Move Six

	Description of Materials	Description of Experimental Steps	Description of Tests	Statistics
Tense	Past	Past	Past	Past
Voice	Passive	Passive	Passive	Passive
Explicit Lexemes	*stained* *exposed* *injected* *subjected*	Ordinals, Adverbs of frequency, time and manner	*tested* *performed*	*Analysis* *t-test* *p-value* *percentages*
Stylistic Features	Tends to be focused on what is being used to affect the subjects; uses medical terms, names of medical devices.	Usually written as a procedural list, presenting a narrative-like framework; uses acronyms and technical jargon.	Jargon-rich, with names of tests that other experts would recognize as being the best possible science.	Numbers and technical jargon used in statistics.

The different Steps of this Move have many similarities, ranging from tense and voice to the use of complex or elite medical and statistical terms and jargon. Based on Teufel and Moens' (2002) classification of rhetorical principles, I assert that the authors appeal to the reader through *scientific argumentation, problem-solution* and *intellectual attribution.* For *scientific argumentation,* the authors claim that they used the best materials and best tests, both medically and statistically, to help solve the problem previously identified, in order to convince the reader that the authors' work is of the highest standard. This claim feeds the *problem-solution* category, as the authors are explicitly stating their procedures for finding and measuring their solution to the problem. In stating what they as author-researchers have done, they imply a difference between their work and that of others or they offer an explicit statement of how the current research is based on previous research. With the latter, it is as if the author is trying to protect their research from criticism by tying their procedure to previous work, as in this example:

(27) For this study, we developed a questionnaire to investigate knowledge and attitude concerning opioids and pain management at the end of life, based upon previous studies [19,20,28- 34]. Questions were adapted and added to be able to answer the research questions, and to fit the specific situation in the Netherlands. For this purpose we used Dutch guidelines, original articles and review articles [17,22,23,35- 53]. (Text 109)

4.3.1.3 Moves and Steps in the Results section

Table 4.9: Moves and Steps in the Results section

Results section						
Skelton (1994)	**Nwogu (1997)**	**Kanoksilapatham (2005)**	**Li and Ge A (2009)**	**Li and Ge B (2009)**	**Fryer (2012)**	**Present study**
*8. Adjustments and exclusion from the original population 9. Represent results in tables 10. Discussion of data in words 11. Assessment of the data	7. Indicate consistent observation *8. Indicate non-consistent observations	8. State procedures A. Describe aims and purposes B. State research questions C. Make hypotheses D. List procedures or methodological techniques 9. Justify procedures or methodology A. Cite established knowledge of the procedure B. Refer to previous research 10. State Results A. Substantiate results B. Invalidate results 11. Start comments on the results A. Explain the results B. Make generalizations or interpret results C. Evaluate current findings by those from previous studies or with regards to hypotheses D. State limitations e. Summarize	7. Indicate consistent observation *8. Indicate non-consistent observations	7. Indicate consistent observation *8. Indicate non-consistent observations	7. Observations A. Reference to non-verbal materials B. Main findings C. Associations/ correlations D. Adjustment to analysis	7. Report the findings A. State the outcomes B. Support with evidence C. Direct readers to figures or tables 8. Review the experiment A. Review of the subjects B. Review of procedures C. Review and/or compare to previous studies 9. Describe statistical tests and state findings A. Report statistical data B. Analyse statistical data *10. Summary report of the overarching findings

NB: * denotes Moves or Steps that the authors coded as being optional; Li and Ge A and B are

different corpora within the same article

The Results section seems straightforward in purpose. According to Ng and Peh (2008, p.967), editors of the *Singapore Medical Journal*:

> The purpose of the results section is to present the main data collected and the observations made during the research. It provides interpretation of the analysed data and does not contain details on the methods, materials or discussion.

The on-line resource from Bates College (2002)[21], *The Structure, Format, Content, and Style of a Journal-Style Scientific Paper* describes the function of the Results section as follows:

> The Results section is to objectively present your key results, *without* interpretation, in an orderly and logical sequence using both text and illustrative materials (Tables and Figures).

From these two statements of purpose, it could be gleaned that the Results section is merely an objective report of findings and data collected. However, a review of the Moves found in the current research, as well as in studies of Moves by others, hints at there being more rhetorical functions involved in this section, including how it is presented to the readership.

In this section, I found three main Moves and an occasional fourth, with a minimum number of Steps. The first was a Move of *Report the findings*, which was comprised of three Steps: (A) *State the outcomes*, (B) *Support with evidence,* and (C) *Direct readers to figures or tables*. Next was a Move to *Review the findings*, which has these Steps: (A) *Review of the subjects*; (B) *Review of the procedures* and (C) *Review and comparison to previous studies*. The third Move in the section was *State and describe statistical findings*, which comprised two Steps: (A) *Report of statistical data* and (B) *Analysis of statistical data*. The fourth Move, considered optional, was the *Summary report of the overarching findings*; although this may also appear in the Discussion sections, it did occur in 55 of my texts and is also found by Kanoksilapatham (2005).

---

[21] http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html

4.3.1.3.1 Move Seven: Report the findings

Move Seven, *Report the findings,* occurs in every article in the corpus of MRAs and is comprised of at least two and usually all of these three Steps: (A) *State the outcomes,* (B) *Support with evidence,* and (C) *Direct readers to figures or tables.* The Step, *State the outcomes,* appears in all 250 articles in the corpus; a typical example is:

> (28) Univariate analysis showed that all the variables listed in Table 1 were highly associated with SES level based on area of residence at the time of diagnosis. (Text 10)

The excerpt illustrates a report of findings drawn from mathematical analysis of the data. The tense is normally the past simple, with the Step usually being written in active voice. Explicit lexemes are verbs: *found, showed, discovered,* and *indicated;* and nouns: *analysis, data, results.* This Step is part of the Moves identified by all of the other researchers; however, the next Step, *Support with evidence,* appears only in the works of Nwogu (1997) and therefore Li and Ge (2009) who use Nwogu's framework, and also Kanoksilapatham (2005). *Support with evidence* is the Step in which the authors attempt to explain how they determined or interpreted the results as shown in the following:

> (29) Cells with two apparently segregated nucleoids and, most frequently, a constricting division septum (herein called double-cells, examples are tagged with red dots in Figure 2B) are over-represented in XL151 populations growing under permissive conditions (55% compare to 21% for wt; Figure 2A). These data suggest that the main effect of the inversion on cell growth is to delay cell division. (Text 179)

This Step presents the explanation underlying the determination of the findings, to help diminish subjectivity and furnish the reader with a way to see how the authors may have concluded that the data or findings actually do represent the findings that they are reporting. It is written in the simple past tense and is presented in the active voice. Explicit lexemes characterizing this Step are words and phrases that indicate judgment on the part of the author, such as *suggests, possibly because, due to,* and *from the data*.

The last Step of this Move, *Direct readers to figures/tables* is one in which the authors explicitly direct the readership to examine tables or figures. This can be accomplished in two ways: the first is by stating the result with a parenthetical reference to the figure or table, an example of which is:

> (30) The tracheostomy was decannulated 2 months after injury. Despite frequent retrograde dilatations, the patient developed recalcitrant strictures (Figure 5). (Text 163)

In the second way, the author requests that the reader look at a specific table or figure that reports data:

> (31) For details, see figure 2. (Text 18)

Explicit lexemes are *table* or *figure,* the voice is active, and the simple past tense is typically used. While this Move is primarily concerned with reporting findings, it does have some rhetorical function, for the authors need to persuade the readers that these findings are accurate and, indeed, the only findings that could have occurred. This is most often accomplished by Step B of this Move, *Support with evidence*. In applying Teufel and Moens' (2002) framework of rhetorical principles, we see that this Move is centred on two of their principles: *problem-solution* and *scientific argumentation*. *Problem-solution* is represented because these findings are what the authors will display as being the solution, and *scientific argumentation* through elaborating to the reader how the findings are both accurate and valid. An argument can be made that this Move, because of the emphasis on layout in both tables and figures, is closely tied to the evolving field of visual rhetoric (Tufte 2001, pp.13-16; Foss 2005, p.304).

4.3.1.3.2 Move Eight: Review the experiment

Move Eight, *Review the experiment,* comprises three Steps: *Review subjects; Review procedures*; and *Review and/or compare to previous studies*. The first Step, *Review subjects,* is also listed by Kanoksilapatham (2005) and Skelton (1997). In it, authors review who or what was treated. As previously mentioned (see 3.3.4, Doctor 2,

personal interview March 20th 2012), the repetition may be an effort to remove ambiguity or to provide additional precision. Repetition is in itself a well-known rhetorical device, usually but not exclusively used for emphasis. In Example (32) an instance of this Step is from the Results section of Text 63 below, an excerpt which one would normally expect to find in the Methods section, or possibly even in the Introduction:

> (32) The core lab received total RNA from ocular tissues, primary cultures, and cell lines collected or derived from three mammalian species: human, monkey, and rat. (Text 63)

The tense used in the first Step is the simple past tense, the voice is active, and typical lexemes include *from, total(s)*, *subjects*, and *groups*.

Step B of Move Eight reviews procedures, often as repetition of part of an earlier segment in the Methods section, as illustrated by:

> (33) In a first step, we investigated viewing behavior relative to bubbles. Subjects made, on average, 6.2 fixations in each trial where bubbles were presented. (Text 174)

The tense is typically the simple past, in active voice; lexemes are those found earlier in the Methods section for the procedures used in a particular study, including ordinals, adverbs of frequency, time and manner, and content-appropriate nouns. The seminal work by Burke, *A Rhetoric of Motives* (1969, p.67) describes repetition at length, characterizing its different forms as being different rhetorical devices. Though it is important to note that both Fryer and Skelton comment on the relative absence of personal pronouns in MRAs, it is particularly in this Step that my corpus evidences the use of *we* and *our.*

The third possible Step, *Review and/or comparison to previous studies*, is one in which authors review results and/or procedures of previous studies by themselves or by others, as exemplified in:

> (34) Some of these aromatic residues, notably in the vicinity of Trp552 have previously been implicated in capped RNA binding by cross-linking studies

[31,32]. (Text 224)

> (35) This result agrees with previous observations that the transcript and protein expression levels of SodB are repressed in the fur mutant of S. oneidensis [10]. (Text 96)

Sometimes the authors wish to align their work a bit more closely with that of others, as shown in the rationale presented ('previously been implicated') for citing the 'cross-linking studies [31, 32]' in Example (34) above, while in other instances the authors simply wish to present a quick review, as in Example (35), with 'agrees with previous observations' but listing only one citation. The voice is typically active, the tense is past, and frequent lexemes and phrases include *previous studies, as found by, previously*, and *notably.*

The example below illustrates persuasion using comparison to previous works as well as presenting a review:

> (36) Efforts to demonstrate biochemically the direct binding of Staufen to oskar mRNA are notably absent in the literature. Therefore, in order to investigate the role of Staufen protein in the transport of osk, we co-visualized the two molecules in live oocytes and studied their mobility during oocyte development. (Text 203)

Move Eight is one of repetition; following Teufel and Moens (2002), the rhetorical functions include *intellectual attribution* and *attitudes toward other people's works*, as there is a comparison to previous work coupled with a review of the authors' own work; and *scientific argumentation* and *problem-solution*, as the authors present a claim with logical steps toward their solution.

4.3.1.3.3 Move Nine: Describe statistical tests and state their findings

In Move Nine, *Describe statistical tests and state their findings*, we find a mixture of review and analysis in two successive Steps, first reporting the data and then analyzing the data. I have chosen to keep the two Steps in a single Move as they are so closely tied together. In the examples that follow, Example (37) illustrates the review combined with the analysis, while Example (38) illustrates analysis alone:

(37) Supporting the hypothesis that plus-end transport may be quite sensitive to BicD levels, we note that in phase II, minus-end motion in the t = 1 and t = 3 backgrounds is approximately the same (t-test, p = 0.56; Fig. 3A), but the small amount of additional BicD present in the t = 1 embryos is enough to significantly improve Phase II plus-end motion above that found in the t = 3 embryos (t-test, p = 0.049; Fig. 3B). Droplet-bound BicD levels are temporally changing, and regulated. (Text 198)

(38) Differences in the constitutive and stimulated release of both chemokines between HBMEC from different donors; however, the pattern of chemokine release after cytokine or LPS treatment was similar in all experiments. (Text 116)

Both Steps are in the active voice while the review may be in the simple present tense, the analysis is more frequently in the simple past tense, and the two Steps differ in their characteristic lexemes. In Step A, readers find *p value, average, significant correlation*; in Step B, readers are more likely to find *due to* and *from*.

4.3.1.3.4 Move Ten: Summary of findings

The final possible Move in the Results section is a *Summary of findings*, as shown below in the extract below (example 39) from Text 118, and is considered optional as it only occurs 55 times in this corpus, or in less than 25% of the texts, and may be a stylistic choice of the authors. It is expressed in the active voice, past tense, and includes phrases such as *in summary* and *to sum up*:

(39) In summary, the nonparetic arm of the patients showed similar properties as both arms of the control group, although being slower in the performance of the task. (Text 118)

The table below shows the Moves and Steps found in Results sections in the MRA corpus. From the table, we can see that *Statistics* and *Reporting findings* are more prevalent, and therefore are considered obligatory, than either the Review or the Summary sections. While my findings support the conclusions of other researchers that the Results section is primarily for the objective reporting of findings, nonetheless, the presence of both Review and Summary sections suggests that authors may have increased their emphasis on rhetorical functions of their reports, thereby

introducing additional subjectivity. Alternatively, perhaps journals have increased their expectations for authorial clarity since 2000, the date of the earliest articles in my corpus, though none of the members of the sample DC explicitly stated a date for a shift, but several mentioned that they have found that in their time there has been more emphasis placed on data (see 3.3.3).

Table 4.10: Moves in the Results section by number and percentage of occurrence

	Report the findings (100%)			Review the experiment (76%)			Describe statistical tests and state findings (91%)		Summary (22%)
Steps Observed	State the out-comes	Support with evid-ence	Direct readers to figures	Review subjects	Review proced-ures	Review and/or com-pare to pre-vious studies	Statistics	Tests	---
Number observed	250	249	153	87	178	75	214	197	55
Total percent	100	99	61	35	71	30	86	79	22

Note: --- indicates no Steps observed

4.3.1.4 Moves and Steps in the Discussion section

Table 4.11: Moves and Steps in the Discussion section

Discussion section						
Skelton (1994)	**Nwogu (1997)**	**Kanoksilapatham (2005)**	**Li and Ge A (2009)**	**Li and Ge B (2009)**	**Fryer (2012)**	**Present study**
12. Limitations 13. Central achievements of the study 14. Contextualize the results 15. Recommendations	9. Highlight overall research outcome 10. Explain specific research outcomes A. State a specific outcome B. Interpret the outcome C. Indicate significance of outcome D. Contrast present and previous outcomes E. Indicate limitations	12. Contextualize the study A. Describe established knowledge B. Present generalizations, claims, deductions, or research gaps 13. Consolidating results A. Restate methodology B. State selected findings C. Referring to previous literature, D. Explain differences in findings E. Make overt claims or	9. Highlight overall research outcome 10. Explain specific research outcomes A. State a specific outcome B. Interpret the outcome, C. Indicate significance of outcome D. Contrast present and previous outcomes E. Indicate limitations of	*9. Highlight overall research outcome 10. Explain specific research outcomes A. State a specific outcome B. Interpret the outcome, C. Indicate significance of outcome D. Contrast present and previous outcomes E. Indicate limitations of outcomes	8. Main Findings A. Compare main findings in relation to hypothesis B. Compare with the literature C. Possible mechanisms/ causes D. Back-ground for the study 9. Limitations 10. Conclu-sion: A. Reiteration of main findings	11. Discoveries from research A. Main findings B. Second-ary findings 12. Treat-ment review 13. Ration-ale for the solution 14. Compari-son to previous

of outcomes 11. State research conclusions A. Indicate research implications B. Promote further research	generalizations, F. Exemplify. 14. State limitations of present study A. Limitations of the study about findings B. The methodology C. The claims made 15. Suggest further research	the outcomes 11. State Research conclusions: A. Indicate research implications B. Promote further research	11. State Research conclusions: A. Indicate research implications B. Promote further research	B. Implicat-ions C. Recommenda-tions for future research	studies *15. Limit-ations 16. Suggest-ions for future research *17. State-ment of study importance

NB: *denotes Moves or Steps that the authors coded as being optional; Li and Ge A and B refer to different corpora reported in the same article.

Given that the Discussion sections are the culmination of the Introduction, Methods and Results sections in each MRA, it stands to reason that they will show the greatest variety and range in Moves and Steps. As I have shown, an MRA is written in a problem-solution framework. The Introduction presents the problem, the Methods describe the way the authors seek a solution; the Results present the findings; and the Discussion section briefly reviews and highlights each of those sections before elaborating on how the solution responds to the problem. Accordingly, this section will often have overt or explicit rhetorical aims and purposes, for it is the authors' final opportunity to suggest their credibility as researchers, to assert the lucidity and appropriateness of their approach to the problem, and to advance the correctness of their solution.

I have found seven possible Moves in the Discussion sections of the MRAs in my corpus. These are: *Discoveries from the research* (which includes two Steps), and *Main findings* (A) and *Secondary findings* (B); the Move of *Treatment review* is used to report the procedures of any kind, both experimental and statistical, and is followed by a Move that proposes a justification or gives a *Rationale for the solution* to the problem. The next Move offers a *Comparison to previous studies*; the subsequent Move notes *Limitations* within the study or within the field as a whole; this is followed by a Move of *Suggestions* for future research. Finally, an optional Move is an explicit *Statement of the importance* of the study and its findings.

4.3.1.4.1 Move Eleven: Discoveries from the research

The first Move, *Discoveries from the research*, incorporates two Steps: *Main findings* (A) and *Secondary findings* (B). *Main findings* present the most salient results that attempt to fill the gap of knowledge identified in the Introduction section, as illustrated below

> (40) On the basis of phylogenetic analysis in conjunction with historic and recent epizootiologic data on rabies, we found no evidence of enzootic dog rabies in the United States for the past 13 years, [1,15,16,18,19,30]. (Text 1)

The Step is typically written in past tense, active voice, and includes lexemes such as *found, discovered, identified,* and *findings.*

In *Secondary findings*, the authors include additional or supplemental findings, or unexpected outcomes that the authors feel warrant inclusion, such as:

> (41) We further found that both SNPs had differential binding efficiency to nuclear proteins in the gel shift experiment. (Text 223)

This example illustrates a point made earlier in Move Eight: the use of *We* signals that the authors can and will also use personal pronouns in this section, in contrast to claims made by other researchers. This Step is written in simple past tense and active voice, with characteristic lexemes and phrases being *further found, in addition, additionally,* and *moreover*.

The Move itself is a combination of *problem-solution* structure and *intellectual attribution* (Teufel and Moens 2002), in which authors present their solution to the problem they earlier identified. Here, the attribution is to them, as they put themselves and their solution in the spotlight.

4.3.1.4.2 Move Twelve: Treatment review

The Move, *Treatment review,* highlights the measures taken to solve the problem identified, whether it be of procedures or anything to do with the experiment, as seen in:

> (42) We used the novel QSAR method DF to develop two classification models to predict ER binding. We furthermore objectively and quantitatively assessed the applicability domains of the models by computing prediction confidence and domain extrapolation for predicting unknown chemicals with an extensive cross-validation. (Text 135)

It is written in the simple past tense, in the active voice, and its characteristic lexemes include *done, used, assessed, measured,* and *analysed.*

4.3.1.4.3 Move Thirteen: Present a rationale

In the Move, *Present a rationale*, the authors shift to using the present and future simple tenses and modals to suggest how their solution could become knowledge once the study has been accepted for publication. In the two examples that follow, the voice remains active, centred on the authors' findings, but the tense shift and use of modals suggest either some uncertainty or some notion of what has not yet been accomplished. This Move can incorporate either a further explanation of purposeful action or the rationale underlying the potential use of the findings:

> (43) The reason that removal from the membrane is important is not clear, but Rac1-GDP present in the membrane may be transiently reactivated, reversing the effects of the RhoGAP. (Text 222)

> (44) The effects of mutations upon protein-protein affinities of these domain-host factor complexes can then be measured and structural studies will help in understanding the nature of the interaction interfaces and the contributions of the surface-exposed host determinant residues. (Text 224)

Treatment review is frequently used to illustrate the authors' good practices in *scientific argumentation* (Teufel and Moens 2002) in an effort to persuade the readers of the validity of both process and findings. Explicit lexemes include *reason, purpose*, and modals *will, can* and *may.*

4.3.1.4.4 Move Fourteen: Compare with other studies

The Move, *Compare with other studies*, repeats some of the findings of other studies, in order to focus more explicitly on methodological differences that have resulted in

the new findings of the current authors, or on the use of similar methodologies which have brought about different results. This Move is one of *intellectual attribution* (Teufel and Moens 2002) in its examination of potential relationships to other work in the field, but it also serves to extend the *scientific argumentation* from the preceding Move, reminding the readers of the validity and correctness of the current study. In addition, this Move includes *attitude toward work by others* (Teufel and Moens 2002), as it allows authors to pass judgment on previous research. Characteristic lexemes and phrases include *difference, different, differ, current, present, between, comparison, previous studies* and *in contrast*, as well as surnames of specific authors of prior studies. As Thompson (2005, p.36) comments, this use of intertextual reference allows the authors to use their own words to express a proposition while attributing its validity to other writer(s), as explained by Hunston (2000, p.192) who called this anonymous referral a 'sourced averral'. Both sourced averral and trust in or support of citations reported by authors will be discussed in more detail in Chapter 5. The simple present tense is used, since other studies have been published and can be treated as knowledge, as shown in:

> (45) However, a critical difference between the current experiment and EMDR is that EMDR is used for treating existing flashbacks in PTSD (at least one month post-trauma), but is not intended to be used during the memory consolidation phase targeted in the current study. (Text 200)

4.3.1.4.5 Move Fifteen: Limitations

I consider the Move of *Limitations* to be optional. Although other studies of MRAs and of RAs assert this Move to be obligatory, I have found it in fewer than half, or 48.8%, of the articles in my corpus, which suggests that the expectations for the explicitness of the Move may be beginning to shift within the discourse community. Not all professional journals require it in their style sheets, such as *The Lancet*.[22] Announcing the limitations of a study, or what the study could have considered, is a hedge; an instance of tentative language. The study could have, for example,

---

[22] http://www.thelancet.com/pb/assets/raw/Lancet/authors/lancet-information-for-authors.pdf

limitations stemming from the methodology, such as in the size of the sample, which could affect the results of the current study, as seen in this example:

> (46) There are potential limitations related to measurement imposed by the need to be parsimonious in questioning due to use of telephone interviews. (Text 249)

Alternatively, it could point to limitations across the field, as illustrated in the following example:

> (47) One of the most important strengths of the present study in comparison with earlier cohort studies is the precision of the risk estimates because of the large number (n=156) of women with ovarian cancer studied; in most previous cohort studies, the number of women ranged from one to 12.[4 7 8 12 13 16-18] More cases (n=45) were included in another study, [19] although that study consisted of a mixture of non-epithelial, epithelial, and borderline ovarian tumours. Thus all the earlier cohort studies were limited by imprecise risk estimates, especially in subgroups of fertility drug users. (Text 75)

The *Limitations* Move allows the authors to demonstrate skill in *scientific argumentation* (Teufel and Moens 2002) and to signal their awareness that areas and issues remain to be studied, beyond the confines of the present work. Not all models use the headword of *Limitations* to signal such awareness or such skill, but will instead explicitly state something like:

> (48) There are potential limitations related to measurement imposed by the need to be parsimonious in questioning due to use of telephone interviews. (Text 249)

The voice is active and the tense is typically simple present, as the authors have current and present knowledge that certain issues are not covered. Characteristic lexemes beyond the headword include *shortcomings* and *drawbacks*.

4.3.1.4.6 Move Sixteen: Suggestions for future work

In *Suggestions for future work*, the authors not only discuss new research that could be developed, particularly if based on the acceptance of current findings, but also

identify potential uses for the findings themselves. This Move can be signalled with headwords such as *suggestions* or *implications*, and is often but not necessarily tied to any statements of limitations of the current work. According to a personal interview with Doctor 6 (March 8th, 2012), authors often use this section to establish a gap they intend to fill with subsequent research and publication. In other words, they can use this section to stake a claim for future work. Such a function would be one of *intellectual attribution*, but tied to self/selves rather than to other scholars:

> (49) The PBDE-induced changes in TH-regulated gene transcripts seen here indicate that neurogenesis and brain development may be impacted by PBDE exposure, and they highlight the need for future investigations into how PBDEs influence TH-mediated neural function. (Text 136)

4.3.1.4.7 Move Seventeen: Statement of study importance

The Move, *Statement of study importance* is the explicit statement of the importance to the field of the study and its findings, and may occur in a form such as this:

> (50) This study is the first to investigate the therapeutic impact of re-scripting social phobia-related traumatic memories. (Text 164)

This is an optional Move; journals may question the use of expressions such as 'This study is the first to do Y' as, given the expansion in publication outlets, particularly via the Internet, it is often impossible for an author to be fully confident of such a claim. For example, a recommendation in the Guidelines of the ICMJE (2014b, p.15) states: 'Avoid claiming priority or alluding to work that has not been completed'.

In the Introduction, the authors have, presumably, indicated a research gap that they intend to fill, and have suggested the importance of such research to the field. The entire article is designed to persuade the reader that its topic is important, and one with which the audience can identify, not just that its findings are significant either clinically or statistically (see Ioannidis 2005 for extensive discussion of significance). Swales and Feak (2004, p.195) add that by the time the scholar has arrived at the Discussion section – which assumes the expert has gone through the article in linear order of IMRD sections – the reader can be assumed to have a considerable amount of

shared knowledge with the authors. This Move is one of *intellectual attribution* (Teufel and Moens 2002) and an implicit comparison with work by others; it is written in the active voice and the present tense, with characteristic lexemes such as *first, most,* and *important.*

4.3.2   Moves distributed by subjects

During the initial coding of the corpus, I noticed what seemed to be a trend of certain Moves occurring with greater frequency when an article dealt with a particular type of experimental subject. To check my perception, I divided the corpus into categories based on type of subjects, consulting with medical experts such as Dr. Sara Campbell (see 4.2) on the labelling of subject type. Accordingly, I categorized articles as falling into one of seven types of subjects: human; cellular organisms; animal; bacteria; DNA; plants; and data, as displayed in the table below:

Table 4.12: Subject category types and kind of study

Subject category	Number	Definition of type of study
Humans	103	A study reporting direct testing of a patient or group of people, through either surveys or medical experiment.
Cells	50	A study that looked at parts of a cell and/or tumors.
Animals	33	A study on any other living creature that was not human: mice, pigs, dogs, and flies or other insects
Bacteria	6	A study looking at the cellular level, such as strains of bacteria or viruses
DNA	38	A study on genetics, DNA or RNA
Plants	3	A study that analysed plants, trees, or flowers
Data	17	A study on previously collected data, or on statistical models or on possible computer programs to be used in the future

I first reviewed the full collection of data and used Excel to identify the percentage and frequency of each Move and its Steps, which gave me observable data, displayed below in Table 4.13.

Table 4.13: Observable data across the sections

a. Moves and Steps found in Introduction sections of corpus and percentage of the total 250 articles

Moves	1. Describe the study situation		2. Describe the problem		3. Describe way to the solution			
Steps	Framing Statement	Background work by others	Gap	Hypothesis	Outline procedures	Subjects	State aims	Preview findings
Number	250	250	244	66	199	239	71	66

| Percent | 100 | 100 | 97 | 26.4 | 79.6 | 95.6 | 28.4 | 26.4 |

b. Moves and Steps found in Methods sections of corpus and percentage of the total 250 articles

Moves	4. Ethics statement	5. Describe what is being treated			6. Describe the experiment			
Steps	---	Area	Bio-data	Rationale	Materials	Procedures	Tests	Statistics
Number	120	22	243	153	250	250	247	202
Percent	48	8.8	97.2	61.2	100	100	98.8	80.8

Note: --- indicates no Steps observed

c. Moves and Steps found in Results sections of corpus and percentage of the total 250 articles

Moves	7. Report findings			8. Review experiment			9. Statistics		10. Summary
Steps	Outcome	Evidence	Figure	Subjects	Procedures	Previous work	Statistics	Analysis	Summary
Number	250	249	248	87	178	75	214	197	55
Percent	100	99.6	99.2	34.8	71.2	30	85.6	78.8	22

d. Moves and Steps found in Discussion sections of corpus and percentage of the total 250 articles

Move	11. Discoveries from research		12. Treatment review	13. Rationale for the solution	14. Comparison to previous studies	15. Limitations	16. Suggestions	17. Statement of importance
Steps	Main	Secondary	---	---	---	---	---	---
Number	250	236	223	212	247	122	191	61
Percent	100	94.4	89.2	84.8	98.8	48.8	76.4	24.4

Note: --- indicates no Steps observed

Next, using Table 4.13, I used Excel to investigate whether there were significant patterns in any particular Move, based on frequencies of the subjects used in the experiments, which allowed me to compare observable with expected data in the first three sections. Comparisons for the fourth section, Discussion, were not significant, as they took their meaning and interpretation from the previous sections. A statistician conducted a spot-check analysis of my data to verify my results.[23]

Table 4.14: Correlation of observed Moves and Steps with subject types in Introduction section

Subjects	Statement	Background	Gap	Hypothesis	Outline	Subjects	Finding	Aims	Total
Human	103	103	100	31	77	97	12	44	567
Cell	50	50	49	11	38	50	18	9	275
DNA	38	38	37	3	30	36	18	5	205
Animal	33	33	33	14	28	31	12	8	192
Data	17	17	17	7	17	16	2	3	96
Bacteria	6	6	5	0	6	6	2	2	33
Plants	3	3	3	0	3	3	2	0	17
Totals	250	250	244	66	199	239	66	71	1385

[23] Dr. Michael McEwan, University of Glasgow

With a Chi-square test p-value of 0.1893, there is no significant pattern for any of the Moves as they relate to the subject types in the Introduction section. A proportions table allowed the identification of a generalization.

Table 4.15: Proportions for Moves and Steps relating to subject types in the Introduction section

Subjects	Statement	Background	Gap	Hypothesis	Outline	Subjects	Findings	Aims
Human	1.00	1.00	0.97	0.30	0.75	0.94	0.12	0.43
Cell	1.00	1.00	0.98	0.22	0.76	1.00	0.36	0.18
DNA	1.00	1.00	0.97	0.08	0.79	0.95	0.47	0.13
Animal	1.00	1.00	1.00	0.42	0.85	0.94	0.36	0.24
Data	1.00	1.00	1.00	0.41	1.00	0.94	0.12	0.18
Bacteria	1.00	1.00	0.83	0.00	1.00	1.00	0.33	0.33
Plants	1.00	1.00	1.00	0.00	1.00	1.00	0.67	0.00

The only generalization for the Introduction section that can safely be drawn from the proportions table is that the first three Steps are not influenced by subject matter. The Step for *Hypothesis* in the Introduction does, however, seem to be more prevalent in articles dealing with animals, humans and data than with DNA, plants or bacteria. The *Outline of procedures* and the *Review of subjects* are also not influenced by subject type, although the Step for the explicit *Statement of aims* is more likely to appear in articles about humans or bacteria. Articles in which plants are the focus are the only ones likely to contain the Step of previewing the findings.

A Chi-square test on the Methods section returned a p-value of 0.0279, indicating a significant pattern. The following table displays the observed data and their occurrence in the corpus as relating to the different test subjects.

Table 4.16: Correlation of observed Moves and Steps with different subject types in the Methods sections

	Ethics	Area	Bio-data	Rationale	Materials	Statistics	Procedure	Tests	Totals
Human	77	15	102	87	103	91	103	101	680
Cell	13	2	47	21	50	37	50	50	271
DNA	5	0	38	9	38	23	38	38	189
Animal	21	1	32	20	33	29	33	33	202
Data	4	3	15	13	17	15	17	16	101
Bacteria	0	0	6	2	6	5	6	6	31
Plants	0	1	3	1	3	2	3	3	16
Totals	120	22	243	153	250	202	250	247	1490

Because the significance is strong (p=< 0.0279), it is possible to state that, for example, research on humans will typically include an Ethical Move, while in plant or

bacteria experiments, such a Move may not be necessary. Another way of interpreting these findings is that if there is a difference in the proportion between Moves, it is directly related to the nature of the subject being tested. The following proportions table will help to highlight the difference between the Moves by subject types.

Table 4.17: Proportions of Moves and Steps in the Methods sections as they relate to subject types

Subjects	Ethics	Area	Bio-data	Rationale	Materials	Statistics	Procedure	Tests
Human	0.75	0.15	0.99	0.84	1.00	0.88	1.00	0.98
Cell	0.26	0.04	0.94	0.42	1.00	0.74	1.00	1.00
DNA	0.13	0.00	1.00	0.24	1.00	0.61	1.00	1.00
Animal	0.64	0.03	0.97	0.61	1.00	0.88	1.00	1.00
Data	0.24	0.18	0.88	0.76	1.00	0.88	1.00	0.94
Bacteria	0.00	0.00	1.00	0.33	1.00	0.83	1.00	1.00
Plants	0.00	0.33	1.00	0.33	1.00	0.67	1.00	1.00

It is fair to say that studies involving humans and animals need *Ethics statements*, while others do not. Unless authors feel it is relevant, regardless of subject, they do not need the study's *Area* (geographic location). All articles, regardless of the test subject, need to discuss the *Bio-data* of their experiment, along with the *Procedures*, *Materials* and *Test*s to be followed, used and undertaken. *Statistics* are also mentioned in most articles, regardless of subject type. However, the *Rationale for experiments* with DNA, bacteria, plants, and cells seems to be self-evident, so it is less frequent.

Similar patterns emerged in the Results section. However, unlike the Introduction and Discussion sections, the Results section has a strong relationship between the Moves that appear and the subject of the experiment. The Chi-squared test resulted in a p-value of p =<0.016.

Table 4.18: Correlation of observed Moves and Steps with different subject types in the Results sections

Subject Type	Outcomes	Evidence	Figures	Subjects	Procedures	Previous	Statistics	Analysis	Summary
Human	103	103	101	63	63	13	86	83	18
Cell	50	50	50	8	38	12	37	49	7
DNA	38	37	38	1	31	20	31	31	16
Animal	33	33	33	6	26	13	28	27	8
Data	17	17	17	7	13	12	15	13	5
Bacteria	6	6	6	1	5	3	5	5	1
Plants	3	3	3	1	2	2	3	2	0
Totals	250	249	248	87	178	75	205	210	55

The proportions table that follows helps to show the Moves that are found for each type of experimental subject.

Table 4.19: Proportions of Moves and Steps in the Results sections as they relate to subject types

Subject Type	Outcomes	Evidence	Figures	Subjects	Procedures	Previous	Statistics	Analysis	Summary
Human	1.00	1.00	0.98	0.60	0.60	0.15	0.86	0.81	0.17
Cell	1.00	1.00	1.00	0.16	0.78	0.28	0.82	0.70	0.14
DNA	1.00	0.97	1.00	0.08	0.82	0.58	0.82	0.79	0.42
Animal	1.00	1.00	1.00	0.15	0.79	0.42	0.91	0.82	0.24
Data	1.00	1.00	1.00	0.41	0.76	0.29	0.88	0.82	0.29
Bacteria	1.00	1.00	1.00	0.17	0.83	0.50	0.83	1.00	0.17
Plants	1.00	1.00	1.00	0.33	0.67	0.67	1.00	0.67	0.00

From the table it can be seen that, regardless of subject, an MRA's Results section needs to include the Steps of having an *Outcome*, with *Evidence*, *Statistics*, *Analysis*, *Figures* and a *Review of the procedures*. Only articles about humans need to *Review the subjects*. Also, only those articles that cover DNA, plants and bacteria need concern themselves with discussing previous studies in the Results section. As for the occasional Move of *Summary*, the only subject type that uses it and then less than half the time, is DNA.

In the Discussion sections, a different picture emerges. There was no significant correlation between subject matter and Steps that appeared. The p-value from the Chi-square test was $p = <0.9886$, or almost 1 (meaning that there is no relationship between the Moves and the subjects, so regardless of subject, the Moves will appear).

4.4 Comparison of present work with studies by others: an overview

Each of the previous researchers (Skelton, Nwogu, Kanoksilapatham, Li and Ge, and Fryer) held slightly different notions about what comprises a Move (see 4.1.6), so it is not surprising that there are subtle differences between our findings. The following sections will look at the different findings by author across each section and suggest possible reasons for any major differences.

4.4.1   Comparison: Introduction section

The Moves identified in the Introduction section (see Table 4.3) were quite similar across all the studies of Moves in the corpora of research articles (RAs). Although many of the Moves and Steps may have had somewhat different names, they had similar underlying constructs. For example, all of the researchers identified either a Step or a Move in which previous research was discussed in terms of the gap identified in the study or the problem the study was designed to solve. Li and Ge (2009) followed Nwogu's (1997) framework; Skelton (1994) did not use the idea of Steps or elements in his description of the Moves. The current research was the only one to identify the description of the findings of the experiment as a Step, although there were relatively few instances of this Step (66 instances in my corpus). Since my corpus contained a larger number of texts in which to identify Moves or Steps, and also held the most recent articles, any emerging trend might be more readily seen.

As part of what were only slight differences in nomenclature, it may be that the few minor differences in the frameworks offered by each researcher would be due to the impact of Swales' (1990) work on the field. His CARS framework has been so influential, that it would be hard to not be affected or influenced by his original findings.

4.4.2   Comparison: Methods section

A major different between my findings and those of the other researchers whose frameworks I have been considering is that the current research and that by Fryer (2012) are the only studies to identify a Move for *Ethics approval*. Granted, this is generally an optional Move, but it becomes obligatory if the research study involves animals or humans. In the case of my research, this could be due to the size of my corpus, which is four times larger than those of the other researchers, so that such a Move would be more likely to surface in my corpus. Another difference comes from problems with distinguishing between Step and Move. For example, Kanoksilapatham (2005) establishes an optional Move called *Details of equipment* while three of the other researchers consider that to be a Step. The difference may be due to her corpus

being comprised of biochemical research articles: perhaps the discourse community for this discipline places greater value on identifying the specific tools being used. In addition, all of the other researchers divide descriptions of experimental procedures and of statistical procedures into two Steps, while in the current research I consider these to be Steps in a larger Move focused on procedures of any kind.

4.4.3 Comparison: Results section

This section shows variance among all authors for the identification of Moves and Steps, and for their number. Nwogu (1997), Li and Ge (2009), and Fryer (2012) all identify a single obligatory Move in this section, which concerns observations, and is more in line with the familiar description of this section being one that presents only the results. However, Skelton (1994), Kanoksilapatham (2005), and the current research all find three or more Moves. In the present study, statistical and observational results are divided into two separate Moves, reflecting the subtle rhetorical functions within this section, to persuade the reader of the validity and credibility of the researchers and of the study, as keyed to different kinds of findings. In addition, both Kanoksilapatham (2005) and the current research are the only studies to identify a Move of *Review* as a precursor to justifying a chosen methodology. The current research also identifies an optional Move of *Summary*, which, when present, gives an overall view of the findings. Because this Move occurs so seldom, and because the articles in my corpus reflect publications from 2001 to 2011, it is possible that this finding signals a small but growing trend in medical journals.

4.4.4 Comparison: Discussion section

This section also demonstrates variance in the identification of Moves and Steps. A major difference is in the various ways of classifying Moves and Steps, suggesting the subjective nature of this kind of research. All researchers identified a Move related to discussing the overall research outcomes or main findings. However, the current research study is the only one to identify a Step for secondary, or distal, findings.

While Skelton (1994), Nwogu (1997), and Li and Ge (2009) identified Steps for describing the significance of the results, I identify this as an optional Move, which may be based on recommendations from medical journal editors about avoiding a claim of priority (ICJME 2014b, p.15). It is in the analysis of Moves in the Discussion section that the widest variance between the present study and that of others occurs, in that I identify a greater number of Moves and Steps in the Discussion section. This is probably because, as I coded shifts in information and noted the presence of discourse boundary markers such as discourse markers or key words and key phrases, I was working across 250 discussion sections from 250 articles. Their number may have presented a wider range of topics, clinical and experimental problems, and subject types than studies working with smaller samples.

4.5   Conclusion

While the study of Moves can be considered subjective, it is nonetheless important to attempt to stratify how a genre is rhetorically structured. The current research identified seventeen possible Moves, and includes three optional Moves, each of which has a rhetorical function, specific grammatical structures and characteristic diction. Attempts at Move analysis based on corpora are acts of description, not of prescription, and as such, will present wide variety in findings as genres change over time, no matter how slightly. A genre analysis cannot be completely replicated, if done on a different corpus of texts. However, such a study could serve as guidelines for a near-replication since there should be enough information about the texts in the corpus for another researcher to build the same corpus, and enough detail on methodology to enable him/her to replicate the research. As additional tools and methodologies are developed that support the analysis of increasingly large corpora, the task of genre analysis should become easier. One of the limitations of this research was the lack of inter-coder reliability, described in 4.1.5.4: this remains a future task.

There are many reasons why, over time, there may be shifts in the Moves that appear in a genre. According to Li and Ge (2009, p.102) 'Genres are capable of modification over time in response to socio-cognitive needs of the users and genres will change with the modification of the discourse community and its members' perceptions of the

world'. As discussed in Chapter 3, many members of the sample felt that the Internet is having and will have an even more prolific effect on the genre (see 3.3.6). As the genre changes over time, journals will change their 'house styles' to better reflect the needs of their readership. An example of this is urging authors to not claim priority, or the way in which certain journals, like *The Lancet* (see 3.3.7), now require authors to include a section within their discussion explicitly placing their research into context of previous studies. The search engine, Google, has changed how researchers and readers alike access articles: it was not uncommon in the past to see something along the lines of 'We were, to the best of our knowledge, the first to … ' in papers. However, in 2014, with a range of search engines and services like *Science Direct*, and the National Institutes of Health's own warehouse of articles, *PubMed,* finding other examples of research on a particular topic has become much easier, and thus the likelihood of being the first and being able to claim priority has become more unlikely.

The current research had two goals. The first was to present a recent snapshot of the genre of medical research articles, following the initial framework by Swales (1990). The second was to investigate the rhetorical grounding underlying the Moves. It is hoped that this research will help novice contributors understand the genre and its expected components. Understanding the diction and the reasons behind writing eludes many second-language writers. Accordingly, the following chapter will discuss word choice in MRAs. Understanding the metastructures for the rhetoric underlying Moves may be just as important.

In his study of everyday English texts, Jordan (1984, p.20) claims that most texts follow a similar metastructure in answering these questions:

> What is the situation?
> What is the problem?
> What is the solution?
> How well did the solution overcome the problem?

Jordan's questions and his metastructure look very similar to Swales' CARS framework, with one exception: the last question in Jordan's metastructure is evaluative, while Swales' CARS framework at first seems to lack the element of

evaluation. However, when I review the Moves I described for the Introduction section, there is an optional Step (*Preview the findings*) in Move Three *(Way to the solution)* that has an evaluative component, as it is a summary finding, and thus gives the reader a preview of the outcome of the work.

Jordan begins by explaining the framework of ordinary English texts as being made up of four kinds of information, which parallels the structure typical of fiction. These are *Situation – Problem – Solution – Evaluation*, which Jordan sees as parallel to the *Setting and Exposition – Complication – Resolution – Dénouement* of a play or novel. Swales and Feak (2004, p.59) use a similar four-part framework in *Academic Writing for Graduate Students*. In scientific texts, the IMRD framework offers another four-part parallel. That is, the Introduction in a scientific article establishes both the situation and the setting, which, as the noted anthropologist and linguist Dell Hymes (2003, p.55) has explained, is the necessary initiator of any communicative act, be it spoken or written. The evaluation or dénouement in a scientific article is typically carried out in the Discussion section, which is both evaluative and persuasive in nature. An article may also include what Jordan (1984, p.110) calls pre-evaluation, in which the effectiveness of the solution is suggested before its amplification, which can be seen in the optional Step in the Introduction section as well as in the Discussion section.

Hymes' view of communication was influenced by his long acquaintance with the rhetorician, Kenneth Burke (Jordan 2005), and with Burke's concept of rhetoric as being 'the use of words by human agents to form attitudes or to induce actions in other human agents' (Burke 1969, p.40). Both persuasion and evaluation are involved in Burke's well-known theory of identification:

> As for the relation between 'identification' and 'persuasion': we might well keep it in mind that a speaker persuades an audience by the use of stylistic identifications; his act of persuasion may be for the purpose of causing the audience to identify itself with the speaker's interests; and the speaker draws on identification of interests to establish rapport between himself and his audience. So, there is no chance of our keeping apart the meanings of persuasion, identification ('consubstantiality') and communication (the nature of rhetoric as 'addressed'). (Burke 1969, p.46)

The reader of an MRA is practising evaluation throughout the process of reading, whether doing so in a linear order or skipping from one section to another and even omitting one or more sections. Even in the act of choosing to read an article, there is some evaluation. The readers of an MRA are looking to fill clinical or research needs. The doctor or scholar who choses the article seeks either information that can be immediately adapted to clinical practice or, in the case of research, can either support or refute their work, which is a form of identification.

How readers locate the clues beyond headwords that signal Moves or Steps is primarily by the identification of predominant variations in the basic IMRD pattern (including tense and voice), and by its characteristic lexemes. While they may not be particularly aware of the rhetorical functions underlying Moves and Steps, or may not notice specific diction, such as the authors' use of reporting verbs, those word choices work to suggest evaluation on the part of the authors, and expand the persuasion directed to the readers.

4.5.1   Summary

This chapter began with a definition of a Move as a shift in information flow which has a communicative function; its overt boundaries are signalled by headwords or phrases, spacing and paragraph indentation and its other, less openly signalled, boundaries are demarcated by content and key words or phrases that act as discourse markers, including coordination or subordinating adverbs. I then presented a brief bibliographic overview of studies of Moves in individual sections, and provided brief descriptions of five other research studies of Moves in all sections of RAs and MRAs: Skelton (1994), Nwogu (1997), Kanoksilapatham (2005), Li and Ge (2009), and Fryer (2012). The section concluded with a tabulation of Move frameworks, adapted from Swales (1990), by these authors and the present study, highlights of the differences in Move assignments, and a description of the rhetoric underlying the Moves.

Next, I discussed in turn the Moves and Steps located in the successive major sections of each article: Introduction, Methods, Results and Discussion, prefacing each section with a comparative tabulation of Moves found across five previous studies and the

present study. Comparing findings allowed me to differentiate my own framework for deciding what constituted a Move and its components. In addition, each Move and Step that I identified was individually discussed for grammatical patterns of voice and tense, and for most characteristic and frequent lexical patterns. While the use of the passive voice in my corpus continues the trends illustrated by Swales and Feak (2004) in Table 4.6, my adaptation of their Table in 4.3.1.2, about tense usage in the MRA corpus, diverged from their findings in some interesting ways, primarily in the Discussion section. Swales and Feak note that present tense occurs in the Discussion section with high frequency, and simple past tense occurs with mid frequency. However, in the MRA corpus, simple past tense is predominant in two Moves, present tense predominates in three Moves, the two tenses vary in one Move, and in the Move *Present a rationale*, present tense and future constructions with modals are used.

The sequence of individual discussions and exemplifying excerpts from texts supported the identification of rhetorical functions for individual Moves and Steps identified in the present work. This section also examined Moves according to whether they were optional or obligatory, and looked closely at potential associations between Moves/Steps and the type of subject used in the study, i.e. human, animal or other. While an example of a complete article's analysis would be helpful, there are many complications with regards to copywrite. The corpus is composed of articles in the public domain, but only as an agreement between the original journal and pubmed and any agreement that I have with pubmed may not extend to the whole article meaning that I could only show small pieces of each article.

It should be noted that the analysis of Moves and Steps and their discussion were not based on a sample of the MRA corpus, unlike the five previous studies, but instead on the full corpus, which contained 250 articles. I then summarized the comparison of the present work with the five previous studies, working section by section, to identify the relative amount of variance in Move identification. A limitation of the present study is that, unlike Kanoksilapatham (2005), inter-coder reliability was not conducted, leaving that a task for the future. The focus of 4.4 was on the twinned rhetorical goals of persuasion and evaluation by comparing work by Jordan (1984) on what he calls the four-part structure of ordinary text (*Situation – Problem – Solution – Evaluation)* to other four-part structures Jordan identifies, including that for fiction

(*Setting and Exposition – Complication – Resolution – Dénouement).* The latter model is linked both to the ethnographic approach to communication of Hymes (2003) and the motives for language use identified by Kenneth Burke (1969). Burke's emphasis on persuasion, in which hearers/readers are led to identify with some part of a text and its producer, can lead in turn to closer analysis of diction, embracing both evaluation and persuasion.

Chapter 5: An examination of thirteen reporting verbs in medical research articles

5.1   Introduction

A key feature of academic writing is an author's reporting and offering commentary on previous work on a topic or subject of investigation, both by himself/herself as well as by others. A major expectation for academic writing is its being situated within the larger discourse community, and to do this, authors must refer to the works of others (Hyland 2002, p.115). Reporting information goes beyond the attribution of citations to include incorporating verbs chosen to report on research. Reporting verbs such as *state* or *claim* are used to report ideas, discoveries, or findings by self or other researchers. They may be used to mark what words and ideas are taken from other researchers and which findings come from the current work. If describing works by others, the authors might say, 'Previous studies have found', while if they are reporting a finding of their own, they could say, 'Our data have shown'. They are a part of the evaluation the authors offer throughout a text. Part of professional and scientific writing is the appropriate use of sources and citations, for which it is necessary to use reporting verbs, clauses and phrases. A reporting phrase can be described as the combination of an adverb of manner with the reporting verb, which combination can then suggest a judgment on the findings being reported, such as *weakly attempted, strongly refuted,* and *minimally confirmed* (Tseronis 2009; Biber & Finegan 1988; 1989).

Earlier studies of reporting verbs by Thompson and Ye (1991) and Thomas and Hawes (1994) are the most frequently cited by researchers on scientific and medical writing. Some recent additions to the field include Bloch (2010) and Jirapanakorn (2012), who develop their own corpora to examine how reporting verbs are used by different groups of writers. Bloch focuses on comparing papers in *Science*, one of the top academic journals, with those written by students, while Jirapanakorn analyses papers written in English by Thai medical professionals. Both researchers categorize reporting verbs according to their functions but do not examine any modality or colouring of the word in further defining the verbs.

The challenge of understanding the pragmatic meaning behind a word has been observed by others, such as Bloch (2009; 2010). Many novice professionals and second-language writers have difficulties with identifying or using reporting verbs, and with having a sense of their pragmatic meaning, such as knowing the differences between those that are factive and non-factive (Tadros 1985, p.5). Basically, the choice of a factive verb suggests that the authors think that what they are reporting is true, and that of a non-factive verb intimates that the authors believe what they are reporting is not true. Factiveness also distinguishes between the choice of reporting verbs used for reporting data and those for reporting thoughts, as discussed below (see 5.1.3).

Citations are another aspect of reporting information. According to Small (2010, p.192), since the early twentieth century the need and number of citations and references used in an academic article has continuously increased and the reason, 'in part, emerges from the drive for priority and originality which entails demonstrating that others have not done before what you have done'. As described in 1.1.1, one of the problems facing less experienced or novice second language authors of medical research articles (MRAs) is how best to report on work done either by others or themselves. As discussed in Chapter 4, authors of MRAs also need to utilize rhetorical approaches, particularly those that persuade the reader that the authors are experts and can offer something of value to the field. Part of achieving this goal is discussing others' work and framing the current research as being unique or as having priority in terms of newness, relevance, or some other characteristic, and part of that framing is the choice of reporting verbs. According to Biber et al. (1999, p.196), presenting information or findings is done by means of reporting clauses, which are 'direct reports of somebody's speech or thought'. Section 4.1.7 of Chapter 4 described how rhetorical studies and genre analysis could be intertwined in the exploration of possible reasons for word choice of reporting verbs, and how they can show if and when authors support or agree with a particular work.

This chapter will employ corpus-based analysis to categorize selected reporting verbs according to whether the verbs in a specific context indicate disagreement or agreement with the findings of others or, as I will later argue, by being factive,

counter-factive or non-factive. I will examine 13 verbs, utilizing the lenses of semantic prosody (SP) and lexical priming (LP), to separate the verbs into those that show agreement with the work being reported on and those that do not. The verbs are then further categorized by whether they are being used to describe work done by the authors or focusing on the work of others. The prosody of the verbs will be examined as their use may help uncover possible colouring in a particular usage. For example, several of the reporting verbs examined in this study present a negative or a contrastive colouring, which can intersect with their factiveness. I will also use LP to investigate whether the choice of reporting verb can be affected by its context. I will describe and illustrate how, through the use of LP, authors of MRAs may be drawing on formulaic expressions with regard to word choice as I seek to explain reasons for the selection of certain words and the stances they may index.

One of the major issues in the analysis of reporting verbs is that they seem to have an evaluative function that is not inherent in the word's definition, but is instead part of its prosody. Tadros (1985) demonstrates that when an author reports on something, the text is primed for an evaluation by the authors' words regarding the information or ideas being reported. This evaluation indicates the writer's stance (Englebretson 2007).

Hyland (2012) states that stance is equivalent to voice; therefore, understanding how or why authors are reporting something can 'encode' the author's own stance. When authors report something, they signal their own bias or stance about the topic. As discussed in 4.1.7, I suggest that medical research articles (MRAs) are persuasive texts by nature, and in reporting work by others, their authors need to show how their work is in agreement with a previous study or how it is different, and must evaluate previous work accordingly, thereby presenting an evaluative stance. The assumption that reporting verbs contribute to showing the author's stance has underscored the persuasive nature of MRAs and contextualized the following research.

5.1.1   Stance and reporting verbs: a brief overview

Stance is an established area of study in discourse analysis (Englebretson 2007; Jaffe

2009), especially with regard to academic writing and speaking. According to Junqueira and Cortes (2014, p.92), to look for evaluation and stance, a researcher needs to look at 'words, collocations, and phrases,' which I do in this work using corpus analysis along with SP and LP (see 5.3).

Hunston and Thompson (2000, p.5) define evaluation as 'a broad cover term for the expression of the speaker or writer's attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about'. Stance is the authors' feelings or attitudes about the information being presented (Hyland 2012, p.148). Thompson and Ye (1991) assert that reporting verbs go beyond stating authors' findings but also can be employed to show the authors' attitudes towards what is being reported. Hyland (2008, p.70) describes how often those who are non-native English writers (NNW) have trouble understanding differences among reporting verbs, leading to 'fuzziness' and perceptions of ambiguity on the part of the reader. This means that while an NNW or novice academic writer may be able to choose a reporting verb that is grammatically acceptable, it may lack an appropriate rhetorical meaning. While part of the problem may be a lack of academic vocabulary, it can also be due to the ways in which some English words are polysemous and can carry pragmatic meanings that are not fully understood by a NNW (Hyland 2008; Kecskes 2014).

According to Hyland (2005), stance is one of the two ways writers can show evaluation:

> This can be seen as an attitudinal dimension and includes features that refer to the ways writers present themselves and convey their judgments, opinions, and commitments. It is the way that writers intrude to stamp their personal authority onto their arguments or step back and disguise their involvement. (p.176)

Hyland also presents caveats for authors on how to persuade (Hyland 2014), warning, for example, that even though the author may agree or disagree with something or somebody, it is ultimately up to the readers to decide whether they feel the same as the author (Hyland 2014, p.3). The interpretation of facts is never simple and is 'filtered' through many lenses (Hyland 2009, p.299). In addition, one of the major

differences between expert and novice writers is the level of criticism manifested towards other's works in that the more experienced writer is likely to exhibit a stance of disagreement more forcefully than a writer who is new to academic writing (Hyland & Guinda 2012, p.7). I follow Hyland (2005, p.175) with my claim that it is the act of persuasion through the choice of rhetorical devices that helps authors to convince readers, specifically: 'to understand what counts as effective persuasion in academic writing, every instance of evaluation has to be seen as an act socially situated in a disciplinary or institutional context.' Reporting verbs are one of the ways through which authors can situate their evaluation of other research, whether by themselves or by others.

Stance has been examined by numerous researchers in different areas of functional and corpus linguistics and has been given a variety of names reflecting slight differences in emphasis, such as: 'affect' (Ochs 1989), 'appraisal' (Martin 2000; Martin & White 2005) 'evaluation' (Biber & Finegan 1989; Hunston & Thompson 2000; Biber 2006), 'evidentiality' (Chafe & Nichols 1986), 'footing' (Goffman 1981), 'hedging' (Hyland 1998), 'intensity' (Labov 1984), 'point of view' (Simpson 2003), and 'positioning' (van Langenhove & Harré 1999). In each case, an effort is being made to show authorial confidence in information about a topic, issue or person and some form of evaluation or action toward whichever it is.

According to Jaffe (2009, p.9), stance at its most basic is a comparison between the views of author/speaker and others, in a juxtaposition that shows either agreement or disagreement:

> Stance saturates talk about others, in which speakers engage in both explicit and implicit forms of social categorization and evaluation, attribute intentionality, affect, knowledge, agency to themselves and others, and lay claim to particular social and/or moral identities. (2009, p.9)

She further argues that stance can characterize patterns that emerge in a collective rather than on an individual basis:

> The linguistic systems indexed by stance are all embedded in political, social, ideological, and cultural fields of action. (2009, p.13)

Her discussion is highly relevant to my present work, as I have attempted to show that the sections in an MRA have different stances inherent in their composition. For example, since the Introduction section is primarily where authors will attempt to mark their work as different from that of previous scholars (see 4.3.1.1.2), it typically presents a negative stance and the use of counter-factive reporting verbs associated with the earlier works being reported. In the following example the authors are using *claimed* as a counter-factive reporting verb of controversial claims made by previous studies.

> (1) Effective models that enable comprehensive prediction of health protective behaviours remain limited mainly to two overlapping theoretical paradigms: the Theories of Reasoned Action/Planned Behaviour (TPB) [14-16] and Banduraís concept of self-efficacy [17 -19] (the belief that one can successfully execute some behaviour), particularly regarding the core TPB concept of perceived behavioural control, which controversially is claimed by some to be largely synonymous with self-efficacy [19-21] and by others to be indistinguishable from intent [22] (the intention to execute a particular behaviour), the key predictive element of TPB[16]. (Text 249)

Thompson and Ye (1991) suggest that reporting verbs in research articles (RAs) can show stance; however, there are several differences between their work and the current study, particularly in relation to the time period covered by data, the disciplinary area, the size of the collection, and the lexis. Thompson and Ye's work (see 5.1.3) contains articles from a wider variety of RAs, while the current study focuses exclusively on MRAs. Thompson and Ye (1991, p.366) analyse a corpus composed of over a hundred RAs from a variety of disciplines, including linguistics, geology and engineering, but not medicine. Their study begins with an interest in how evaluation may be expressed with reporting verbs, as based on Sinclair (1985) and Tadros (1985), but quickly establishes that evaluation is far more complex than simple agreement or disagreement (1991, p.380). Their focus is on ways to identify the denotation of the reporting verbs; although that focus is excellent, and this study incorporates their findings whenever appropriate, this study also utilizes SP and LP to help determine authorial stance from reporting verbs, as stance can also suggest rhetorical function of the verb in the context of its passage.

5.1.2   Reporting verbs and averral

Citations to sources in academic writing can identify existing knowledge about a topic or subject, and can thereby suggest different aspects of evaluation (Bednarek 2008) by presenting attribution. Citation can be used 'to integrate the ideas of others [….] or to point out the weaknesses in others' arguments, aligning themselves with a particular camp/school/grouping' (Thompson & Tribble 2001, p.92; see also Thompson 2002). As Swales (2014, p.119) comments, citations manifest both dialogism and intertextuality, and 'also operate rhetorically to strengthen arguments and claims in various ways'.

Citation is more than a listing of sources of knowledge attributed to prior work. Thompson (2002) distinguishes between attribution and averral: attribution identifies information prior to the current study being read or proposition being advanced; averral is the current author's implicit promise to the discourse community that what they say in their study is true

According to Tadros (1993, p.101) averral is the 'default condition' of all information being presented. That is to say that unless authors explicitly state that they do not agree with something (attribution), they are supporting what is being reported. This further supports the claim by Teufel and Moens (2002, p.413) that every work mentioned carries a judgment from the authors citing them.

5.1.3   Using the construct of factiveness to classify reporting verbs

Reporting verbs are prevalent throughout the different sections of medical research articles, although they vary in frequency and purpose according to the section in which they occur. Since it holds 250 articles reporting medical research, the present corpus should enable an investigation of the functions of reporting verbs in this area of scientific writing, and thereby increase comprehension of how corpora can assist in understanding genre.

For this study, I chose to select and categorize thirteen reporting verbs on the basis of their factiveness. Tadros (1985) first introduced the idea of categorizing verbs based on their factive and non-factive qualities. Factiveness is a slippery concept: averral is truthful reporting while factiveness is the belief that something is true: these *could* be two different things. For example, I can aver what a study says, but the study might be statistically flawed, which would be counter-factive. Originally, Tadros (1985) defined factiveness as being when the authors agree with the information being put forth and believe it to be true, while non-factive means they disagree with it and believe it to be untrue. Thompson and Ye (1991, p.372) expanded the definition of factiveness when they added a third category, and their three categories will be used in this study:

> *Factive*: the writer portrays the author as presenting true information or a correct opinion; for example, *acknowledge, bring out, demonstrate, identify, improve, notice, prove, recognize, substantiate, throw light on*.

> *Counter-factive*: the writer portrays the author as presenting false information or an incorrect opinion; for example, *betray (ignorance), confuse, disregard, ignore, misuse*.

> *Non-factive*: the writer gives no clear signal as to her attitude towards the author's information/opinion; for example, *advance, believe, claim, examine, generalize, propose, retain, urge, utilize*.

Field (1997) draws on a corpus of spoken British and American English to explain the connection between factive constructions and stance:

> In factive constructions, the juxtaposition or comparison is between the epistemic stance of the proximal speaker with the epistemic and/or affective stance of the grammatical subject. Thus, factive constructions contribute to the multidimensional nature of the indexical ground of a speech event, as they signal information at two levels (i.e., referentially as well as indexically), about participants' stances towards the propositions contained in *that*-complement clauses. (1997, p.800)

In essence, the factive construction can be epistemic, pertaining to the confidence the author has in the information, such as *discover, know, realize*, or it can be affective, such as *regret, resent, be surprised* (Field 1997, p.802). An epistemic predicate indexes the speaker/writer's certainty about the information in the complement clause

of the predicate (Field 1997, p.803) and can also index the speaker/writer's stance in relation to the person or thing referred to by the grammatical subject.

Sinclair, Louw and others have claimed that words can have a positive or negative prosody. When considering MRAs, these two choices seem limiting, especially since I am treating the words as having a local prosody. For example, Stubbs (1995, p.250) described *cause* as having a negative prosody since it often collocates with negative words like *cancer*. However, the purpose of an MRA is to identify a problem (usually a negative topic from a lay perspective) and detail what others have done with the problem. Negative prosody is thus not necessarily a fair characterization of the words, so that the prosody may need additional description along the continuum from positive to negative. This study will note examples of negative prosody wherever appropriate; however, examining SP in terms of factiveness may also deepen our understanding of how a word is primed.

Factiveness has been used as a construct in several fields. An example would be Wyse's (2009) work in computer programming that primarily focuses on helping to create computer software that could use human language to ask and respond to queries in programs to help tutor students. As a computer programmer, Wyse recognizes the considerable complexity in trying to code language (Wyse 2009, p.39). Without claiming any particular familiarity with corpus linguistics research methodologies, he does, however, conduct a corpus-based analysis by choosing to work with the entire data set across multiple disciplines from the academic discourse collection, *OpenLearn*, a tool connected to The Open University.[24] His research is accomplished by writing a software algorithm to perform factive/non-factive recognition. Wyse was enabled to do this by using a list of factive and non-factive verbs produced by Hooper (1974), which Wyse then expanded by using a thesaurus. The expanded list allowed him to determine frequency of occurrence for factive/non-factive indicators and thus analyse overall impact.

Wyse's study is perhaps the most recent, thorough look at reporting verbs in a corpus and genre. His classification of verbs is not built on his location and

---

[24] Listed in 2014 as http://www.open.edu/openlearn

identification of the verbs from a single genre, but from two sources: Hooper's listing (1974) and *The Oxford American Writer's Thesaurus*. He then develops a list of verbs and writes an algorithm to locate the verbs in text. Wyse's findings derive from his algorithm and do not take SP or LP into account, which explains part of the differences in our methodology and our occasional differences in findings. In addition, our works looked at different genres, which may (see 5.3.2) explain differences in our findings.

Wyse is concerned with factiveness as it is related to question generation that is answerable. Wyse (2009, p.38) shows the difference in factiveness with the verbs in two sentences: I *think* that X… and I *know* that X… Here, *know* establishes the that-clause as a true, or factive, predicate. Accordingly, his Question-Generating System had specific design aims (Wyse 2009, p.42):

> These design aims were that the system would focus on single sentences only, and that the answer to the generated question would be contained within the sentence. In fact, the system would also allow the generation of the answer in addition to the question.

The remainder of his study focuses on mathematical approximations of certainty in the truth, or factivity, by the system's production of questions and answers. While my work for the most part has confirmed his findings, there are some differences in how we code words as belonging to classes of factiveness, as I use the three categories Thompson and Ye (1991, p.372) implemented in their study whilst Wyse uses two categories. Wyse's dichotomy does not allow for the fact that some verbs can fall into the category that Thompson and Ye (1991) identified as counter-factive, those verbs that have an oppositional stance to the source being reported in the work by the authors. Although my findings and Wyse's are not truly comparable, since he is working with a different genre, his classifications of factiveness for specific verbs are highly useful, and I will draw on them.

5.2   Using corpus analysis to gather information about word choice

McCarthy and O'Keeffe argue that corpus linguistics is a 'means to an end' as well as 'an end in itself' for interacting with large collections of text in order to explore and analyse words (2010, p.7). They further add that corpus linguistic techniques or tools enable a discourse analysis to be 'better assisted' (McCarthy & O'Keeffe 2010, p. 7), and that utilizing corpus tools is a mechanism that allows researchers to examine large bodies of text in a more efficient way. Moon (2010, p.197) stresses that while a corpus does not tell us how many words there are within a language, a corpus can tell us which words are used. This is an important distinction, which enables corpus linguistics to offer commentary and describe not only which words are being used but also some notion of why they have been chosen. Numerous studies have looked at word frequencies and how words appear, such as those by Sinclair (1987; 2004b), Stubbs (2001), and O'Keeffe, McCarthy and Carter (2007). Sinclair (2004b, p.101) stresses that the meaning of a word cannot be found by looking at an individual word but must be discovered by looking at the word in context, regardless of its frequency, as in a different context the word may change its meaning. Stubbs (2001, p.35) argues that connotation of words is important as it gives an indication of which community the user belongs to; that is, different communities may use terms differently and it is through a word's collocation patterns that we can start to understand its meaning to that group.

For example, non-native student writers (as well as native speakers) often misuse verbs of high frequency (Altenberg & Granger 2001); more recently, I have noticed the misuse by my Taiwanese university students of the term *prove* in their writings: they often write that an author has 'proved' something with their findings, when actually the findings may only have been listed. According to Thompson and Ye (1991, p.366), the non-native language user may not understand all the weight the different reporting verb carries. Academic writing is typically cautious writing, and a word like *prove* has no tentative qualities. If the author of an MRA chose to use the word *prove*, invariably it was used in my corpus not as a finite reporting verb but rather in its infinitive form, where the proof is something to be achieved, or accompanied by some kind of moderation, such as *may*. For example:

> (2) Firstly, targeting a specific group of women *may prove* difficult when scaling up this intervention to programme conditions. (Text 191)

O'Keeffe, McCarthy and Carter (2007, p.68) describe a word's 'fingerprint', that is, the clues that the collocation gives to the word's meaning and likely user. A word by itself, taken out of context, could not give us the necessary clues to discern the meaning, the community the author belongs to or any other pragmatic information such as attitude or stance toward the topic (O'Keeffe, McCarthy & Carter 2007, p.77). The word needs to be looked at in context. The current study addresses the function of different reporting verbs in the corpus but with only brief discussion of frequencies for the reporting verbs and an emphasis on how they are used as opposed to how often.

5.2.1   Word choice in medical research articles

Tribble (2002, p.133) outlines ten components of his framework for doing genre analysis by utilizing a corpus. He separates the framework into two parts: *contextual*, which is more concerned with function, and *linguistic*, which looks more at the form and language. The basis for the work in this chapter was derived from his framework because his three steps for linguistic analysis, the examination of lexico-grammatical features, text relations/textual patterning, and text structure, give the focus needed to conduct the linguistic analysis of reporting verbs. I will chiefly follow the linguistic part of his framework, and will limit the construct of context to the IMRD section in which the verb is used.

Table 5.1: Framework for genre using corpus tools (adapted from Tribble 2002, p 133)

Contextual analysis: Steps	Description	Linguistic analysis: Steps	Description
1. Name	What is the name of the genre of which this text is an exemplar?	8. Lexico-grammatical features	What lexico-grammatical features of the text are statistically prominent and stylistically salient?
2. Social context	In what social setting is this kind of text typically produced? What constraints and obligations does this setting impose on writers and readers?	9. Text relations/textual patterning	Can textual patterns be identified in the text? What is the reason for such textual patterning?
3. Communicative	What is the	10. Text structure	How is the text organized

purpose	communicative purpose of this text?		as a series of units of meaning? What is the reason for this organization?
4. Roles	What roles may be required of writers and readers in this genre?		
5. Cultural values	What shared cultural values may be required of writers and readers in this genre?		
6. Text context	What knowledge of other texts may be required of writers and readers in this genre?		
7. Formal text features	What shared knowledge of formal text features (conventions) is required to write effectively in this genre?		

In Step 8, 'the examination of lexico-grammatical features', Tribble means the identification of words or phrases that occur statistically either frequently enough or too little to warrant further investigation. That is, if a specific word or phrase occurs more frequently than others throughout the text, this may need to be accounted for, while if a word or phrase only appears once, but was expected to appear more often, it also needs review. Either case asks for further exploration, and would need a reference corpus (see Table 2.1) with which to be compared. Even without the reference corpus, however, as in this study, word choice can be examined.

Tribble's next step, the examination of text relations/textual patterning, expands to include collocations: how the word connects to other words in the text, and whether there is a reason for this pattern. It is here that the constructs of SP and LP, as developed by other scholars, become useful in analyzing first the co-text, i.e. words that appear near the node, and then the context.

The last part of Tribble's lexicogrammatical framework is the analysis of text structure, which examines the role the word plays in the larger picture of the text. For example, when a reporting verb occurs in the introduction section of an MRA, it can be hypothesized that the verb is being used to show disagreement with previous studies, as the authors will use the verb to employ a rhetorical perspective in order to

distance or set apart their work from previous work or that by others. Key words typify nearly every Move. For example, in Chapter 4, key words were identified for Moves and/or Steps, as in Move Two (see 4.3.1.1.2) where *needs, shortcomings, however, determining, discovering, reason*, and *problem* suggested the emphasis of the Move. Specific reporting verbs in Move Two include *hypothesized, predicted* and *anticipated*, in which authors discuss how others had set up parameters for the issue, problem or topic being investigated.

5.2.2   Textual patterning: What it means to be formulaic

The formulaicity of textual and lexical patterning as well as the specific and non-formulaic lexis that is inherent in medical writing is one of its major characteristics. The definition by Wray and Perkins (2000, p.1) is useful: formulaic language is 'a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (see also Durrant & Mathews-Aydınlı 2011). This definition, when overlaid onto the ninth step in Tribble's framework (2002, p.133), illustrates how the identification of textual patterns can include formulaic language.

As with other scientific writing (Hyland 2008; Allen 2009), MRAs have a formulaic quality due to their number of repeated collocations and lexical bundles, and it is that assumption that lays the groundwork for seeing that the reporting verbs being used are part of formulae. For example, looking at Tadros' (1985, p.25) claim that a reporting verb signals a need for some evaluation in the text, then we may see the verb as fitting a pattern that has been previously engrained in the author's writing repertoire. This is similar to Becker's (1975, p.73) category of *sentence builder* from his taxonomy of forms, where Becker gives this example: Person A reports on Person B's findings, which can be mapped onto reporting in this way:

> (Previous study) (reporting phrase) (topic of findings of previous study) and (possible evaluation). (Becker 1975, p. 73)

An example is:

(3) Previous studies using the BI-RADS scale reported moderate agreement, with kappa statistics of 0.43 to 0.59 for intra-observer studies [16,17]. (Text 13)

Note that when I use the term, *reporting phrase*, it will typically refer to the reporting verb, its headword(s) and its auxiliary verbs.

Depending on whether the reporting phrase is used by the author to evaluate a finding, there may be an evaluative aspect to the sentence, unless, as Tadros says, the author feels the statement is in fact a truth or as she terms it, an *aver*. An example of a reporting phrase is *attempted to demonstrate*. In this phrase, we find the headword *demonstrate* combined with *attempted to*, so that the combination is used to show some modality or evaluative qualities of the phrase. Using Tadros' idea that the reporting verb signals some degree of evaluation, I suggest that if the author uses *attempted to* in the reporting phrase, then the author may be presumed to be evaluating work by the previous scholars who 'attempted' something, and I see the evaluation here as being at least slightly negative (X tried but did not succeed; X tried but did not finish; X tried and was prevented by Y from succeeding; and so on). An example from my corpus is:

(4) Qiu and co-workers attempted to resolve this issue using a Bayesian model that allowed ancestral inheritance of introns, gain of introns and loss of introns (intron gains and losses were assumed to be completely reversible). However, this model used the unrealistic assumption that the sites actually occupied by an intron in at least one family member comprised the total set of protosplice sites in which multiple independent gains of introns could occur without restriction. (Text 143)

The first sentence in this excerpt shows that a previous study had tried to demonstrate something using a particular model, but the second sentence is the authors' comment about why and where the other researchers had failed in their attempt.

5.2.3   The co-occurrence of formalized structure and formulaic language

Durrant and Mathews-Aydınlı (2011, p.58) make the point that formulae are created within a discourse community over time, through usage and repetition. They further stress that the use of formulaic patterns can help indicate who belongs to a particular community. Those who are part of a community use jargon specific to their community: after using a certain pattern so often, the use of a formulaic phrase (*in such instances*) or prefabricated sentence frame (*factual claim highlights importance of topic*) becomes similar to the use of jargon. For example, the first sentence of many MRAs is a statement of fact regarding a medical issue:

> (5) Childhood asthma is a serious and growing health concern, with prevalence rates at a historic high and over 9 million children in the U.S. diagnosed in their lifetime [1,2]. (Text 015)

The factual statement helps set the context for the following paper, using a formalized sentence frame (Wray 2000, p.465) that has come to represent to the discourse community that the author is working within a formalized framework.

Stating a hypothesis and assumptions is a fairly common technique, but MRAs tend to bypass stating that a hypothesis or assumption is important, and instead transition into treating the statement as a fact, such as:

> (6) Once the native state is achieved, the protein is believed to be released from its interactions with the chaperone(s). (Text 186)

The findings from the study by Durrant and Mathews-Aydınlı (2011, p.63) are especially appealing here, because they apply the construct of formulaic language to Swales' (1981) idea of Moves, tying formulaic language to their claim that a Move is itself a formulaic structure. It could then be argued that the language used in stating a fact at the outset of an MRA is used to satisfy Swales' first Move of describing background information, and that factual statement, then, becomes or initiates the formulaic structure.

The extension of formulaic language to include Moves as well as phrases can support Sinclair's (1991, p.100) idea of the 'idiom principle': that is, that writers use an obligatory Move (such as the factual statement at the outset) or a pre-constructed

phrase that is the accepted norm of the community. However, deciding which term to use from the viewpoint of the idiom principle means the author needs to understand the word or phrase's pragmatic undertones. According to Wray (2002, p. ix), the usage of formulaic phrases is a major problem for non-native English users, even though chunks of formulaic language may be how learners first interact with language. However, as Kecskes (2014, p.83) comments, 'knowing the 'core' (dictionary) meaning of an expression is one thing, but knowing both the core meaning and the pragmatic, socio-cultural load attached to that expression in a given culture is another matter'. Both formulaic and idiomatic phrases are pragmatically rich and are so nuanced that it is difficult for those not familiar with the community or the language to use them effectively (Kecskes 2014). It becomes necessary to look at the words and phrases and study their pragmatic qualities: we are thereby led to examine discourse prosody as one way to study the words and phrases and to better describe them. Such examination will involve us with various aspects of lexical priming.

5.3   Semantic prosody and lexical priming

A word is defined by several features, including its grammatical function, its collocates, and its larger context. Defining a verb based on its function may help suggest how the word is partially being used: that is, the use of one specific word or phrase will establish a need for another word. After using the processes of lexical priming (LP) and semantic prosody (SP), there may be more involved in the selection of a word than just a word's grammatical function. The authors may also be searching for a way to use their voice or signal their stance.

5.3.1   The perspective of semantic prosody

Since the appearance of Louw's (1993) work on the subject, SP has become a frequent form of lexical analysis in corpus linguistics. It has been used to study words across a variety of genres and in numerous studies on word selection. Its importance to corpus analysis is, as many have found (Kay 1995; Moon 1998; Channell 2000),

that the evaluative qualities of words can only be found through empirical research and large amounts of data are needed for their discovery. SP is defined by Louw (1993, p.157) as when two words often appear together and one of the words 'colours' the other: that is, if a word often collocates with another word that is positive or negative, the word will be imbued with similar feeling. However, colouring can usually only be noticed by a native speaker of that language, which speaks to the 'hidden' qualities of the word that have occurred from the pairing.

According to Bednarek (2008), a researcher begins examining SP by assuming semantic preference: when two words often appear together, they become 'married' in the collective readership. Bednarek (2008, p.122) stresses that semantic preference looks only at collocation, while SP needs the researcher to infer and critically analyse the word to see how and why it is being 'coloured' semantically.

Sinclair (1996, p.78) urges researchers using corpus linguistics for text analysis to look closely at SP. One of the major criticisms of it, however, is that SP is typically used to label words as having only two possible prosodies: positive or negative (see Louw 1993). As Hunston (2007, p.256) states:

> Such meaning is often not reducible to a simple 'positive or negative'. It is essentially linked to point of view, so that there is often not one indisputable interpretation of attitude.

However, Morley and Partington (2009, p.141) argue that the reason for using SP to define words in terms of only two conditions is based on human nature: humans like to separate things into two terms, good or bad. The present study will employ a multi-faceted characterization: instead of using negative and positive as the descriptors, as done by other researchers such as Wyse (2009), Hooper (1974), and Williams (2004), I will emphasize the use of *factive, counter-factive* and *non-factive* (following Tadros 1985 and Thompson & Ye 1991) to describe qualities of the verb for reporting stance. This will allow me to incorporate the notion of authorial stance, in the sense of using the choice of verb itself to suggest a point of view, which can include support for or disagreement with the findings of others.

Another major criticism of SP (Whitsitt 2005, pp.285-286) is that SP is too similar to

connotation and collocation to warrant a separate treatment. Morley and Partington (2009, p.143) rebuff this claim by saying that while both collocation and SP deal with Sinclair's construct of 'co-selection', SP looks at how meaning changes due to its collocation and context. Sinclair posits that the meaning of a lexical item has five components, which include the '*core*, which is invariable, and constitutes the evidence of the occurrence of the term as a whole, and the *semantic prosody*, which is the determiner of the meaning of the whole' (1998, p.15). Accordingly, if authors choose a term, they will want one that naturally fits together with others (collocation) but also can express or suggest their feelings about the subject, which can indicate authorial stance. An example might be that an author is reporting ideas from previous studies by other researchers, and therefore needs an appropriate word and a correct grammatical structure. Accordingly, the author may choose from several reporting verbs, such as: *argued, claimed, stated, believed*, all of which are non-factive, or they may choose *suggest* or *discover* which are factive, as defined by Tadros (1985, p.32). The verb needs to be in the past tense to follow the genre-based constraint on the word, but there is also a pragmatic issue of how authors show evaluation of the idea held by previous scholars, something that is part of the act of reporting upon others' work (Tadros 1985, p.5). If authors agree with the idea held by previous scholars, they will want their reporting verb to be factive and will want to choose a word that feels the most factive in its prosody as it will then be most likely to reinforce the truthfulness of research and the veracity of the authors, hence contributing to the persuasiveness of the work. As my research suggests, they will often choose the word *indicate* (see 5.4.2.5)*.* However, if the authors want to show disagreement or disbelief, they will need to select a word that is non-factive, so they might choose *claimed,* which suggests tentativeness and the potential for falsification. Note that in such instances, positive and negative co-text may intersect with the factiveness. It is in the selection of the verb, which reveals their stance toward previous work, that SP can be seen in relation to co-selection. Though SP can offer a motivation for the selection of a word, the fact that the reporting verb is needed comes from the context, and as such it is primed for use.

When, through repeated use, word A becomes associated with word B, it becomes affected by its collocate. Subsequently, even when word A is not used with word B, it has nonetheless typically been coloured by some of the meaning of its 'partner'. An

English example, which differs by country, is *mastermind.* In American English, this word is often seen in collocation with the noun modifier *criminal*, a negative word so that even when used on its own, *mastermind* often continues to have a negative connotation. When looking at the term *mastermind* in the Corpus of Contemporary American English (COCA) (Davies 2008), the first 20 instances of *mastermind* all relate to stories of criminals or terrorists. However, when *mastermind* is entered into the web version of the British National Corpus[25] the word relates to business leaders and notable scholars as well as to the UK game show *Mastermind*. Although the two corpora may be slightly skewed by having specific types of texts, the example illustrates that SP, much as Tribble (2002) described, may be affected by factors such as genre or diatopic variety.

5.3.2   The perspective of lexical priming

Hoey (2005) describes the idea behind LP thus:

> As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context (Hoey 2005, p.8).

While Tadros (1985) initiated the idea that word-prediction in text is keyed to the occurrence of other words, Hoey (2005) takes it farther. He develops the concept and coins the term, *lexical priming* (LP). In LP, a word is primed not only by its relationship to other words, but also to grammatical needs and the context in which it is being used. When selecting a word, an author is primed by previous history with the word according to at least one of the following ten criteria for priming, the first of which is similar to the basis on which SP is identified:

1. Every word is primed to occur with particular other words; these are its

---

[25] (http://bnc.bl.uk/saraWeb.php?qy=mastermind&mysubmit=Go) Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

collocates.
2. Every word is primed to occur with particular semantic sets; these are its semantic associations.
3. Every word is primed to occur in association with particular pragmatic functions; these are its pragmatic associations.
4. Every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations.
5. Co-hyponyms and synonyms differ with respect to their collocations, semantic associations and colligations.
6. When a word is polysemous, the collocations, semantic associations and colligations of one sense of the word differ from those of its other senses.
7. Every word is primed for use in one or more grammatical roles; these are its grammatical categories.
8. Every word is primed to participate in, or avoid, particular types of cohesive relation in a discourse; these are its textual collocations.
9. Every word is primed to occur in particular semantic relations in the discourse; these are its textual semantic associations.
10. Every word is primed to occur in, or avoid, certain positions within the discourse; these are its textual colligations. (Hoey 2005, p.13)

A word needs to fit into its communicative aim, and in the case of reporting verbs, that is the reporting of data or information. Hoey's (2005) study of LP provides clear examples and in-depth analyses of how LP occurs from multiple positions. After looking at lexical and semantic aspects of word choice, he examines word choice first pragmatically, then grammatically, and follows that by discussing how LP affects synonyms. Hoey (2005) illustrates the effect of LP on polysemous words, i.e. words that have two or more meanings, and shows that even the rare meanings of a word are similarly primed to the most commonly accepted definition. Hoey (2005, p.83) gives the example of *consequence*, as it has two meanings. The first is the more common meaning of result, i.e. the consequence of not doing your homework is that you will fail the class. The second and less common meaning is that of importance, for example: 'It is of great consequence that you complete the mission'. Hoey argues that even though the word is polysemous, the rarer form still will be primed as if it were the usual form.

Priming can also help in understanding what words are not appropriate in a particular context. Recently I assigned my second language learners the task of writing about a chart that shows that for every hour spent studying, exam scores went up exponentially. One of my students wrote, 'The graph clearly shows that the time spent

will obviously affect the scores of the student on their grades'. I felt the choice of the word *obviously* violated Hoey's third condition. Pragmatically, this word was poorly chosen, in that native speakers are taught to use modals and other language features to show tentativeness about their findings, as in 'the time spent probably had/could well have had a major impact on the scores'. This kind of issue may not be apparent to a second language learner because if they accept only the first and most frequent definition in learner dictionaries, the word *obviously* is primed for use when there is no doubt as to the causation.

A second example comes from a junior high school student in Taiwan who said he was doing poorly, because he had 'cats and dogs homework'. Without a background in scholarship, he was trying to follow Hoey's principle that a word can be primed through its collocates. He had taken the idiom 'raining cats and dogs' to mean that cats and dogs can replace the idea of *heavy* or *lots,* and instead of stating he had a heavy load of homework, he dressed it up and tried to use an amalgamation of the idiom and a different head word, in this case with *homework* replacing *rain*. However, while the new idiom he coined was creative, it upset Hoey's first principle of semantic priming, by going out of the expected set of collocates, and thus could not work in terms of the desired register. LP can be used to help show what reasons an author may have for selecting certain words and in this study, can identify which of Hoey's criteria are met in research on reporting verbs.

Both SP and LP are closely tied to collocation (Sinclair 2004b). Word choice may be affected by what Tribble calls Local Prosody (2000, p.86). Stewart (2010, p.17) gives examples of local prosodies found in other scholars' works that had been previously considered 'sector-specific' prosody. That is, the genre can also affect word choice, as described earlier in this chapter by the example of *mastermind*. For example, the term 'rhetoric' as used in the genre of news reporting usually presents a negative connotation, as in 'the politician's presentation of the issues was mere rhetoric'; however, the use of rhetorical techniques in both news reporting and academic writing is both common and positive. The present research will be focused on only the LP and SP of selected reporting verbs found in the genre of MRAs.

5.4   The analytic process

It is important to look for the potential influence of either SP or LP by beginning with concordances as it is Hoey's first criterion for looking at LP. According to Hoey, it is through identifying a word's collocates that the affects of SP can be examined. To create concordances and key word lists, which calculate the frequency of lemmata in a corpus, I chose WordSmith Tools (Scott 2008) because, as Hoey (2005, p. xi) himself notes, it is a program that allows a researcher to interact with a corpus in a variety of seemingly painless and efficient ways, such as reviewing collocates and co-text through the Concordance feature described below.

After entering my corpus of 250 MRAs into the WordSmith Tools (Scott 2008) program, I then used the Concordance feature that allowed a search of all my collected texts for a specific term or phrase. The Concordance feature produced a spreadsheet showing the term as the key word in context (KWIC). This is a way of looking at a search term (type, or 'node') in the concordance program with the key word centred so as to see the patterns created by the other words, its co-text. This feature allows the user to click on the example, and retrieve the entire surrounding text, in order to examine the text further for collocates. For example, from a longer concordance of *argue*:

```
1. Further, Zhou and Leydesdorff both argue that China is
2. results from the two-hybrid screen argue for an interaction
3.  shelterin components, our results argue for an association
```

In the above sample, *argue* is the node. This small glimpse of the node *argue* suggests that *argue* can collocate with both *that* and *for* and is a colligate with that phrase. It would be hard to assign a prosody from the immediate collocates that surround the node in the above example, or to claim that *for an, that China, our results,* or *two-hybrid screen*, have any evaluative qualities. In his study of SP in works by the author James Joyce, Stewart makes a persuasive argument that the two word boundary set down by Sinclair is too restrictive and that on occasion, one must look at a larger portion of the text to find the prosody (2010, p.115). Indeed, the reader will need additional context to discover whether *argue* is positive in nature. In addition, according to Hoey (2005, pp.81-82), the colligates of a word transfer to its different

forms, so that if the noun form of the word is contrastive then the verb form will be as well, or the reverse. This is of interest for future research, since many of the reporting verbs I examined have nominal counterparts.

5.4.1   The thirteen selected reporting verbs

To accommodate space restraints, my sample is limited to 13 verbs. In order to have a range from commonly used to less often used verbs. I selected 12 of the verbs from a list of reporting verbs from the British Academic Word List (Gillett 1999).[26] I added *prove*, which was not on Gillett's list, because in teaching second language writing, I have found that my students either misuse or overuse *prove*. I wanted to see how that verb was utilized in my corpus in an effort to better inform my students, and to give an example of how writing instructors might examine specific word use by their students.

Table 5.2 lists the 13 reporting verbs I have chosen for analysis, with their level of frequency according to Cobb's *VocabProfile*[27] program based on Nation's word frequency lists (Laufer & Nation 1995), as well as the number of instances of each word in its base, past/past participle and third person singular present forms within the MRA corpus. The third column indicates the verb's factiveness as defined by Wyse (2009, Appendix B).

I chose these verbs based on the following criteria: first, the most frequently used reporting verb (*find*) and then the least frequently used reporting verb (*point out*) in the corpus as identified by simple frequencies across the total of all sections, to show the effects of SP and LP on a wide range of verbs, those more likely to be used, and those less likely. Next, I chose both four factive and nine non-factive verbs (as originally designated by Wyse [2009]), as I wanted to have a mixture with which to perform the analysis to determine the priming or colouring based on the verb's function. I added *point out*, *say,* and, as noted above, *prove,* as those are three verbs

---

[26] www.uefap.com/writingforapurpose/vocabulary/reporting_verbs.doc
[27] www.lextutor.ca/vp/eng/

frequently used by my second-language students, hoping to substantiate my claim that professional medical writing tends either not to use them or to use them only in certain circumstances. I also wanted to include words from different levels of frequency of use. Laufer and Nation (1995) categorized words as belonging to different levels: a word that is one of the thousand most frequent words in English is said to be classified as being in the K1 category; a word found in the second thousand most common words is classified as being in the K2 category. There is also a category called Academic which is used to list the 550 most common academic words (Coxhead 1998) and a fourth is the category called Other, i.e., a word that is not on any of the other three lists (see also Morris & Cobb 2004). Accordingly, I included verbs from both the K2 and Academic categories.

Table 5.2: Thirteen words to be examined

Word		Frequency	Wyse's Coding	Current works coding	Base	Past tense	3rd person present	Totals
1.	Argue	K-2	Non-factive	Counter-factive	33	19	6	58
2.	Claim	K-1	Non-factive	Counter-factive	13	11	13	37
3.	Discover	K-1	Factive	Factive	9	30	0	39
4.	Find	K-1	Factive	Non-factive	103	1158	11	1272
5.	Indicate	Academic	Non-factive	Non-factive	242	264	153	659
6.	Observe	K-1	Factive	Factive	66	1094	1	1161
7.	Point out	K-1	Non-factive	Non-factive	3	8	1	12
8.	Prove	K-1	Factive	Factive	9	26	3	38
9.	Report	K-1	Non-factive	Counter-factive	237	812	144	1193
10.	Say	K-1	Non-factive	Non-factive	12	28	1	41
11.	Stress	Academic	Non-factive	Non-factive	299	47	14	360
12.	Suggest	K-1	Non-factive	Non-factive	330	214	284	828
13.	Think	K-1	Non-factive	Counter-factive	33	95	0	128

It would not be feasible at this point to tag all reporting verbs in my 250-article corpus, particularly since it is their function in context that is crucial. For example, the verb *collapse* would not typically be seen to act like the verb *indicate*, as it is often associated with some kind of physical structure, as in *the bridge collapsed* or with a metaphorical structure, as in reporting and evaluating the structure of an argument, as in *Their argument collapses when the genome is reviewed* (Text 183). Nor do reporting verbs form a substantial part of lexical studies in MRAs based on Coxhead's Academic Word List (Chen & Ge 2007; Wang, Liang & Ge 2008), although they are one of the components in a study of finite/non-finite verbs in four clinical as opposed

to experimental MRAs where, not surprisingly, they occur more frequently for both categories in the Results and Discussion sections (Williams 1996).

Some words such as *say* have a built-in quality of being related to people, since *say* typically needs a human subject. Collocates of *say* can be identified within the Concordance program of WordSmith Tools (Scott 2008), which allows the researcher an overview of the word's possible prosody. However, the word's location within a particular section of the text might have some bearing on the word's use and its collocates. Swales (2012) demonstrates how language use in a research article can be affected by appearing in a particular section; for example, text in the Methods sections contains more passive constructions than other sections. He adds that another factor in explaining which words are chosen is how 'senior' the scientist is; by this, Swales means that those researchers who are established in a field are also more comfortable with using the discipline-expected language.

A number of the reporting verbs I have selected have a high frequency in the total corpus: some of the words, such as *find/s/found* occur more than a thousand times. I conducted a random collection from different sections of the corpus, to help ensure a representative sample. The total size of the sample varied by the number of instances of a word as it occurred in my corpus. For example, there were 214 instances of *suggested*. Selecting every sixth example from my corpus provided 33 occurrences across all four sections of IMRD, which allowed me to mirror the proportions of the reporting verbs in each section. *Suggested* only occurred 13 times in the Introduction, 22 times in the Methods, 75 times in the Results and 104 in the Discussion. By doing an every nth selection I am still able to show that there were more occurrences in my corpus from the Discussion section, thus preserving proportionality across the different sections (IMRD) of the text. For the analysis, I chose 33 instances of a verb as being the threshold for trying to implement any categorization. According to statistical advice I sought, any sample over 30 approaches a normalized distribution, meaning that one can have a confidence of 96% in any finding (p<.05). [28] While 50 - 100 instances would provide a result with greater confidence, many of the verbs in my study did not have those numbers of occurrences, so I used *argue,* the first verb I

---

[28] Dr. Michael McEwan, University of Glasgow

looked at which had 33 instances, as a lowest common denominator. Since I was not conducting a primarily statistical analysis, 33 would be sufficient to assure confidence. The following table shows the number of instances in the corpus of each form of the thirteen words, the size of the sample I looked at and the number of times it was a reporting verb along with the percentage of sample that was a reporting verb.

Table 5.3: Observable data of the thirteen words

Word	Total of instances in the corpus	Size of sample	Number of reporting verbs	Percentage of usage as a reporting verb in the sample
Argue	33	33	23	70
Argued	19	19	12	63
Argues	6	6	6	100
Claim	13	13	7	54
Claimed	11	11	11	100
Claims	13	13	1	8
Discover	9	9	2	22
Discovered	30	30	13	43
Discovers	0	0	0	0
Find	103	33	16	48
Found	1158	33	32	97
Finds	11	11	0	0
Indicate	242	33	27	82
Indicated	264	33	29	88
Indicates	153	33	33	100
Observe	66	33	19	58
Observed	1094	33	24	73
Observes	1	1	0	0
Point out	3	3	1	33
Pointed out	8	8	8	100
Points out	1	1	1	100
Prove	9	9	0	0
Proved	26	26	1	4
Proves	3	3	3	100
Report	237	33	14	42
Reported	812	33	32	97
Reports	144	33	2	6
Say	12	12	4	33
Said	28	28	28	100
Says	1	1	0	0
Stress	299	33	0	0
Stressed	47	33	0	0
Stresses	14	14	1	7
Suggest	330	33	26	79
Suggested	214	33	27	82
Suggests	284	33	33	100
Think	33	33	30	91
Thought	95	33	24	73
Thinks	0	0	0	0

5.4.2 Classification of thirteen reporting verbs

This section discusses 13 reporting verbs as they are found throughout my corpus. I will address any possible colouring of the words, such as factiveness conferred by the co-text, using SP, and propose reasons for their selection using LP. Tables for individual verbs can be found in Appendix 4, Tables 1-13. They will be separated into one of three categories based on the classification by Thompson and Ye (1991, p.372): factive, where the authors' verb choice agrees with and/or avers the information or claim; counter-factive, where the authors' verb choice shows disagreement with the information or claim; and non-factive, where there is no clear judgment expressed by the verb choice towards the sources being described. In addition, I will also investigate if the discourse level has any impact on the stance, i.e. whether or not the reported information is coming from the field (previous studies or commonly held beliefs) or the current experiment.

5.4.2.1 *Argue*

According to Wyse (2009, Appendix B), *argue* is non-factive, but I would instead classify it as being counter-factive based on the limited number of instances found in the corpus of MRA. *Argue* does not occur in the Methods section of any text in the corpus in any form. Only four times in the sample does *argue* report something by a named scholar; that is, only four times does the name of the source appear with the reporting verb: the verb is not used only for reporting. In nine instances of the 33 listed in Table 1A in Appendix 4, *argue* was used in a presuppositional clause along with a modal to indicate a hypothetical situation and its possible outcome, as for example:

> (7) As with conventional meta-analyses, some will *argue* that we have not compared like with like. Our model, however, was based on relative treatment effects (differences between groups expressed as effect sizes [23]), and variations in patients' characteristics between trials are fully accounted for in the analysis by maintaining randomised comparisons within each trial. (Text 77)

This is almost exclusively done to create a straw man that the author can use to promote the researcher's own stance or work. The word is used mostly in the Introduction and Discussion sections, to describe both the field and the experiment. The past tense form *argued* appeared 19 times in the MRA corpus (Appendix 4, Table 1B), and it is used as both a reporting verb and as part of a presuppositional clause. When used in a presuppositional clause, it is used to advance or protect the author's own work. It also follows a similar pattern to *argue*, being primarily used in the Introduction and Discussion sections. The only difference between the two forms was that *argued* was primed as a reporting verb used more frequently for disagreeing with the field, and for presenting an evaluative stance negating what was being reported:

> (8) It has been *argued* that the Wolbachia that induces parthenogenesis may be functionally restricted to organisms which have a haplodiploid mechanism of sex determination [15]. (Text 45)

A disagreement with the field can be illustrated by the previous example, in which the authors state the previously held convention regarding Wolbachia, but then add the following phrase, 'Whereas we now have many demonstrations of Wolbachia-induced parthenogenesis in haplodiploids . . . ', thus differentiating their work from the field.

Argues occurs six times (Appendix 4, Table 1C) in the corpus used for the current research and follows similar patterns to *argued*, inasmuch as it appears mainly in the Introduction and Discussion sections; however, with only six examples it would be rash to make any claims regarding possible priming or prosody.

5.4.2.2 *Claim*

The word form *claim* appears 37 times (Appendix 4, Table 2A) in the corpus: 31 of the uses are reporting verbs with seven uses of *claim* in the base form, 11 as past/past participle and only 13 occurrences as a third person present singular verb. According to Wyse (2009), Hooper (1974), and Williams (2004), *claim* is a non-factive verb, which raises the question of whether negative words have coloured the SP of this word. Some collocates include *although, in spite of, in sharp contrast* and *but,* (to introduce a counter-claim)*, controversially, despite* (to introduce a contrastive phrase)

and *seems*. Of the 31 uses, approximately half were part of negative collocations, given repetition of words such as *not*. The following example of the verb form shows how an environment of negativity is achieved through the repetition of the word *not*.

> (9) These potential problems are not relevant for our study since we do not *claim* that certain bubbles would be used by the observers to solve the classification task on full fields, whereas other bubbles would not. (Text 174)

In this sentence, the first two uses of *not* seem to draw a strong line between the researcher and those who might question the author's results. It can be argued that due to the verb forms *claim/claims/claimed* collocating so often with negative words or in negative contexts, it is counter-factive, and therefore authors would not be expected to use it to report their own findings or ideas. Looking at whether the verb is being used in the author's own experiment as opposed to when the author is reporting on the field enables us to understand why authors would use it about their own research.

When authors use the reporting verb *claim* in connection with a respondent or participant in their experiment, they appear to use *claim* almost as self-protection:

> (10) Most of the respondents *claimed* that the modern breed is poor in disease and stress tolerance (86%) and in the ability to escape predators prevalent in their village conditions (96%). The modern breed generally required higher level of management (83%) often hard to afford and are poor scavengers (86%) compared to indigenous chickens. In addition, 77% of the farmers in Horro and 90% in Sheka *claimed* that hatchability of eggs obtained from the modern breed is inferior to eggs from indigenous chickens. Likewise, most of the respondents have the opinion that the eggs (90%) and meat (92%) obtained from modern breeds have poorer taste (Table 6). This was also confirmed by the lower market preference for eggs from exotic chickens. In the opinion of 98%, 74%, and 93% of the total respondents pooled over all regions RIR chickens were rated superior in egg production, meat yield, and egg size, respectively, to the indigenous chickens (data not shown). (Text 62)

Using *claimed* is one way an author can question the results found by others and thereby suggest how thorough, in contrast, is his/her own research, as shown in the next example. Here, *claimed* is used to show contrast between perception and reality: between the perception that the participants feel they have been trained to become skilled and the reality that they are not skilled or trained:

(11) The observed poor knowledge is in sharp contrast to the 93% who *claimed* they received AMTSL training either during pre-service education, or with in-service training. (Text 41)

In the following example the authors do not provide names of those who presumably (derived from the use of the adjectival form *claimed)* did the work with which they disagree, but instead assert that recent studies have improved the situation:

(12) Recent studies stress the importance of using alternative test procedures (e.g. direct RNA transfection) in conjunction with a combination of sensitive RNA analysis for discerning IRES-containing sequences in eukaryotic mRNAs (41,49). In fact, several previously *claimed* IRESs activities were discovered to be due to promoter activity present in the same region of the 50-UTR (37–40). (Text 144)

From these examples, we see that authors may use *claim* to distance themselves from the answers given by a respondent, or to show that they do not wholly agree with what someone else reported. Since in professional writing authors may only want to include information that they are sure of, that may explain why *claim* was only used 19 times as a reporting verb.

Patterns emerge in reviewing the data (Appendix 4, Table 2A). The verb does not occur in the Methods section. However, when the verb *claim* is used in the Introduction section, it seems to be primed for doubt about some aspect of the field, as in all instances throughout the corpus, the usage reflected findings of previous studies. In the Results section, *claim* is used only once to present the author's disagreement with the field, and the other five instances were describing findings in the experiment. However, *claim* is not being used to cast doubt on the actual experiment, but rather on what participants said in interviews or on questionnaires. There is no clear difference in the use of the verb in the Discussion section, as it can be used to evaluate both the field and the experiment. When considering the fact that one feature of the Discussion section is the Move *Limitations*, which is where *claim* is used to describe the experiment, its usage now becomes more understandable. The *Limitations* Move is the point where the author is offering a *meta-hedge* (my term), in which they are identifying a potential concern with the whole experiment.

5.4.2.3 *Discover*

Discover, which Wyse labels factive (2009, Appendix B), only appears in two forms, *discover* and *discovered*. The third person singular does not appear in the corpus; since *discover* is commonly an action verb, it is hard to use it for reporting, which has an inherent past tense quality. As *discover*, it is used eight times as the base form for the infinitive, the act of discovery, in the Introduction, Methods, Results and Discussion sections, where it can refer to either participants in an experiment, or authors of an experiment, as in:

> (13) To *discover* misconceptions about transmission of HIV/AIDS the women were asked to include/exclude ways that HIV/AIDS can or not can be transmitted. (Text 18)

> (14) Our study is the first to *discover* that humans cannot necessarily resolve model uncertainty. In our experiment, many participants failed to recognize the presence of unexpected uncertainty. Consequently, in the exit questionnaires they often took the arms to be ''random'' [in our language, risky] which illustrates the antagonistic relationship between risk and unexpected uncertainty – jumps were confounded with realization of risk. (Text 175)

In the past tense, *discovered* is used in reporting on what was done, not thought, or found, but actually as a result from a set of performed actions:

> (15) We also *discovered* a new intermediate phase of transport, phase IIb, wherein there is a large up-regulation of both plus- and minus-end directed run lengths in wild-type (Fig. 3C). (Text 198)

As for occurrence, what has been *discovered*/what authors *discovered* (active vs. passive constructions vary) occur in all sections, but least often in Methods (Appendix 4, Table 3B). In the Introduction section, where authors report the findings of other researchers, it occurs seven times. In the Results section, where the authors report their own findings, it appears 11 times, and in the Discussion section, where the authors report findings from the field as well as their own, it occurs nine times. Only three instances are found in the Methods section, where the authors are more likely to use the participle to describe some part of the experiment.

The above data may indicate the presence of examples of priming related to occurrences in different sections of an MRA. For instance, in the Results section the word is primed to describe the author's own findings, while in the Introduction section it is used for reporting upon findings in the field, which is again much like other verbs reporting on both the field and the experiment in the Discussion section.

As for SP, *discovered* has a significant collocation (according to WordSmith's collocation function) only with the words *of* and *the*; however, there are ten instances out of 30 where the word appears with a quantifier of time, meaning that it is collocating with words such as: *new, newly*, and *recently*. Such collocation suggests that *discovered* is often used to describe an event that has recently occurred at the time of publication. For example:

> (16) We also *discovered* a new intermediate phase of transport, phase IIb, wherein there is a large up-regulation of both plus- and minus-end directed run lengths in wild-type (Fig. 3C). (Text 198)

Discovery is an important function in the sciences and if the authors are highlighting recent discoveries such as a new phase, then it could be argued that they assume their original finding to be true and timely, which could be one reason why *discover* is categorized as a factive.

5.4.2.4 *Find*

Find in all its forms may be the most important reporting verb in MRAs: certainly it is the most frequently used. Wyse classifies it as factive (2009, Appendix B). The word *find* in its different forms occurs 1272 times throughout my corpus. Since there are 103 instances of the word in its base form, a random sampling of one out of every three produced a sample of 33 (Appendix 4, Table 4A).

In every instance in the sample of 33 where *find* is being used as a reporting verb, there is a contrastive element, regardless of location in text section or sentence. Sometimes the contrast is with previous research in the field:

(17) In contrast with [8], we did not *find* that any of the ten B. montana accessions tested shared a chloroplast haplotype with B. napus. (Text 232)

There can be a contrast with the author's own research:

(18) Previously we *found* that mature collagen is thermally unstable at body temperature [20]. We now find that propeptides, divalent ions, and ER-like crowding with proteins do not increase the stability of the triple helix. (Text 186)

Moreover, there can even be a contrast with the author's own expectations or a happy confirmation:

(19) We were pleased to *find* that as many as 76% answered that you can not tell by looking at a person if she/he is infected with HIV/AIDS. (Text 18)

Find can also be used as a reporting verb in an *if-then* conditional construction:

(20) If fitness is assumed to be a linearly decreasing function of the phenotypic distance to the optimum, then we *find* that the average fitness is given by…. (Text 184)

In the MRA corpus, *found* appears 1158 times; a random selection of every thirty-fifth instance was used to help ensure a relatively broad and representative sample of 33 data points. In the sample, *found* occurs primarily in the Results and Discussion sections, referring most frequently to the authors' findings from their experiment: *found* is typically a self-reporting verb.

Found, which frequently exists in negative contexts or constructions, has in those instances a semantic prosody of contrastiveness, causing it to be most often seen as counter-factive, as in:

(21) Some human myeloma cell lines likewise lack BMPR-IA and BMPR-IB [30]. On the other hand, a significant increase in BMPR-IB levels was *found* in malignant human glioma cells [31]. (Text 152)

When *found* as a reporting verb presents no element of contrastiveness or counter-factiveness, the authors are indicating a relationship between their research and that of others working in the field, expressed here as *these studies*:

> (22) Consistent with *these studies*, we *found* that CUP1, SOD1, FET3 and FTR1 were expressed at higher levels in the presence of 1 mM copper sulfate medium compared to rich medium (Figures 4, 5). (Text 139)

Both *find* and *found* are coloured with a contrastive element, but it is not consistently strong enough to code as either factive or counter-factive, thus leading to a categorisation of non-factive. Nevertheless, something interesting is happening, because not every instance of the verb has a human actor as its agent, which suggests this usage may be a way for the authors to shield themselves from criticism. By tying the reporting of the data to a thing, a model, a type of statistical analysis or some other method, the authors are taking the responsibility for the results out of their hands and placing it on some kind of mechanism instead. In my interviews with the discourse community sample, some doctors asserted they were not secure in their use of statistics and formulae; therefore, this use of *find(s)* can protect them from incorrect claims using statistics.

5.4.2.5 *Indicate*

As a reporting verb, *indicate* is more constrained by its actors, in the sense that there seems to be a data-based connotation in which the data are assumed to be factive and to speak for itself. According to Wyse (2009, Appendix B), *indicate* is non-factive. *Indicate* occurs 242 times (Appendix 4, Table 5A) in the corpus and therefore every seventh occurrence was selected to assemble a representative sample to analyse.

Indicate is almost always used as a reporting verb, even in the few instances of its use as an infinitive. Overall, its actors were predominantly related to data or results; for example, there was an instances of having respondents report their thoughts, which offers a way to show *indicate* as an infinitive used to report, as seen in this example:

(23) Respondents were asked to *indicate* their level of worry over the past one week about contracting influenza A/H1N1. (Text 249)

In nine of the 33 instances I found either a modal auxiliary verb or an adverb of certainty co-occurring with *indicate*, as in:

(24) Oceanographic data *indicate* that currents in near-shore regions are mostly wind-driven, and that the direction of the wind changes frequently [21,40], suggesting that gene flow among mudprawn populations in this region *may* mostly occur close to the coast. (Text 230)

The discourse level of this example is directed at the field, meaning that the information being reported comes from the field in general as opposed either to work by individual scholars, to the current research experiment or to the current research's findings (which identify the discourse levels), so there is no rebuttal of the findings of particular scholars in the field. In most instances, it is the author's own work that seems to be *indicating* something. This might be a way of tempering the analysis of the data in order to act as a buffer between the author and critics. Since the word usually collocates with data or results, *indicate(d)* typically focuses on the experiment reported in the article. *Indicate* appears to be primed to appear in the Results and Discussion sections.

In its past tense/participle forms, *indicated* appears 264 times in the MRA corpus, so I took a sample made up of every seventh example up to the maximum of 33 to ensure a fair representation and reflection of the word across all sections. *Indicate* was predominantly focused on the experiment and was found throughout the different sections, but not evenly, as the Introduction section had far fewer occurrences than the other sections. The actors were people only three times; the other times the actor was an inanimate entity, such as *data*. Only three times was there any sort of modality in the vicinity of the term. On occasion, as in the example below, *indicated* is primed to be the reporting verb for when the author wants to describe results from the process (here, the association with 'GRG') that may be ancillary to the main result of the experiment, which in this case was the identification of lab markers that were predictors of risk:

(25) Thus, the results of these unstratified analyses indicated that the posttest probabilities and LR estimates of developing AIDS associated with a low, moderate or high GRG were remarkably similar to those conveyed by strata of laboratory markers that are known to prognosticate a low, moderate or high risk of AIDS. (Text 196)

Indicates occurs 153 times (Appendix 4, Table 5C), 111 fewer than *indicated*, so this time every fifth word was collected. Every instance in the sample has *indicates* used as a reporting verb only, with no human actors. *Indicates* is primed as the reporting verb for third person singular situations for when authors want to tie the result to a process, instead of to themselves. There were only three examples suggesting semantic prosody of counter-factiveness, and there was only one instance of its occurrence with a tentative word, *may*, which does not support a claim for collocation. *Indicates* was primarily found in the Results and Discussion sections, where it typically focused on the current experiment. I define *indicates* as being non-factive, which is in agreement with Wyse's findings. It also has both counter-factive and factive collocates; an example of the word's being non-factive is:

(26) Recent information *indicates* that worldwide the percentage is even higher than previously thought, ranging from 30-39%. [3] (Text 41)

while an example of its being both negative and having counter-factiveness is:

(27) Lack of correlation with actometric findings in NIP subgroup *indicates* that tremor may not be the core feature of NIP. (Text 4)

In Example (26) the authors agree with the recent information and use it to stress that the situation is even more dire than previously thought, while in Example (27) the authors are using their findings to indicate that what was previously thought is wrong.

5.4.2.6 *Observe*

The next verb to be addressed is *observe*. On the surface, this appears to be an action verb as it reports the actions of the author or others observing a result. Wyse (2009, Appendix B) categorises *observe* as being a factive and it is found 1161 times throughout the corpus in its three forms (base, past/participle, third person singular

present)*, Observe* is found a total of 66 times, and therefore every second occurrence forms the sample. It appears 58% of the time in the sampling as a reporting verb and an example of it as a reporting verb is with the first person plural:

(28) We did not *observe* them in H&E stained sections (Table 2, R). (Text 87)

Usually the actor is the author, and this bears remembering for analyses of other forms of the word. Only once is there a counter-factive colouring of the word, with some notion of contrastiveness, which one would expect with counter-factive; it seems to be a straightforward recounting of actions or outcomes, and tends to support Wyse's (2009, Appendix B) classification of it as a factive. As for priming, since only two instances in the sample were not related to the current experiment, it seems to be primed for referring to the author's own observations. Because authors are unlikely to disbelieve themselves, this also supports the classification of the verb's being factive.

There are 1094 occurrences of *observed,* so every thirty-third instance was selected to develop a representative sample. *Observed* is primarily used as a reporting verb as it occurs as such 73% of the time in my sample.

Once again, the actor is predominantly the author, with a few instances in which other research in the field is the actor; in one instance, a scientific model was an actor. There was almost no evidence of the word's being linked through semantic prosody to negative or contrastive elements. The discourse is primarily at the experiment level but there was one anomalous use, in which the experiment was used to support the field and vice versa:

(29) The reduction in T4 *observed* in both sexes is consistent with previous studies showing that PBDE mixtures and single congeners can depress plasma T4 (Fernie et al. 2005; Hallgren et al. 2001). (Text 136)

Almost all occurrences happened in the Results and Discussion sections, which are more focused on the outcome of the experiment; accordingly, the word is primed for reporting on an author's own work as a factive.

In summary, *observe* and *observed* are factive and are primed when authors are reporting on their own research. These forms are predominantly found in the Results and Discussion sections, both of which contain a Move of stating primary and secondary findings (see Chapter 4), which have been *observed,* a verb choice which seems to be used to indicate confidence in findings.

5.4.2.7 *Point out*

Point out is actually a phrasal verb; when used as a reporting verb, Wyse classifies it as non-factive (2009, Appendix B). The verb appears in the MRA corpus in any form only 12 times (Appendix 4, Table 7A). When written in its base form, it only occurs three times, and the one time it does act as a reporting verb, it is reporting what the results found:

> (30) The results in Table 1 *point out* several interesting differences to the
> results obtained with our proposed method (Fig. 5). (Text 79)

There is no evidence of any colouring from collocates, and the focus is on the experiment. This example was located in the Results section. However, as seen earlier, *indicate* could possibly be used when the writer wants to report actions of a non-human; accordingly, *indicate* may be more typical of this professional register, and therefore more useful than *point out*.

Pointed out occurs 8 times and in every instance is used as a reporting verb. It has many different types of actor and is used both to report on the field and the experiment. Half of the examples presented a counter-factive and contrastive semantic prosody to the word, for example:

> (31) Other studies also *pointed out* the existence of different subgroups that
> often do not strictly follow the above-mentioned theories. (Text 97)

However, as there were only eight instances in the MRA corpus, it is difficult to extrapolate from these findings to a reliable picture of the word and its prosody.

Points out only appears once in the corpus:

> (32) Akers [2] *points out* that healthcare personnel, particularly those of childbearing age, are concerned about occupational exposures as they relate to fertility and pregnancy. (Text 26)

This usage is found in the Introduction and has a named actor along with a focus on the field. There is no apparent semantic prosody, but it is a reporting verb. The only possible conclusion to draw from this limited sample size is that the verb is infrequent. *Point out* is eschewed by writers of MRAs in this corpus. There may be a non-factive prosody for the word, but with so few instances of its occurrence in a 250-article corpus, such a claim would be unconvincing at best.

5.4.2.8 *Prove*

Prove is a common choice for my second language writing students. It is, according to Wyse, factive (2009, Appendix B). The base form appears only nine times in my corpus. In this example, *prove* suggests that something in the future may not be easy:

> (33) Firstly, targeting a specific group of women may *prove* difficult when scaling up this intervention to programme conditions. (Text 191)

Since there are only nine examples in total, it seems fair to say that *prove* is not regularly used as a reporting verb by writers of MRAs. *Proved,* however, is used differently: it is seldom used alone, and is primarily used to report when conjoined with *to be*. A concordance was created for the 26 uses of *proved* in the MRA corpus:

```
1     led mutant models—two of which  proved  to be of particular in
2     e PGE2 production by A17 cells   proved  to be sensitive to bot
3     ue dataset. Interestingly, pCR   proved  to significantly under
4     criptional analysis, A17 cells   proved  to be more highly rela
5     Table S2). However, A17 tumors   proved  to be more closely cor
6     ents Since the episomal system   proved  to be suitable for the
7     particle hydrodynamic diameter   proved  to be 91 nm. Fe2O3-NP
8     larly for cell lines that have   proved  useful as models for c
9     imensional grading system that   proved  to be useful in the ev
10    g in the stroma. A17-signature   proved, indeed, to be able to
11    MSCs and stromal phenotype and   proved  to be unexpectedly eff
12    epilepsy seizures [27]. The SL   proved  able to detect seizure
13    theta band SL on the first EEG   proved  to be a significant pr
```

```
14    diac arrest, where hypothermia proved to be beneficial even
15    ocial conflict in rodents have proved a biologically relevan
16    duit creation laparoscopically proved to be very straightfor
17    nd their transcripts. ESP also proved to be a powerful tool
18     a multidimensional score that proved to be useful in predic
19    imensional microarray data has proved problematic as it is n
20    he world [6,30,31]. It has not proved to be very effective a
21    henotypes of dyslexic subjects proved itself relevant becaus
22    e Netherlands, septic abortion proved to be rare. One death
23    owever, the generative updates proved to be too aggressive i
24    tioned serum (ACS, Orthokine®) proved slightly to moderately
25    of overall ED utilization also proved to be a solid basis fo
26    rveys [10,29]. Factor analysis proved that our results were
```

Of the 26, only the final example is a reporting verb, with 18 of the remaining 25 used as part of a phrasal construction, *proved to be*, a phrase which is often used for reporting. When either the verb or the phrase is used to report, it signals that the authors were right and therefore has a prosody of factiveness. It is also primed to report on the current research and is found in a Discussion section. *Proves* appears only three times in the corpus: twice in a Discussion section and once in an Introduction section. In all three of these instances, *proves* is used as a reporting verb; once again it is factive. Much like *point out*, the word is used so sparingly that no firm conclusions can be drawn.

5.4.2.9 *Report*

Report, the namesake for this class of verbs, is non-factive according to Wyse (2009, Appendix B), although I find it is also used to show counter-factiveness. It occurs 237 times and is used as a reporting verb 43% of the time in my sample. (Appendix 4, Table 9A)

In the sample, *report* is used as a noun more than half the time and as a non-reporting verb infinitive once. In its use as a reporting verb 14 out of the 33 times in my sample, it pertained to both the field and the experiment. One of the doctors interviewed (personal interview Doctor 3, May 20th, 2012) said that he would not repeat his results in the discussion, but instead, he would discuss how they could be used in the field. This is a much broader use of the data and his generalisation about typical usage seems to be supported by these findings for *report*.

Concerning SP, a prosody of counter-factiveness was found, inasmuch as the author often offered a phrase or word to show a difference between the current research and the field. In the following example the current research attempted to follow the protocol of other authors, but found that it did not work for their study:

> (34) Using this protocol and column, we could not clearly separate the oleic acid (18:1n-9) and alpha linolenic acid (ALA 18:3n-3) peaks thus we *report* these results as 18:1 + 18:3 ALA. (Text 233)

Reported may occur as a participle as well as the past tense formation. In my corpus *reported* appears 812 times, so the sample consists of every twenty-fourth instance with its being a reporting verb 97% of the time (see Table 5.3).

There were many actors, ranging from tests and findings to studies and results. The word was used in all sections of the MRA text, though most often in the Discussion section. With regard to the verb's reference to the field or the experiment, throughout the Methods and Results sections, *reported* was used to describe both; however, in the Discussion and Introduction sections, it primarily focused (all but once) on the field, as in some cases authors use the experiment's findings to justify the field:

> (35) They show that high-level features and spatial biases make the largest contribution in a mixture model, which is in line with the results *reported here*. (Text 174)

The *reported here* is referring to the author's own experiment, but it should be noted they are using the first clause of the sentence, which is about the field, to help describe the validity of their results. In the next example the opposite occurs:

> (36) However, lung injury caused by 50% of oxygen, as used in our ventilated mice, has not been *previously reported.* (Text 80)

Here, the authors are using *reported* to describe what the field has not found, but what they have discovered. The word *previously* that appears in the position immediately before the verb is interesting: a concordance of the word shows that *previously* co-occurs 73 times with *reported*, and according to the WordSmith Tools program (Scott 2008), in a statistically significant number of times (51 times) it appears as the word

immediately to the left of *reported*. *Reported*, then, is usually factive and signals previous findings that the authors wish to consider.

On the other hand, the key phrase, *to the best of our knowledge*, which can signal some contrast to be presented, appears in the MRA corpus within three sentences of the word *reported*, and 12 times overall. For example:

> (37) Because, to *the best of our knowledge*, the free concentration of inorganic cations in lysosomes had not been *reported*, the presence of potential luminal counter-ions had to be validated first. (Text 157)

This example suggests a semantic prosody of contrastiveness between the author's procedure and the reported procedures by other researchers working in the field, which in turn affected the author's subsequent methodology. This suggests that in the Introduction and Discussion sections, the word may be primed for describing the field and is not always factive, depending on how strongly contrastive the context may be. This may be an example of the effects of local prosody, and how in different contexts the prosody may change.

Reports occurs 144 times in the MRA corpus. Therefore, every fourth example was collected for the sample. *Reports* is only used twice as a reporting verb in the sample, so there is little data to show how it is primed, but what was apparent was that even when used as a noun it carried the semantic prosody of contrastiveness. I did not find whether the word class has any bearing on the semantic prosody, but from what I have found, which echoes Hoey (2005, p.81), if the verb is contrastive, then the noun form is also contrastive. The counter-factive quality of the word leads to an evaluative stance by the authors.

5.4.2.10 *Say*

Say occurred in only 41 instances in any form through the corpus (Appendix 4, Table 10A). According to Wyse (2009, Appendix B), it is non-factive. *Say* as the base form appears 12 times; it is used only five times as a reporting verb, By collocating with *can*, the verb phrase suggests some tentativeness:

(38) Taking into account the significant effects, we can *say* that the difficulty of the task is defined by the Speed, Interval and Range, and by the interactions Speed*Interval, Speed*Range and Interval*Range, and this relation can be therefore quantified by a quadratic model. (Text 118)

Idiomatically, it can be used to define something:

(39) Given a variable, *say* X, which can exist in one of KX states, and its set of associated probabilities {p(X = x1), p(X = x2)...p(X = xK)}'. (Text 192)

Used as an infinitive, as here, it has a quite different transitivity pattern from *tell*:

(40) Statistically significantly more Russian women, 51% (95% CI 36–66) said they would tell their partner as compared with the Kazakh women, 26% (95% CI 19–33), but on the other hand there was a tendency among the Kazakh women to *say* more often that they would tell their friends than among the Russian women. (Text 18)

In quotations, it can be used as a conditional reporting verb:

(41) …there always has been camaraderie with other PRHOs, but at least I get to spend time around now as well, so ahm … you know it is great, when you literally you *say*, will you do this, I will do that, and we meet up at the end [of the day] … that is always really nice. (Text 25)

The last example is transcribed from spoken discourse and clarified as such in Text 25. Unfortunately, *say* is used so infrequently as a reporting verb in MRAs that it is hard to draw reliable conclusions.

Said is used 28 times in the corpus; in almost every instance it was used as a reporting verb, and there was a corollary with spoken discourse, as indicated by the excerpt from the concordance:

```
1from diagnostic equipment, DDs    said they needed facilitation
2  elderly care physicians (62%)    said they never observed this
3 anti-emetic. Many physicians     said they waited to see wheth
4enough options for education      said that offerings for educa
5    (5%). Some physicians who     said there were not enough op
6     'somewhat helpful'; 80 (85%)  said they were 'very comforta
7   viewing the video; 89 (95%)    said they would 'definitely'
8 portunities. I'd be lying if I   said (money had no influence
9 sort of squashed with work. I    said 'I'd better get out of t
```

```
10    seeming useful then; 38 (46%) said it had not seemed useful
```

This reporting verb is primarily used to show that someone gave an answer orally; it is also worth noting that the actor is usually not the author, but instead is typically the set of participants in a study. It is used to introduce a quotation three times, but the quotes are part of the experiment and not quoting a fellow researcher. In addition, it is not found in the Introduction section of any of the texts in the MRA corpus.

Says is only used once:

> (42) This would include thinking of all the reasons why children bully other children and what this *says* about the children who did the bullying, rather than him. (Text 164)

On the whole there are so few instances of this word it is difficult to offer any descriptions in terms of stance or SP.

5.4.2.11 *Stress*

The next word, *stress*, has multiple form classes. The word appears 299 times so collecting every eighth occurrence forms the sample, of which none are reporting verbs (Appendix 4, Table 11). According to Wyse (2009, Appendix B), it is classified as a non-factive.

Every usage of the base form of *stress* is a noun in the sample. Accordingly, it seems that *stress* in the base form is not a reporting verb for this genre.

Stressed appears only 47 times and all instances are participles appearing in nominal phrases. *Stresses* is only used as a reporting verb once out of its 14 occurrences:

> (43) Whereas HBM *stresses* the importance of treating all fevers as malaria, caretakers are aware that fever is a symptom of several illnesses and often treat symptoms as separate diseases with a variety of drugs [25,26]. (Text 112)

It is connected in this example to 'importance'. However, since there is only one instance, the only conclusion I may draw is that it is used in the genre of MRAs only seldom as a reporting verb, and could suggest a stance showing high confidence.

5.4.2.12 *Suggest*

Counting all of its forms, the next verb, *suggest*, appeared the second most frequently in the corpus. According to Wyse (2009, Appendix B), it is a non-factive. Since *suggest* appears 330 times, a random selection of every tenth entry in the MRA corpus was collected, with 26 out of the 33 instances being a reporting verb. *Suggest* is complicated in terms of its factiveness and its semantic prosody, as seen below in Table 5.4; tables for other forms are found in Appendix 4.

Suggest is predominantly used as a reporting verb, though in several instances it is used to recommend instead of report, as in:

> (44) Thus, we suggest that the FTPV could provide an auditory analogue of the well known N170 [44, 49]. (Text 231)

Suggest is used with many actors, and there is no noticeable pattern that emerges; it typically is used to discuss the experiment and can be found in all sections, although only once in the Methods section. It appears more frequently in the Discussion section. As for its semantic prosody, I did find what appears to be some element of validation and certainty, and a stance showing some degree of epistemicity as in:

> (45) . . . our data tentatively suggest, that transcript degradation could at least play a possible role in silencing expression of Cre recombinase. (Text 244)

This indicates a prosody of factiveness; however, with the use of so much modality by the use of modal auxiliary verbs and adverbial mitigation, I suggest that the verb could be classified as having qualities of both factiveness and counter-factiveness and therefore any categorization would depend on usage within specific contexts.

Table 5.4: Semantic prosody for *suggest* as reporting verb

Suggest	Class	Actor	Factiveness	Semantic prosody	Level	Location
1	AV	Others	Counter-factive	attempted	Field	Introduction
2	AV	Others	Non-factive		Field	Introduction
3	RV	Data	Counter-factive	strongly	Experiment	Introduction
4	AV	Others	Counter-factive	possibly	Experiment	Introduction
5	RV	Data	Counter-factive	additional complications	Experiment	Introduction
6	AV	Designs	Non-factive		Experiment	Method
7	RV	Results	Factive	so far	Experiment	Results
8	RV	Data	Factive	May	Experiment	Results
9	RV	Data	Factive	despite, significantly	Experiment	Results
10	RV	Studies	Counter-factive	also	Field	Results
11	RV	Results	Non-factive		Field	Results
12	RV	Results	Non-factive		Experiment	Results
13	RV	Results	Factive	tentatively/ could at least play a possible role	Experiment	Discussion
14	RV	Results	Factive	support	Experiment	Discussion
15	AV	Authors	Counter-factive	Could	Experiment	Discussion
16	RV	Studies	Factive	but there is scant available data	Experiment	Discussion
17	AV	Finding	Counter-factive	may	Experiment	Discussion
18	RV	Data	Counter-factive	may	Experiment	Discussion
19	RV	Data	Non-factive		Field	Discussion
20	RV	Others	Factive	seemingly	Experiment	Discussion
21	RV	Findings	Counter-factive	may not	Experiment	Discussion
22	RV	Patient	Counter-factive	may	Experiment	Discussion
23	AV	Authors	Counter-factive	might	Experiment	Discussion
24	RV	Others	Factive	are possible	Field	Discussion
25	RV	Results	Non-factive		Experiment	Discussion
26	RV	Others	Factive	can	Field	Discussion
27	RV	Authors	Factive	for now/ may	Experiment	Discussion
28	RV	Data	Factive	may in fact	Experiment/ Field	Discussion
29	RV	Results	Factive	Correct	Experiment	Discussion
30	RV	Results	Non-factive		Experiment	Discussion
310	RV	Authors	Counter-factive	in support	Experiment	Discussion
32	RV	Results	Non-factive		Field	Discussion
33	RV	Results	Counter-factive	may	Field	Discussion

With regard to *suggested*, of which there are 214 occurrences in the MRA corpus, every sixth entry was collected for the sample. There were many similarities between *suggested* and *suggest*; both are often used as reporting verbs and both have many possible actors. A sense of disagreement or non-factiveness is more apparent with this form of the verb, particularly in the Discussion section, which is where the majority of its uses occur. The major difference between the two forms is that *suggested* was primed more frequently for disagreeing with the field, for example:

(46) There was no association between aneuploidy and HIV infection in our cohort but previous reports *suggested* higher proliferation and lower DI among HIV-associated lymphomas [44]. However, these reports are scanty, and further documentation is needed [48]. (Text 12)

Here, the authors not only say where their findings differ from what was suggested by other researchers, but also characterize reports by others as 'scanty'.

There are 284 instances of *suggests* throughout the corpus; the sample comprises 33 examples. *Suggests* is primed differently from the other forms: in the Results section it is used exclusively for discussing the experiment but in the Introduction and Discussion sections it is primed for describing the field. The following example shows how it is used for results in the experiment:

(47) Aberrant silencing of the 14-3-3 gene in many types of epithelial malignancies strongly *suggests* that it may function as a potent tumor suppressor gene. Indeed, the tumor suppression effects of… (Text 9)

The example below shows contrastiveness with the field:

(48) It was previously suggested that G1 arrest occurs when the expression is down regulated, but our data suggests that down regulation only defines the competence for this response to Dpp. (Text 86)

When describing the experiment, *suggests* is coloured as a factive, but when it pertains to the field, it is a counter-factive.

5.4.2.13 *Think*

Think appears in its different forms 128 times, but never in the third person singular form. Wyse (2009, Appendix B) classifies *think* as a non-factive. The base form and the first person plural form occur 33 times, although entries 2 through 10 are from the same article and are quotations from participants on the experiment. This aside, there still remain many instances of *think* being used as a reporting verb throughout the MRA texts, not just for spoken transcription. As a reporting verb, *think* expressed a

semblance of uncertainty as well as a feeling of contradiction. When *think* was used in the text, the actors were human, suggesting that in MRAs, inanimate objects like reports or findings are not expressed as 'thinking'. There were also a couple of instances of an implied suggestion through its use in an idiomatic expression (my italics), as in:

> (49) In our experimental set up, cell proliferation was, however, not affected by BMP4 treatment irrespective of the culture time and concentration and *one might think* of different reasons for this result. (Text 152)

Concerning LP, in most cases, *think* was related to the experiment, and even when it focused on the field, the experiment was on the periphery:

> (50) We *think* that increased density of mast cells in these cases should not be overlooked and it may contribute to clinical manifestations in some way. (Text 50)

The fact that *think* is primed for discussing the author's own work is interesting and noteworthy as it does have a quality of being counter-factive, suggesting that this could be a verb authors use when they are not fully confident about their own claims. These findings are in addition to the main finding or issue of the experiment: should authors find themselves unable to offer substantial support for their main finding, they will need to shield themselves and their argument. To do this, they can use *think* for either their primary or secondary findings. An example of this from the corpus is:

> (51) We think that this difference in the speed could be related to a general slowing down in movements that has been reported in stroke patients [50,51]. (Text 118)

There were 95 occurrences of *thought* throughout the corpus so every third entry was chosen for the sample. *Thought* seems to be far more primed to what researchers across the field think and have thought, rather than what the authors think because of the reported experiment. Once again, there was a contrastive feeling with this word, but it was mostly centred on the field, for example:

> (52) Our results show that category-specific processing in high-level category-sensitive cortical areas already occurs during the first 100-ms of visual

processing, much earlier than previously *thought*, hereby shedding a new light on the early neural mechanisms of visual object processing. (Text 245)

The actors are always humans and *thought* does appear in all sections of the texts, signalling a stance of slight uncertainty. There are no occurrences in the MRA corpus of *thinks*. I would classify this reporting verb as being counter-factive.

5.5   Discussion

This chapter had as its goal to establish whether it is possible to determine some of the reasons behind an author's selection of a particular reporting verb in medical research articles (MRAs). The question is approached through two lenses, semantic prosody and lexical priming, particularly as supported by corpus analysis. By providing some measure of evaluation, these lenses allow readers to infer the author's stance. The distinction between factive and non-factive was based on Wyse's (2009) definitions, and this distinction has been studied by numerous scholars: see Tadros (1985), Hooper (1974), and Williams (2004). I included the distinction of counter-factive, which would map onto Wyse's coding of non-factive, following Thompson and Ye (1991), while in the research non-factive is closer to neutral as in it can be used to show a stance of agreement or disagreement depending on the context.

Table 5.5 displays each of the 13 verbs in their different forms with semantic prosody and lexical priming, based on location within the IMRD structure; asterisks denote difference between my findings and those of Wyse (2009), while an X denotes either no occurrences found in my corpus as a reporting verb, or those verbs with too small a sample to draw any conclusions. Tables 5.6 through 5.9, below, group specific verb forms by factiveness and focus in specific IMRD sections.

Table 5.5: Semantic prosody and lexical priming across four sections of MRAs

Word	Introduction	Methods	Results	Discussion
Argue	Counter-factive No specific focus or actor	X	Counter-factive Focus on experiment	Counter-factive No specific focus or actor

Argued	Counter-factive No specific focus or actor	X	Counter-factive Focus on experiment	Counter-factive No specific focus or actor
Argues	X	X	X	X
Claim	Counter-factive Focus on field	X	Counter-factive Focus on experiment	Counter-factive Focus on both field and experiment
Claimed	Counter-factive Focus on field	X	Counter-factive Focus on experiment	Counter-factive Focus on both field and experiment
Claims	Counter-factive Focus on field	X	Counter-factive Focus on experiment	Counter-factive Focus on both field and experiment
Discover	Factive Focus on field	Factive Focus on experiment	Factive Focus on experiment	Factive Focus on both field and experiment
Discovered	Factive Focus on field	Factive Focus on experiment	Factive Focus on experiment	Factive Focus on both field and experiment
Discovers	X	X	X	X
Find	Non-factive Focus on field	Non-factive Focus on both field and experiment	Non-factive Focus on both field and experiment	Non-factive Focus on both field and experiment
Found	Non-factive Focus on field	Non-factive Focus on experiment	Non-factive Focus on experiment	Non-factive Focus on both field and experiment
Finds	X	X	X	X
Indicate	Non-factive Focus on field data	Non-factive Focus on experiment data	Non-factive Focus on experiment Data	Non-factive Focus on both experiment/field data
Indicated	Non-factive Focus on experiment	Non-factive Focus on experiment	Non-Factive Focus on experiment	Non-Factive Focus on experiment
Indicates	Non-factive Focus on experiment	Non-factive Focus on experiment	Non-factive Focus on experiment	Non-factive Focus on experiment
Observe	Factive Focus on field	Factive Focus on experiment	Factive Focus on experiment	Factive Focus on experiment
Observed	Factive Focus on	Factive Focus on	Factive Focus on experiment	Factive Focus on experiment

	field	experiment		
Observes	X	X	X	X
Point out	X	X	X	X
Pointed out	X	X	X	X
Points out	X	X	X	X
Proved	X	X	X	X
Proves	X	X	X	X
Report	Counter-factive No clear priming related to focus or location	Counter-factive No clear priming related to focus or location	X	Counter-factive No clear priming related to focus or location
Reported	Counter-factive Focus on field	Counter-factive No clear priming related to focus or location	Counter-factive No clear priming related to focus or location	Counter-factive Focus on field
Reports	Counter-factive Focus on field	Counter-factive No clear priming related to focus or location	Counter-factive No clear priming related to focus or location	Counter-factive Focus on field
Say	X	X	X	X
Said	X	X	X	X
Says	X	X	X	X
Stress	Factive Used as a noun	Factive Used as a noun	Factive Used as a noun	Factive Used as a noun
Stressed	Factive Participles appearing in nominal phrases	Factive Participles appearing in nominal phrases	Factive Participles appearing in nominal phrases	Factive Participles appearing in nominal phrases
Stresses	X	X	X	X
Suggest	Non-factive Focus on the experiment	Non-factive Focus on the experiment	Non-factive Focus on the experiment	Non-factive Focus on the experiment
Suggested	Non-factive Focus on field	Non-factitive Focus on experiment Only one occurrence	Non-factive Focus on experiment	Non-factive Focus on field
Suggests	Non-factive Focus on field	X	Non-factive Focus on experiment	Non-factive Focus on field
Think	Counter-factive Focus on field	Counter-factive Focus on experiment	Counter-factive Focus on a quote or experiment	Counter-factive Focus on both experiment and field

Thought	Counter-factive Focus on field	Counter-factive Focus on a quote or experiment	Counter-factive Focus on a quote or experiment and the field	Counter-factive Focus on field
Thinks	X	X	X	X

Some of the word forms were not used as reporting verbs, such as *thinks*, *stress, stressed, prove, observes,* and *discovers*. That some of them did not appear could be due to grammatical reasons; for example, *discovers* describes a present discovery being done by others, and it would be hard for an author to use it in this genre of self-reporting. Others, such as *prove*, may not be used for editorial reasons, such as being considered overly inappropriate for reporting experimental findings. A few of them, such as *stress*, are used as nouns more often in MRAs. Others may simply not occur frequently enough to appear in a corpus of this size.

In support of the idea of local boundaries, 5.2.1 offers several examples of a word's semantic prosody and lexical priming changing when the word is located in different sections. *Found* has a prosody of non-factiveness in the Introduction, Methods, and Discussion sections of MRAs in my corpus, but switches its prosody to factiveness for the Results section. The reason for this may be due to a switch in focus of what is being reported. A pattern emerged that when authors were discussing the field they would generally use a verb that is less factive, which supports the idea that writers need to separate or distance their work from others in a rhetorical manner. The following series of tables is arranged to show which verbs, separated by textual sections (IMRD), are factive and non-factive, or both, based on my research and whether they are being used to describe the field, the experiment, or if there were no clear focus as to factiveness.

Table 5.6: Verbs based on factiveness and focus found in the Introduction sections

	Field	Experiment	No clear focus
Factive	discover, discovered		observe, observed
Counter-factive	claim, claimed, claims, reported, reports, think, thought		argue, argued, argues, report, say
Non-factive	find, found, suggests, suggested	indicate, suggest	indicated, indicates

From the table, it can be seen that a majority of the words were either used for focusing on the field or lack a definite focus, which can be possibly explained by the situation: in the Introduction section, an author is generally describing what others have done, and very little of the text is about the current research. (See the description in Chapter 4 of the Moves in the Introduction section.) Following this line of thought, more verbs showed a prosody of non-factiveness, as part of the author's objective to show the difference between his/her work and that of others, in order to persuade the reader that the previous studies were in some way flawed.

Table 5.7: Verbs based on factiveness and focus found in the Methods sections

	Field	**Experiment**	**No clear focus**
Factive	discover, discovered		observe, observed
Counter-factive		think, thought, suggest, suggested	report, reported, reports, say
Non-factive		finds, found, indicate, indicated	find, indicates

In this table, we see a shift from the greater focus in the Introduction section on the field, to a greater focus in the Methods section on describing the experiment. There are also fewer verbs in total in this section, due in part to the situation. In the Methods section, writers are usually outlining what they did, as opposed to what they found. (see 4.3.1.2 for a more detailed discussion of this.) Two verbs that I do not include in this table are *said* and *says*, as their focus was neither on reporting findings nor on the field but only on what participants had spoken orally.

Table 5.8: Verbs based on factiveness and focus found in the Results sections

	Field	**Experiment**	**No clear focus**
Factive		discover, discovered, observe, observed	
Counter-factive		argue, argued, claim, claimed, claims, found, think	report, reported, reports, thought
Non-factive		indicate, indicated, suggest, suggested, suggests	find, indicates

For the Results section, the verbs are either solely about the experiment or used for describing the authors' own focus; throughout the samples by sections, no reporting verbs are used only for describing the field, which supports the classification of the Results section, in Chapter 4, as being primarily focused on the experiment. As this

section is focused mostly on the author's own work, that may explain why there were more verbs that exhibited factiveness, since authors are presumably unlikely to doubt their own work, particularly at the end of their effort to persuade the reader of the essential correctness of their argument.

Table 5.9: Verbs based on factiveness and focus found in the Discussion sections

	Field	**Experiment**	**No clear focus**
Factive		observe, observed	discover, discovered
Counter-factive	reported, reports, thought		argue, argued, claim, claimed, claims, find, found, report, think
Non-factive	suggested, suggests	Suggest	indicate, indicates

The Discussion section presents a swing back towards having more to say about the field than either the Results or Methods sections. The Discussion section has been described in Chapter 4 as being the most balanced from a focus standpoint, inasmuch as the section can describe the findings of the experiment and then compare it to others.

Some of my findings relating to the factiveness of specific verbs were different from Wyse's (2009) and that could be for two main reasons. The first was that he only described factiveness in two dimensions, factive and non-factive, while I added a third, counter-factive. The other reason is that he did not employ SP to analyse the words but rather relied on previous definitions of the words by others.

5.5.1   Conclusion

By employing corpus linguistic techniques, much can be examined with regard to the impact of semantic prosody and lexical priming. Future study could review corpus data across genres to find if there are differences in regard to semantic prosody and lexical priming for multiple word types. It would be of great benefit in helping both readers and writers understand the reasons why authors in certain genres often co-select words, not only from a pedagogical standpoint. The better we understand word selection, the better we may understand the meaning that the author is trying to convey.

This chapter set out to examine if it was possible to discover why authors of MRAs used a particular reporting verb. To do this, I utilized corpus linguistic tools and approaches, by looking at the verbs through the lenses of semantic prosody and lexical priming. I cannot present any conclusive claims regarding the verbs in terms of their semantic prosody and lexical priming as I could not run statistical tests on their varying numbers of occurrences, but from my results I can conclude that it is possible to describe words based on these techniques. I can, in addition, state that certain verbs are used more often than others: for example the instances of *find* and *suggest* far outnumber those of *prove*, *point out*, and *say*, and this is useful information for second language users who write in this genre.

Traditionally, semantic prosody is categorised in terms of positive or negative feelings of the word being examined. Accordingly, for this research, I reviewed the selected reporting verbs in terms of the authors' possible stance toward what is being reported, choosing the terms factive, counter-factive and non-factive to describe the colouring of the words, and incorporating consideration of their positive and negative polarity.

Overall, it appears that the idea of defining the prosody of words by types of factiveness may be fruitful and can be used as a way to describe verbs in the writing of MRAs. Lexical priming helps to support the idea that words are used differently and primed by more than text or grammatical constraints, so that for a word to be used, an author may need to consider a word's potential semantic prosody as well as its location within a text.

Chapter 6: Selected discourse markers in medical research articles

6.1   Introduction

Previous chapters have defined the genre of medical research articles (MRAs) (Chapter 3), mapped their Rhetorical Moves (Chapter 4), and analysed reporting verbs in terms of how and why authors may use them within MRAs (Chapter 5). This chapter has as its primary goal to examine how cohesion can be maintained within a text using discourse markers (DMs); that is, it will attempt to discover if and how DMs connect segments of text and manage information flow in this genre. It will begin with a discussion of what DMs do, followed by a description of the coherence properties exhibited by DMs, and a brief discussion separating DMs and pragmatic markers (PMs). Section 6.1.3 will present an overview of previous research on DMs in general, leading to the examination of four selected DMs and one newly discovered DM, together with their distribution and functions in the MRA corpus. These DMs are *however* (beginning in Section 6.2.2), *moreover* (beginning in Section 6.2.3), *therefore* (beginning in Section 6.2.4) and *so far* (beginning in Section 6.2.5). *As mentioned*, the newly identified DM, will begin in Section 6.2.6.

6.1.1   Examining discourse markers in medical research articles

As previously mentioned (see 1.3), it is necessary to study how certain aspects of discourse, such as coherence or stance, are achieved in a particular genre. Discourse markers are multifunctional, and have been studied by researchers in philosophy, pragmatics, semantics and sociolinguistics, with approaches from systemic functional linguistics (Halliday & Hasan 1976), discourse coherence (Schiffrin 1987; 2001), relevance theory (Blakemore 1992) and pragmatics (Fraser 1999; 2006; 2009a).

The study of DMs is important, as authors use them to connect segments of text to other segments, and it is this pragmatic linking of ideas that this chapter will examine. Authors use a DM to act like a bridge or a signpost, and are able also to draw on potential inherent evaluative, epistemic or pragmatic qualities that may not be readily

apparent in the grammatical use of the word. Beyond text-division at a meta-structural level by section titles, such as Introduction, Methods, Results and Discussion (see Chapter 4), DMs can mark information boundaries, mark information links for discourse coherence, signal stance (including inference and evaluation of prior sources or claims), and tell readers how to read the text via rhetorical functions such as signposting, evaluating and persuading (Schourup 2011). According to Fraser (1999, p.932), DMs perform a variety of duties and have thus been given many names:

> Unfortunately, the term has different meanings for different groups of researchers, and we find work on DMs done under a variety of labels including, but not limited to cue phrases (Knott and Dale 1994), discourse connectives (Blakemore 1987, 1992), discourse operators (Redeker 1990, 1991), discourse particles (Schorup [*sic*] 1985), discourse signaling devices (Polanyi and Scha 1983), phatic connectives (Bazanella [*sic*] 1990), pragmatic connectives (van Dijk 1979; Stubbs 1983), pragmatic expressions (Erman 1992), pragmatic formatives (Fraser 1987), pragmatic markers (Fraser 1988, 1990; Schiffrin 1987), pragmatic operators (Ariel 1993), pragmatic particles (Ostman 1995), semantic conjuncts (Quirk et al. 1985), sentence connectives (Halliday and Hasan 1976).

His early list of sources has greatly expanded. In 2006, Fraser lists key citations on his website (www. http://people.bu.edu/bfraser/) under various categories of DMs. Šimčikaitė (2012, p.28) updates the list with the following:

> Discourse particle (Schourup 1985; Fischer 2000), connective (Salkie 1995; Axelrod & Cooper 2001; Celle and Huart 2007), insert/ discourse marker (Biber et al. 1999), connector (Copage 1999; Stephens 1999; Frodesen & Eyring 2000), discourse marker/ utterance indicator/filler (Pridham 2001), linker (Foley & Hall 2003), pragmatic marker/discourse marker (Aijmer 2002; Aijmer, Foolen & Simon-Vandenbergen 2006; Carter & McCarthy 2006)

Following major scholars, such as Halliday and Hasan, Schiffrin, Fraser, and Blakemore, numerous researchers have set out to define and describe discourse markers (DMs) and their functions in spoken and written text. Because many DMs are multifunctional, this is still a fuzzy area, leading to disagreements and differing terminologies among those who study them from various perspectives. This chapter will employ Blakemore's early (1992) categorisation of types of DMs (see 6.1.3.4) to describe five DMs in the MRA corpus, and attempt to locate them within Fraser's two classes of DMs (see 6.1.3.3). The work will use corpus analysis and the lenses of

semantic prosody (SP) and lexical priming (LP). I will attempt to show how authors of MRAs tend to use the five DMs selected for analysis in terms of their colouring and priming, whether they are being used to agree or disagree with previous information, and whether their location within a sentence, paragraph, or larger chunk of text affects the ways in which the DMs are used.

6.1.2   Major functions of discourse markers

As part of publishing academic and scientific articles, researchers need to demonstrate that their research is both reliable and relevant. In so doing, their language use needs to be moderated so that the authors do not over-reach themselves concerning either their new claims or their disagreements with previous research. Swales (1990) notes that one part of writing a research article is its rhetorical function of convincing the reader/reviewer that the author's work is the most authoritative and correct but, to do this, authors need to show contrast or comparison between their work and that of others, which is a persuasive function. It is difficult to negotiate the path of academic hubris, as Salager-Meyer (1994) comments, since researchers must not only refer readers to their current work, but also to their own and others' previous studies. The preceding chapter focused on the ways authors report findings and moderate their claims; this one will look at how authors may use strategically-placed discourse markers in an attempt to soften their claims, or to possibly act as barriers between the author's claims and potential criticism.

This chapter will first characterise DMs, and will touch briefly on modality and subjectivity to lay the groundwork for a longer discussion using *however, moreover, therefore, so far* and *as mentioned* (see 6.2) in their roles as DMs to illustrate a sustained distinction between the broader field and the specific experiment in different sections of MRAs. Establishing a niche for current research often means identifying a gap in the existing knowledge; in MRAs, the gap can be something in the field as a whole, or it may be in the particular experiment being reported. I will first examine four lexemes traditionally classified as DMs, *however, moreover, therefore,* and *so far,* and then identify and define a fifth class of DMs that I have found, using *as mentioned* as an example of what I call a meta-discourse marker. As

Hyland and Tse (2004, p.158) argue, metadiscoursal comments can be textual, organizing the text in various ways, and interpersonal, which suggests the writer's stance to the reader; they call these endophoric markers. While the present work emphasizes the textual, its examination of writer stance in various moves supports the identification of a metadiscoursal discourse marker. I will look particularly at the shifts they signify and where they typically appear within paragraph and sentence level contexts, as well as at their location in the larger IMRD structure.

6.1.3   Working with discourse markers: research highlights

DMs have long been a popular area of linguistic research: Fraser (1999, p.301) calls the study of DMs 'a growth market in linguistics', and while most studies focus on spoken texts, DMs have started to be examined in written communication as well (Teufel 1998). What DMs are, and ways to identify them, warrant clarification.

Similar to Halliday and Hasan's (1976, p.276) ideas regarding how conjunctions and other conjunctive devices work, DMs may act like signposts, tying words together or sections of text to each other, and may mark discourse boundaries. According to most scholars (Schiffrin 1987; Fraser 1990, 1999; Blakemore 1992), DMs signal a shift in the flow of information and may also be used to indicate the extent to which the text that follows, while being connected to the previous clause or section of text, may or may not be automatically in agreement with what had previously been claimed.

6.1.3.1 Halliday and Hasan: cohesion and coherence

Over the last forty years, studies of those features impacting the coherence of a text, particularly cohesive features, have come to the forefront of functional linguistic research. Halliday and Hasan (1976), in their highly influential work on cohesion in English, describe five ways that text and discourse can have cohesion between segments, namely reference, repetition, substitution, ellipsis and conjunction. Their discussion of the semantic function of conjunctions (and the larger category of conjunctives) is close to that which this work calls discourse markers, or DMs.

Accordingly, Halliday and Hasan posit the idea of conjunctions as being a semantic class rather than a purely structural relation, as can be seen in this passage, which focuses on the process of conjunction:

> …in describing conjunction as a cohesive device, we are focusing attention not on the semantic relations as such, as realized throughout the grammar of the language, but on one particular aspect of them, namely the function they have of relating to each other linguistic elements that occur in succession, but are not related by other, structural means. (Halliday & Hasan 1976: 227)

Halliday and Hasan view conjunctions as the means by which coherence is maintained throughout different segments of text (1976, p.227). They propose that conjunctions (called DMs in this work) can have more than just a connective quality, as in this example:

> I went to the restaurant you recommended to me *and* it was awful.

In this example, *and* is connecting the two segments; however, given that the second clause expresses a negative reaction to the recommendation cited in the first clause, we may infer that the use of *and* here highlights the contrast, and is more than simply a connective device. Although Halliday and Hasan (1976) do not use the term *discourse marker*, their notion of a cohesive device plays a prominent role in subsequent accounts of discourse markers. Their list of cohesive devices includes many of the expressions that have been classified as apposition markers; for example, *that is, in other words, at least, in short* and *for instance.* Halliday and Hasan also assert that there can be adversative qualities within a conjunctive device, or conjunction (1976, p.237); for example, *but* may connect two segments (I like candy *but* it is bad for me); however, the cohesion is that of contrastiveness. *But* has always been used to show contrast, but Halliday and Hasan argue that the same contrastiveness can be found for words like *yet, so* and *then*. Their framework claims that a conjunction could belong to one of four categories: *additive, adversative, causal* and *temporal* (1976, pp.238-239). *Additives* are those words that add information, *adversatives* show contrast between segments, *causals* show that something happened because of something else, and *temporal* words show shifts in time.

The notion that the conjunctions could have additional meanings or functions beyond their role of connecting opened the door for researchers to examine the ways in which this could occur, and led to a major work on DMs, the sociolinguistic study by Schiffrin (1987). A second aspect of Halliday and Hasan's findings is that conjunctions have different roles that today we classify as pragmatic in nature. The present work will draw on and expand their findings to examine how and where authors may show stance with their choice of conjunctions and other conjunctive devices.

6.1.3.2 Schiffrin and discourse coherence

In her early work on discourse markers, Schiffrin (1987) focused on developing a discourse coherence model. According to Schiffrin (1987, pp.24-25), there are five levels of coherence within discourse, which are typified by the different functions that DMs may employ:

- exchange structure, which includes adjacency-pairs such as questions and answers,
- action structure, which is where speech acts are situated,
- ideational structure, which is viewed from a semantic point of view as an idea exchange,
- participation framework which is the interaction and relation between the speaker and listener, and
- information state, which focuses on participants' cognitive capacities.

Exchange structures lend themselves to spoken discourse more than they do to written texts. These function as DMs that manage whose turn it is to talk within a conversation (Schiffrin 2006, p.16). Action structures are replies from one speaker after hearing something from a different speaker (Schiffrin 2006, p.16). Ideational structures are used more frequently in writing and are for the relaying of information, similar to Halliday and Hasan's additive class. A participation framework explains how the relationship between the author and audience may impact how the information is presented. Schiffrin's information state refers to the fact that the speaker wants to pass information on to the audience, and signals that there is more

information to come. While crucial to the study of DMs, Schiffrin's work is more applicable to spoken discourse as it relies heavily on turn taking and acknowledgement, something that is difficult to process with a written, informative text.

However, according to Yang (2012, p.39), studies that employ Schiffrin's framework are typically focused more on coherence and less on 'local context'. Schiffrin herself notes that her earlier work was influenced by her belief at the time that discourse was not only a unit of language, but was also a process of social interaction (2001, p.56). In her subsequent work, Schiffrin makes adjustments to her earlier framework, stating that a word can sometimes be a DM and sometimes not, and that a scholar needs to examine the data, as meanings could change (2001, p.66). Her clarification is important for my study, as I will use corpus analysis to review a sample of DMs and will draw on her more recent construct to determine if a word is a DM. This needs to be looked at on a case-by-case basis, which provides some flexibility in determining whether or not a word is being used to bridge two segments of text. Schiffrin (1987, pp.267-292) considers some DMs to refer to parts of the discourse that are not contiguous.

This last is a major difference from the framework developed by Fraser, who states that DMs are used to connect two segments that are corresponding and contiguous (2006, 2009a; personal correspondence 11 October 2014).[29] The fifth DM I will examine in this study, *as mentioned*, cannot be classified as a DM using Fraser's definition, while Schiffrin's framework allows for it.

6.1.3.3 Fraser and his location of DMs within pragmatics

Fraser (1999) places DMs within pragmatics, defined by Coupland (2009, p.852) as the study of how meaning accrues in practical activities of talk in social contexts. Fraser is one of the most prominent scholars in the area of DM research. Between 1999 and 2009, Fraser's ideas of the nature of DMs have changed, along with his

---

[29] Dr. Fraser was kind enough to allow me to use our personal communication through email in this work.

definition of DMs. At first, Fraser (1990, p.394) proposed that DMs were 'not a random group of expressions, but rather that they are a type of pragmatic (as opposed to content) class, specifically a class of commentary pragmatic markers'. He later revised his classification, declaring that DMs are a class of pragmatic markers (PMs) unto themselves (Fraser 1996, p.169). The distinction between these two kinds of markers will be addressed more fully in 6.1.4. Fraser (1999, p.950) sets forth two classes of DMs, which I paraphrase as:

> A. Those that relate the explicit interpretation conveyed by Segment 2 (S2) with some aspect associated with Segment 1 (S1);
> B. Those that relate the topic of S2 to that of S1.

According to Fraser (1999, p.950) a DM is:

> A pragmatic class, lexical expressions drawn from the syntactic classes of conjunctions, adverbials, and prepositional phrases… With certain exceptions, they signal a relationship between the segment they introduce, S2, and the prior segment, S1. They have a core meaning which is procedural, not conceptual, and their more specific interpretation is 'negotiated' by the context, both linguistic and conceptual. There are two types: those that relate aspects of the explicit message conveyed by S2 with aspects of a message, direct or indirect, associated with S1; and those that relate the topic of S2 to that of S1.

Fraser goes on to review the history of DMs as they are defined and explored by scholars prior to 2000. He notes that, because DMs are seen as different constructions by different scholars, they are differently termed. He refers to Halliday and Hasan's seminal work, *Cohesion in English* (1976), citing them for the label of 'sentence connectives' for DMs (Fraser 1999, p.932). However, I would argue that the idea advanced by Halliday and Hasan about types of conjunctive relations (Halliday & Hasan 1976, pp.238-243) more closely illustrates different functions by which to identify types of DM. They assign *nevertheless* to the adversative category (see 6.1.3.1), but *so far* does not enter into their purview.

In one of his earliest works on DMs, Fraser (1999) did not include the requirement that segments bridged by DMs need to be immediately adjacent to fulfil the criteria for being considered DMs but, by 2006, the adjacency of the segments had become one of his rules (Fraser 2006, pp.8-9). In his 2006 work, Fraser offers four classes (see

below) but, by 2009, he had decided, without elaboration, that there were only three. He no longer considers a temporal marker to be a DM (2009a, p.8).

> a) **CONTRASTIVE MARKERS** (CDMs) *but, alternatively, although, contrariwise, contrary to expectations, conversely, despite (this/that), even so, however...*
> b) **ELABORATIVE MARKERS** (EDMs) *and, above all, also, alternatively, analogously, besides, by the same token, correspondingly, equally, for example, for instance, further(more), in addition, in other words, in particular, likewise, more accurately, more importantly, more precisely, more to the point, moreover...*
> c) **INFERENTIAL MARKERS** (IDMs) *so, after all, all things considered, as a conclusion, as a consequence (of this/that), as a result (of this/that), because (of this/that), consequently, for this/that reason, hence, it follows that, accordingly, in this/that/any case, on this/that condition, on these/those grounds, then, therefore...*
> d) **TEMPORAL MARKERS** (TDMs) *then, after, as soon as, before, eventually, finally, first, immediately afterwards, meantime, meanwhile, originally, second, subsequently, when*

In order to be considered a DM using Fraser's 2006 framework, two criteria have to be met. The first criterion is that the proposed DM can only work on a lexical level, thus ruling out Schiffrin's (1987) proposal to include non-verbal gestures, while the second is that the segments being linked 'are single contiguous discourse segments' (Fraser 2006, p.5). His first criterion is important to the current research as I am looking exclusively at written text; however, his second criterion proves to be somewhat problematic, as one of the DMs I will analyse, *as mentioned*, generally connects a present idea, sentence, phrase, clause or utterance with one from a much earlier part of the text (and almost certainly does not connect neighbouring sentences). In a subsequent study, Fraser reduces the number of DMs to three, deleting the temporal marker as a category. Instead, he coins the term '*discourse structure markers*' (2009b, p.893), thus defining a new, fourth pragmatic marker, or PM (see 6.1.4), that is used to 'signal a meta-comment on the discourse'. If the adjacency requirements were removed from Fraser's definition, then this fourth type of PM could be seen as a different form of DM.

Fraser's 2009 definition of DMs (Fraser 2009a, p.5) is as follows:

(1) For a sequence of discourse segments S1-S2, each of which encodes a complete message, a lexical expression LE functions as a discourse marker if, when it

occurs in S2-initial position (S1-LE + S2), LE signals that a semantic relationship holds between S2 and S1 which is one of:
a) Elaboration; *and* and *anyway*
b) Contrast; *but* and *on the contrary*
c) Inference; *so* and *as a result*

Fraser says that as long as the lexical expression, in this case a DM, functions with a corresponding segment, this signals a semantic relationship of elaboration, contrastiveness, or inference, all of which are sketched in broad strokes, while he specifies that the two segments S1 and S2 have to appear in sequence, which is much narrower in scope. Blakemore counters Fraser's definition, stating:

> Fraser's framework for the analysis of non-truth-conditional meaning rests on the unexplained distinction between content or descriptive meaning and meaning which is signalled or indicated: an expression which functions as an indicator (or marker) does so simply on the grounds that it does not contribute to 'content' (2006, p.223)

Here, she asserts that Fraser's focus is too narrow and does not take into account the basic understanding that authors and readers share concerning what, in Relevance Theory, is called *implicature*, in which the reader decides what the author is saying and judges the information based on what the reader determines it to be. Blakemore's point may be that, while Fraser claims a DM may not have a semantic property within the sentence, the reader may choose to think of it as such. I find Fraser's earlier four types of DMs to be supported by the data in my corpus; therefore, I will use these categorisations in this work when discussing my first four DMs, but my data suggests that his later work may be overly narrow regarding how strictly a DM in S2 must be considered adjacent to S1, as suggested in my discussion of *as mentioned*. This is an example of what I call a meta-marker because it is still relating a segment of text to another segment and signals a connection between the two segments of information and thus is a DM; however, it does not follow Fraser's requirement of adjacency.

6.1.3.4 Blakemore and relevance theory

For discourse to be seen as coherent, there must be an understanding between the reader and the author, in that the reader needs to infer what the author means; this is

one of the foundations of relevance theory, of which Blakemore is a founding practitioner. Relevance theory is derived from the work of Sperber and Wilson (1986; Wilson & Sperber 2002) and has its foundation in Grice's (1975) theory of conversation and implicature, wherein a reader/listener has to decode what the author has implicated, not what the author has implied. Grice believed that, for an utterance to work, there must be a tacit understanding between the speaker and hearer that what the speaker is saying is true. Bach (2005, p.3) clarifies the distinction between the two, agreeing with Grice that there is a tacit understanding regarding the veracity of the speaker's utterance, so that to *implicate* is stronger than to *imply*:

> This fundamental difference reflects the fact that what a sentence implies depends on its semantic content, while what a speaker implicates is a matter of his communicative intention in uttering the sentence. That's why implicature is pragmatic in character, hence why in different situations one can utter a given unambiguous sentence and implicate different things. (Bach 2005, p.3)

One of the original, basic tenets of relevance rheory is that lexis has only two meanings, conceptual and procedural. Rouchota (1996, p.5) distinguishes between the two, as follows:

> Words with conceptual meaning contribute to the content of assertions and are analysed as encoding elements of conceptual representations. Words with procedural meaning, on the other hand, encode information about how these representations are to be used in inference, they tell you how to 'take' these representations. So, in the case of connectives, the claim is that connectives do not contribute to the proposition expressed by an utterance or to any other conceptual representation the utterance may communicate; rather they point the hearer to the context in which he is expected to process the utterance and the conclusions he should be drawing from it.

What Rouchota calls connectives derives from Blakemore's (1987) early work, in which Blakemore uses the term *discourse connectives* to refer to DMs. Blakemore (1996, p.326) argues that the greatest difference between her framework and that of Fraser is that her framework enables the separation of the semantic meaning, the truth conditional meaning, and the procedural meaning, which is the pragmatic meaning, to define what the DM is actually doing. Unlike Fraser, who adopts three classifications of DMs, Blakemore (1992, pp.138-141) separates DMs according to linguistically-specified constraints on contexts, and categorises them in four ways which I

paraphrase as follows:

> 1. DMs may allow the derivation of a contextual implication (*so, therefore, too, also*)
> 2. DMs may strengthen an existing assumption, by providing better evidence for it (*after all, moreover, furthermore*)
> 3. DMs may contradict as existing assumption (*however, still, nevertheless, but*)
> 4. DMs may specify the role of the utterance in the discourse (*anyway, incidentally, by the way, finally*).

Blakemore's functions are robust and will be used in my subsequent definition (see 6.1.5), as they open the door to the idea that DMs can have a rhetorical usage. Reviewing the location of the DM can also contribute to our understanding its contexts and its rhetorical usage. For example, although she is working with Rhetorical Structure Theory instead of relevance theory, Taboada (2006) uses two corpora of newspaper articles and scheduling dialogues to find that DMs can occur in both the Nucleus and Satellite parts of an utterance, which supports Fraser's (2006, p.14) findings that DMs can occur in the initial, middle and final parts of a sentence.

6.1.4 Distinguishing between pragmatic markers and discourse markers

In the previous section highlighting Fraser's contribution to the study of DMs (see 6.1.3.3), the term *pragmatic marker* (PM) was introduced; some clarification is warranted as to how a PM may be distinguished from a DM before positing a working definition of a DM. Fraser (1996, pp.168-169) defines PMs as being the signals within sentences that indicate the non-propositional parts of a sentence, or a large part of the pragmatic component of an utterance. According to Fraser (1996, p.169), PMs should be:

> . . . taken to be separate and distinct from the propositional content of the sentence, [and] are the linguistically encoded clues which signal the speaker's potential communicative intentions.

According to Aijmer and Simon-Vandenbergen (2011, p.223) studies of PMs indicate, 'that there were grammatical phenomena which seemed to be dependent on context

rather than on rules formulated in syntax'. The phrase 'potential communicative intentions' suggests that PMs, and hence DMs, which, according to Fraser (1990) are one type of PM, can have a pragmatic quality. That this quality may be missed by some readers/speakers, particularly by non-native English speakers, is part of the reason for the current research.

Aijmer and Simon-Vandenbergen (2011, p.227) further add that there is much confusion regarding what a PM really is; however, they denote the difference between PMs and DMs as follows: 'Pragmatic marker is preferred to discourse marker when the markers have a pragmatic rather than a discourse-marking function'. The problem with this claim is that any signal, be it a PM or a DM, can have pragmatic meaning; to say that PM is the more appropriate term does not allow for the fact that DMs can have pragmatic qualities (as shown by Fraser 1990, 1999; Blakemore 1987, 1996; Schiffrin 1987).

In the present study, I will follow Fraser's 2009a framework for PMs, as it includes DMs as a kind of PM. Fraser initially separates PMs into four distinct classes; however, those classes change over time, as can be seen in the following table:

Table 6.1: Fraser's classes of pragmatic markers taken from his two works, 1996, and 2009a

Class of PM (1996)	Description	Class of PM (2009a)	Description
1. Basic Markers	'signal more or less specifically the force (the kind of message in contrast to its content) of the basic message, include sentence mood and lexical expression' (1996, p.168)	1. Basic Markers	'signal the type of message (the Illocutionary Force – cf. Bach and Harnish 1979) the speaker intends to convey in the utterance of the segment' (2009a, p.3
2. Commentary Messages	'provide a comment on the basic message' (1996, p.168)	2. Commentary Markers	'signal a comment on the basic message' (2009a, p.4)
3. Parallel Markers	'signal an entire message separate from the basic and any commentary messages' (1996, p.168)	3. Discourse Structure Markers	'signal an aspect of the organization of the on going discourse' (2009a, p.5)
4. Discourse Markers	'signal a message specifying how the basic message is related to the foregoing discourse' (1996, p.169)	4. Discourse Markers	'typically signal a relation between the discourse segment which hosts them and the prior discourse segment, perhaps

			produced by another speaker' (2009a, p.4)

In the fifteen years between the two studies tabulated above, Fraser removes the category of *parallel markers* and replaces them with *discourse structure markers.* Were he to relax his requirement that segments connected by markers be immediately adjacent, his *discourse structure markers* could easily be considered to be a type of DM. He offers no explanation of why *parallel markers* are replaced by *discourse structure markers*, but his *discourse structure markers* may be more of an umbrella term, inside which *parallel markers* may reside. In his breakdown of the subclasses of discourse structure markers, Fraser lists three, the first being discourse management markers, or those that are used to signal information that had been covered earlier in discourse, such as *In summary* and *I add* (Fraser 2009a, p.5). I would place *as mentioned* in this subclass but, based on my corpus-based definition below (see 6.1.5), the discourse structure marker in these examples is a DM, rather than a PM. I agree that PMs and DMs have a difference of focus, although DMs are PMs as they signal something the writer wants to express to the reader. While all PMs act as signals, DMs are more specifically related to the connection of information segments, as other PMs signal other things, such as force or an emphasis the author may wish to identify. This leads to the necessity for an expanded definition of DMs.

6.1.5   Working definition of a discourse marker

Since the study of DMs is varied and expansive, as noted in the introductory remarks (see 6.1), it is difficult to find one unifying definition. My attempt here is to create an umbrella definition that can incorporate the many facets brought to the forefront by the works of previous scholars, including both semantic and pragmatic qualities. According to Biber et al. (1999, p.1086), DMs are:

> inserts which tend to occur at the beginning of a turn or utterance and to combine two roles: (a) to signal a transition in the evolving progress of the conversation, and (b) to signal an interactive relationship between speaker, hearer, and message.

There is an issue with this definition, namely the claim that DMs tend to occur in the

initial position of an utterance. Both Fraser (1990; 1999) and Rouchota (1996) have shown DMs to have the ability to occur in several positions throughout a segment of text. However, even without a more inclusive component of location (similar to the problem with Fraser's definition of constricting DMs to occur only between adjoining segments), this is a fairly comprehensive definition. Throughout this chapter, the definition of a DM will be as follows:

> Discourse markers are words or phrases that signal two features in a stretch of text. The first is that the author is signalling a connection between two segments of text, and the second is that the author is sending a message to the audience regarding the importance of the relationship between the two segments.

While the above definition is broad, it is purposefully so as, prior to obtaining empirical evidence, I am hesitant to put constraints on the location of DMs. I will also attempt to show that DMs are used to indicate some feature of positive or negative stance, or of epistemic stance by the authors, which are two of the many possible pragmatic judgments that an author may want to display. The next issue is identifying the classes into which DMs can be categorised. This study will use Fraser's four classes: contrastive markers (CDMs), elaborative markers (EDMs), inferential markers (IDMs), temporal markers (TDMs) (as originally established by Fraser 2006, p.16) and a final set of markers which I will call meta-markers (MDMs), such as, *as mentioned, previously stated* and *as earlier discussed*.

6.1.5.1 Contrastive discourse markers

Fraser (2009a, p.87) updates his definition of CDMs as 'signals that the speaker of S2 considers S1 to be an incorrect representation of some action, state, or property attributed to an aspect of that segment, and offers S2 as the correct representation'. He uses the example of *on the contrary* to exemplify how CDMs work.

6.1.5.2 Elaborative discourse markers

Fraser (2009b, p.9) amended his 2006 definition of EDMs to state that authors may use EDMs to 'signal[s] an elaboration in S2 to the information contained in S1'. He uses the primary example of *and* to illustrate how authors may use S2 as an explanation of why something disagreeable in S1 is nonetheless acceptable (2009b, p.19).

6.1.5.3 Inferential discourse markers

Fraser's revised definition of IDMs (2009b, p.9) indicates that authors may use them as 'signals that S1 provides a basis for inferring S2'. His primary example is *so*, and it is used to exemplify why S1 may have occurred.

6.1.5.4 Temporal discourse markers

According to Grote (1998, p.22) TDMs are used to signal a time reference between the segments and are ways for authors to show that something might have changed over time. Fraser eliminates this category in 2009 (Fraser 2009a) without discussion.

6.1.5.5 Meta-discourse markers

Following Blakemore's (1987, 1999) and Rouchota's (1996) formulations that some DMs can be used to link discourse beyond the contiguous segments, I will attempt to show evidence for another type of DM, which I call Meta-discourse Markers. Teufel (1998, p.45) first coined the term, although she does not consider such items to be discourse markers, but signals of rhetorical intent on the part of the author. Following Teufel (1998), we can see that MDMs are used by authors to act as a tool of persuasion. The present work will treat MDMs such as *as mentioned* (see 6.2.6) as a type of DM from the corpus-based findings that MDMs are used to signal a connection of information between two segments of text; however, instead of being confined to the adjacent position, there is a more 'global' relationship between the two segments (Schiffrin 1987 p.131).

6.2   Examination of specific discourse markers

The following discussion will examine *moreover*, *therefore, however, so far,* and *as mentioned*, as they are examples of each class of DMs. The following table shows how the present work will code the DMs, as compared to Fraser's coding.

Similar to the methodology used in the chapter on reporting verbs (Chapter 5), I will employ corpus analysis to search for the different words in context. Using SP, I will attempt to determine what, if any, stance can be identified by looking at collocates and words that commonly occur around the nodes. By applying the theory of LP, I will attempt to see if the DMs are primed in terms of use when describing information regarding the field as a whole, or how authors tend to use a DM when connecting ideas within their own research.

As mentioned throughout the previous chapters, authors distinguish their work from the work of others in a number of ways, including word choice (i.e. reporting verbs). DMs, which are used to show changes in solutions they have found to problems they have established as important (Fraser's first class of DMs) and repetition. The rhetorical device of repetition in speeches as well as academic writing is well established, and can even be used for propaganda (Vlăduţescu 2014, p.75), in that one of the ways to convince people is to repeat information. It is with this in mind that I believe authors will refer to earlier or later parts of a text, and it is here that my fifth DM will be utilised. Although Johnstone (1987, pp.206-207) notes that repetition is frequently disapproved of as a stylistic tool, she further adds that there are many different forms of repetition and that some authors may use repetition to make 'an idea persuasive even without logical support'. It is this aspect of repetition that prompted the examination of *as mentioned* since, as a DM, it refers to something stated previously in the text.

The five DMs in the MRA corpus; *therefore*, *so far*, *however, moreover* and *as mentioned*, were selected because they were previously defined by Blakemore (1992) as belonging to different categories of DMs (see 6.1.3). Then they will be separated

into Fraser's two pragmatic classes of DMs (see 6.1.2), although Blakemore (2006, p.223) later cautions that Fraser's work is very different from hers since his work is not based on relevance theory:

> However, Fraser's distinction between representational and procedural meaning is not equivalent to the cognitive distinction that has been developed in Relevance Theory […], since it appeals to the role that DMs play in the coherence of discourse. Not surprisingly, expressions that Fraser classifies as procedural (e.g., as a result) are not regarded as encoding procedural meaning in RT (Relevance Theory).

Since the present work derives from an attempt to study language in use, or pragmatics, and will employ SP and LP (to help determine the usage of the DM within the context of the text), I will also follow Fraser's work as his is a closer approximation to the findings from the analysis of the corpus.

6.2.1   Rationale for selection of specific discourse markers

As previously discussed in Chapters 4 and 5, MRAs are a fairly formulaic genre, which employs various rhetorical structures and Moves to persuade readers. As they are one of the means of maintaining cohesion in a text, authors utilise DMs for a variety of reasons. In Chapter 5, SP and LP were used in conjunction with corpus analysis to see if there were any prosodies or priming due to the actors being described. This chapter will conduct a similar analysis, but the priming will be more focused on location on the following levels: sentence, paragraph and IMRD section. With regards to prosody, Fraser (2006, p.194) has already noted that the DMs have their own core meanings and part of their definition includes either negative or positive stance. I will look at collocates and the DMs in context to see if there are any feelings that may be different from Fraser's coding. I will also see if the DMs are marked for showing differences from others' work or from the authors' own work, which is Fraser's first class of DM; or if the DMs are being used to refer to other parts of the same text, in line with Fraser's second class. I will first highlight previous research concerning each specific DM, and will then summarise its occurrences in the MRA corpus, followed by a description of how it may be primed. In addition, the five chosen DMs show the following traits: contradiction, elaboration, inference, temporal

shifts and meta-commentary. For the quality of contradiction this work will analyse *however*; for elaboration, *moreover*; for the trait of inference, *therefore*; for temporal shifts, *so far* will be used and, for meta-commentary, *as mentioned* was selected. Examples four and five are newly analysed: *so far* is seldom discussed, and the notion of a meta-marker is apparently novel.

The following table shows the total number of occurrences for each of the DMs in the MRA corpus.

Table 6.2: The occurrences of the words in the MRA corpus

DMs	*However*	*Moreover*	*Therefore*	*So far*	*As mentioned*
Totals	1263	145	543	39	18

As the table shows, there are far more uses of *however* than there are of any other DM, with *therefore* and *moreover* having the second and third largest tallies, respectively. *So far* and *as mentioned* are seldom used in MRAs.

6.2.2   *However*

However is the 69th most frequent word in the entire corpus, and appears in 548 texts of the 1000-text corpus (i.e. 250 full articles divided into the four IMRD sections. As seen in the table below, *however* seems to be used far more often in the Discussion section, then in the Introduction, followed by occurrences in Results and, finally, in the Methods section. Every use of *however* I examined was a DM as it was linking and commenting on two segments of text; even when it was used as an adverb, it was still used to signal a relationship between two segments of text (see Example (2)).

I looked manually at every instance and used collocation clues looking at how the node may be coloured or primed.

Table 6.3: The occurrences of *however* across the IMRD sections in the MRA corpus

Section	Introduction	Methods	Results	Discussion
Total size of sub-corpus	153257	286251	311394	300466
Number of occurrences	236	65	354	608

Frequency per 100,000 words*	154	23	114	203

*Note: Frequencies have been rounded up.

As previously discussed in 5.1.3, the possible discourse roles of each section probably influence the usage, as there may be some influences from local prosody. For example, in the Introduction section, it is possible that *however* is being used to identify gaps in research in the field. In Results, *however* could signal the presence of unexpected results, and in Discussion, its use could be a combination of the two previous uses.

Fraser (1990; 1999; 2009) consistently maintains that *however* is an IDM and, as such, signals that S2 is in contradiction to S1. He further adds, 'By contradicting an existing assumption (e.g. however, still, nevertheless, but) … have rough parallels to the analysis in Halliday and Hasan (1976) and to Quirk et al. (1985)' (2009, p.10). His S1 and S2 refer to segments; I will employ Fraser's term in the present study. A segment can be a clause or an entire sentence; it may be thought of as a bundle of information. According to Taboada and de los Ángeles Gómez-González (2012), *however* is a sentence adverbial that normally would be found after the Satellite (see 6.1.3.4 for a discussion of Rhetorical Structure Theory). Thus, in their corpus of English texts, *however* usually appears in the final parts of a sentence, which is not supported by the present work. Blakemore (2002, p.122) argued that while *however, but* and *nevertheless* are similar, *however* can only occur under certain circumstances: '*However* signals that the context includes assumptions which come with a guarantee of relevance accepted by the speaker'. This addresses the rhetorical impact that *however* may have, which will be examined later in this study.

6.2.2.1 The uses of *however* in the Introduction section

With regard to the Introduction section, the position of the discourse marker (beginning, middle or end of a section) could be an indicator of how *however* is being primed as tied to the Rhetorical Moves (see Chapter 4).

(1) Respiratory infectious diseases (RIDs) such as influenza are a major public health issue best dealt with by prevention, ideally vaccination. *However*, in the

first six-months or so of a newly emergent RID epidemic/pandemic vaccines are generally unavailable and non-pharmacological interventions can play a major role in minimizing RID spread [2–4] (Text 249).

Here, *however* is being used at the start of a sentence, but the information indicates that a gap exists in the field. In the first sentence, vaccinations are put forth as the best prevention of RID spread, while the *However* used to begin the second sentence signals that vaccines may not be available if an epidemic or pandemic is a new one, a situation that could indicate that the author has a possible, non-pharmacological solution to this gap.

(2) Much has been written in the popular press about plastic ingestion by Laysan albatrosses (Figure 2), *however*, relatively few empirical studies have examined this phenomenon and whether it is species wide, or if this is confined to certain populations (Text 247).

However here appears in the middle of the sentence and introduces a new clause; once again, it is being used to note a gap in the field. In the excerpt, the first clause of the sentence states that a great deal has been written about the subject (birds eating plastic), then *however* is followed by the second segment, which is a criticism of the fact that there have been few previous empirical studies. Therefore, the authors may be claiming a research gap, as they continue by saying that their study is more detailed than previous ones, and list what makes it so.

(3) The role of the RASSF/MST pathway in gastric cancer is not known, *however*. In the present study, we examined the epigenetic alteration of RASSF family genes together with MST1/2 in a panel of gastric cancer cell lines (Text 238).

The DM *however* occurs at the end of the first full sentence, a claim that there is a missing piece in the research on gastric cancer. The second full sentence presents the author's solution to the problem.

The previous examples (1, 2 and 3) show *however* linking two informational segments, occurring in the initial, the middle and the final position in a sentence, which supports Fraser's description of DMs. In each instance, the author is signalling that there is something wrong in the first informational segment, which gives the word

a feeling of contrastiveness. It is also worth noting that the problem in Example (1) is a problem throughout the field and not within the current research, which suggests that, in terms of the Introduction section, the DM is primed to express disagreement with the field. This is reasonable, as one of the major Rhetorical Moves of this section is to *Describe the problem,* which includes the Step of *Identifying the gap* (see 4.3.1.1.2).

6.2.2.2 The position of *however* within sentences and text in the Introduction section

Of the total 236 occurrences of *however* in the Introduction section, 208 (88%) occur at the beginning of a sentence. Table 6.4 displays the numeric breakdown of the occurrences, with the beginning of a sentence defined as being the initial word. For *however* to be treated as being in the middle and not at the end, there must be a full informational component, usually a clause, occurring before and after it. To be considered as being at the end of a sentence, it has to be in the final informational component position, such as 'Here is the crux of the problem, *however*'. In this sentence, although *here* is not a clause, I would classify it as an informational component referring to the preceding situation. Through punctuation, *however* is separated from the information being relayed by the word *here*.

These data are related to the inter-sentence position, but not necessarily within the structure of the text itself. Because the majority of the usage appears in the first part of a sentence, the word may be primed to start a sentence.

Let us assume that, like a sentence, a text may be divided into thirds, with one-third being the beginning, one-third the middle, and one-third the ending. Using these definitions, the following table will display the data in terms of positional occurrences and percentages. WordSmith Tools (Scott 2008) allows the user to see the percentage at which the node appears within the text. It is worth examining whether the DM is being primed differently in the various thirds, and whether there may be elements that help to indicate its semantic prosody.

Table 6.4 shows that, almost half of the time, *however* is used in the middle third of a text, with an almost even distribution between the first and final thirds.

Table 6.4: Distribution of *however* in sentences and texts of the Introduction section

Location	Sentence	%	Text	%
Beginning	208	88	70	30
Middle	24	10	105	45
End	4	2	61	25

This arrangement may mean that the word is primed to express disagreement with something that appears only after some sort of introduction to the gap or problem in segment one, and before a summary of the solution presented in segment two.

The following table displays the occurrences of *however* across the Introduction section, where it is being used to indicate that the gap/problem that has been split across two segments is related to the field of study, or to the authors' own experiment.

Table 6.5: Percentage of use of *however* with the authors' own work or work in the field in the Introduction section

Sections	Gap/Problem in field	Percentage	Gap/Problem in current experiment	Percentage
Beginning	70	100	0	0
Middle	104	99	1	1
End	61	100	0	0

Of the 236 instances, all but one in the Introduction section are used to indicate a gap in the field. The only one to indicate a gap or problem in the author's own study is as follows:

> (4) *However*, this study was inevitably biased by the long (12-year) inclusion period, during which technological and therapeutic changes have occurred (Text 81).

That *however* was used to create a research space was expected, and the fact that it only occurred once in the context of finding a gap in the author's own research is significant. It can therefore be claimed that *however*, when used in the Introduction section, is almost invariably primed to identify a gap in the field, which lays a foundation for the authors' perspicacity in identifying the gap.

6.2.2.3 The position of *however* within sentences and text in the Methods section

However appears in the Methods section 65 times; thus, a reasonable expectation is that the authors may be using it to denote a difference between their procedures and the procedures used by others. In the Methods section, *however* is overwhelmingly used in the first third of a sentence, but is more evenly distributed on the text level, as seen in Table 6.6.

Table 6.6: Distribution of *however* within sentences and texts in the Methods section

Location	Sentence	Percentage	Text	Percentage
Beginning	56	86	20	30
Middle	8	12	19	29
End	1	2	25	40

Based on the functions of the Methods section, or what it must explain to the reader, I had expected to see *however* being used to indicate a difference between the authors' experiment or other research and other research in the field, specifically the gap in the field identified by the authors. The following table shows the difference in the use of *however* for describing problems in the field as opposed to gaps in the current experiment, and the corresponding percentages.

Table 6.7: Percentage of use of *however* with the authors' own work or work in the field in the Methods section

Sections	Gap in field	Percentage	Gap in current experiment	Percentage
Beginning	0	0	20	100
Middle	2	10	17	90
End	1	4	24	96

Clearly then, the bulk of the usage for *however* seems to be highlighting gaps in the experiment, so the next step is to log the different types of gaps. They seem to fall into two categories, namely issues with materials, or issues regarding calculations.

An example of the first, where *however* introduces an issue with materials, is

(5) 400 g of mitochondria extract protein was originally applied to 11 cm immobilized pH gradient strip with linear pH gradient from 3 to 10 (IPG, Biorad) and equilibrated for 16-18 hr at 20°C in rehydration buffer, *however*

we subsequently utilized a pH gradient of 5 to 8 to increase the resolution and separation of protein spots (Text 37).

Here, *however* introduces a justification for the materials that were used in the process.

The second type, in which *however* introduces issues surrounding calculations or statistics, is exemplified here:

> (6) The previous rank reduction based on the genetic correlation matrix gave the same relative importance to all random regression coefficients. *However*, the German combined lactation EBV depends mainly on the first coefficient of each lactation (Text 84).

In this second instance, the author is explaining why his/her data might behave differently from expected.

From the data, I would conclude that, for the Methods section, *however* is usually primed to appear at the beginning of a sentence, but may appear at any point in a text. The DM seems to be used to explain differences within the authors' experiment.

6.2.2.4 The position of *however* within sentences and text in the Results section

In my corpus of medical texts, the DM *however* appears 354 times in the Results section, a substantial increase compared with the number of times it appears in the Methods section. Once again, the expectations are that the DM is used to identify issues either in the research or in the field, although there should be a much higher focus on the experiment, given that the Results section is intended to discuss the results of the experiment.

On the sentence level, *however* usually occurs in the initial position of a sentence as the first word, indicating it is used for contradictory or contrastive segments contained in two separate sentences. On the text level it is often found in the middle of the text, as its average position (as found by WordSmith Tools (Scott 2008)) is roughly a little

over halfway through the text (average 54% but with a wide distribution on either side: anywhere from 19% to 67% of the text) as shown above in Table 6.5.

Table 6.8: Distribution of *however* within sentences and texts in the Results section

Location	Sentence	Percentage	Text	Percentage
Beginning	311	88	67	19
Middle	33	10	233	64
End	10	2	54	17

From the data, I would argue that the word is primed to begin sentences that appear in the middle of a text. This seems to indicate that there needs to be a build-up of an argument before the contradiction occurs, as well as some explanations or a discussion that follows.

The Results section presents two gaps (Field and Experiment) for discussion, and three categories of issues, or types of problems: data-related, outcome-related and procedure-related (see Table 6.8). Data-related problems refer to the fact that *however* introduces problems with findings in regard to numerical, statistical data, as in:

> (7) Furthermore, evaluation of ML anatomical distribution showed that majority (57.7% (153/265) had supra-diaphragmatic, 27.5% (73/265) sub-diaphragmatic and 26.8% (71/265) disseminated lymphoma. *However*, no primary effusion lymphoma (PEL) or primary central nervous system lymphoma (PCNSL), were found in this cohort [Table 3] (Text 12).

This example discusses the results of the experiments although, in this case, it seems to focus on the raw numbers and results of the statistical analysis. An example of outcome-related issues is:

> (8) For both mice strains the lung histopathology score was higher in HVT mice as compared with controls. *However*, no differences were noted between mice strains, MV strategies and fluid strategies (Text 80).

Example (8) explicitly states that there is a gap in the results or the outcome of the experiment. The difference here is that the author of Example (8) does not use *however* to express the results; instead he/she notes a result as if to frame it as an unexpected outcome. This claim of meaning is based on context of the previous section, and while just on its own it could have a variety of meanings, it is due to its

surroundings that I am making the claim, this further illustrates how DMs are affected by more than what is immediately next to them.

For the third type, a problem or gap in procedure, the process seems to offer an explanation of why the authors have chosen to make certain procedural Moves and how that might affect the reading of the data:

> (9) Alternatively, protein-induced deformability may in some regions be diminished upon I and D substitutions. *However*, the substitutions still allow strong CRP binding due to maintenance of direct amino acid–base pair contacts in the major groove of the I and D substituted DNA fragment (Text 149).

From Example (9), we can see that the authors are stating that, regardless of certain effects, they have decided to continue because the effect did not interfere with their procedures. They are informing the reader that they have chosen to interpret the experiment in this manner.

The following table correlates the gap with its level (i.e. whether the gap is within the field as a whole or within the current experiment) and then with the percentages of the occurrences, depending on the type of issue in which the gap appears, across the Results sections of the MRA corpus.

Table 6.9: Types and occurrences of gaps/problems in the Results section

Gap/Problem	Data	Percentage	Outcome	Percentage	Procedure	Percentage
Field	0	0	4	1	56	16
Experiment	57	16	127	36	110	31

As expected, there are some occurrences of *however* being used to identify gaps in the field, but there are far more gaps in the experiment. The data would seem to indicate that, at least in the first third of the results, data or statistics are not as prevalent as is discussing the justifications of either the materials or the procedures employed. Using *however* to preface a justification is the most common usage, and close behind it is the usage of the DM to explain an unexpected outcome.

There are not as many data issues as I would have expected, leading to the suspicion that there may be a field-wide or genre-wide convention to not describe issues with

data or statistics until the Discussion section. A pattern that did emerge is that there seems to be more empirical or data-driven discourse towards the end of the Results section, similar to the Methods section. From the data, I would say that while the SP is contrastive, *however* sets up a contrast with the authors' own work as opposed to other work in the field, possibly acting as a means of proving to the audience that the authors are trustworthy because they disclose unexpected or 'unhelpful' results.

6.2.2.5 The position of *however* in the Discussion section

This last section will look at the Discussion section of the corpus. Here, *however* normally appears in the first part of the sentence throughout this section, with an average placement in the initial position on the sentence level; conversely, it tends to appear in the middle of the Discussion section, with an average of 49% on the textual level.

Table 6.10: Distribution of *however* within sentences and texts in the Discussion section

Location	Sentence	Percentage	Text	Percentage
Beginning	316	52	140	23
Middle	200	33	297	49
End	92	15	171	28

In the MRA corpus, *however* is used 608 times in the Discussion sections. Since the Discussion section repeats much of the information covered in the previous three sections (see 4.3.1.4), similar results should appear here as well.

6.2.2.5.1 *However*: Introducing gaps in the first third of the Discussion section

Due to the large number of occurrences of *however* in the Discussion section, 90 instances were randomly selected to establish a basis on which to form a representative cross selection, with 30 selections for each successive third of the section. The table below correlates the gap or problem with whether it describes a gap with the outcome or the procedure/materials, for the first third of the Discussion section and the 30 randomly selected occurrences. This section is divided into thirds

in order to make the data more manageable, and this should not be seen as indication that the patterns in this section are always the same. Unlike previous sections, at least in the beginning of the Discussion section, there seems to be no discussion of statistical findings; instead, issues are related either to procedure, which includes materials, or to outcomes. In addition, significantly more gaps in the field are identified in the Discussion section than in the previous sections.

Table 6.11: Types and occurrences of gaps/problems in the beginning of the Discussion section

Gaps	Procedure/materials	Percentage	Outcome	Percentage
Field	11	37	9	30
Experiment	1	3	9	30

Materials are included with procedures, since materials typically refer to something that is necessary in order to carry out the experiment, as shown in Example (10).

> (10) IGRA do indeed carry several advantages over the century old TST [26]: testing requires only 1 patient visit and as these are ex-vivo tests, the risk for adverse effects is reduced and potential boosting is eliminated when testing is repeated. *However*, IGRAs are costly, impose blood drawing, and necessitate appropriately equipped laboratory (Text 189).

This example illustrates a gap in the field; the issue is to identify which material might be more appropriate and what its disadvantages are. The beginning of the Discussion section seems to indicate that it is necessary to distinguish the current research from the field, more so than any previous section aside from the beginning of the Introduction section. In the first third of the Discussion section, *however* seems to be primed more for contrasting the field with the current research. Typically, the function of this section is to report the findings (see 4.3.1.4); thus, authors may feel the need to distinguish their results from those of previous works.

6.2.2.5.2 *However*: Introducing gaps in the middle third of the Discussion section

Table 6.12 displays the uses of *however* in the middle portion of the Discussion section, categorising the types of gaps that are identified, and identifying the issues that correspond to that gap.

Table 6.12: Types and occurrences of gaps/problems in the middle of the Discussion section

Gap	Data	%	Procedure	%	Outcome	%	Suggestion	%	Limitation	%
Field	0	0	6	20	2	6	3	10	0	0
Experiment	1	3	7	24	7	24	0	0	4	13

From the analysis of the sample, two new issues that had not shown up previously surface in the middle of the Discussion section; however, the Moves mirror two of the Moves identified in Chapter 4 (sees 4.3.1.4.5 and 4.3.1.4.6). The first can be applied to the field and described as *Suggestions*, which are directed to the field and are therefore defined as belonging to the gap in the field. The following is an example of a suggestion:

> (11) These findings suggest that HCLB-containing channels are candidate components of the Drosophila response to volatile anaesthetics. *However* it needs further investigations to determinate whether they are direct targets for anaesthetics, like some other members of the Cys-loop superfamily, or whether they modulate anaesthesia by other, indirect, mechanisms (Text 168).

The outcome of the experiment was not expected, but the author recommends future research on this matter, thus suggesting to the field an aspect that requires further investigation. The other new issue is *Limitations* with the experiment, which can be coded as belonging to a concern with the experiment, as exemplified below:

> (12) A potential criticism of our ... Catheter venography and magnetic resonance venography are alternative imaging modalities capable of providing greater anatomical detail than ECD. *However*, these techniques are difficult to apply for the large sample sizes required for genetic analyses, e.g., the CV is an invasive exam and value of MRV for diagnosis of CCSVI is limited [21, 22, 23] (Text 220).

At the start of (12), the authors first establish their focus as dealing with possible questions raised, and then offer a rationale as to why they still followed their chosen procedure regardless of potential criticisms. The authors then use *however* to further explain why some may question the chosen procedure.

As seen in Table 6.12, the issue of data (typically, statistics) is found least often, with only one occurrence. The procedural issue is the most common, with 13 examples in

total, followed by the outcome issue. The two new issues are very similar in frequency. In the middle third of the Discussion section, *however* seems to be primed more for use with gaps in the experiment. This seems to follow the Moves typically found in this part of the Discussion section (4.3.1.4).

6.2.2.5.3 *However*: Introducing gaps in the final third of the Discussion section

Thirty texts were randomly selected to generate representativeness of the instances of *however* in the final third of the Discussion section. The following table shows the different gaps and the issues that are connected to them.

Table 6.13: Types and occurrences of gaps/problems at the end of the Discussion section

Gap	Data	%	Procedure	%	Outcome	%	Suggestion	%	Limitations	%
Field	1	3	0	0	1	3	6	20	8	27
Experiment	0	0	0	0	6	20	1	3	7	24

Data issues were linked to a gap in the field. None of the instances of *however* dealt with issues in procedures on either a field or an experiment level. The gap for the issue of outcome also shrank by 6% compared to the middle third of the section, while suggestions increased by 12%. An interesting example occurred when researchers offered a suggestion to themselves, saying:

> (13) *However*, as many of the women in our cohort have not yet reached the usual peak age for ovarian cancer, we will continue to monitor the risk to try to establish a more definite link between use of fertility drugs and risk of ovarian cancer (Text 75).

The issue of limitations also grew by over 30% in the final third of the Discussion section. In this third, *however* could present limitations not only with the current study, but also with the field as a whole. For example, the following excerpt states that there is no current model (for the problem being examined), which is a limitation in the field, but there is almost an assumed suggestion included in the discourse.

(14) In reality, *however*, because of the difficulty of such a task, we are unaware of any model development and test set selection in the literature that incorporates a systematic selection of a representative test set (Text 135).

As the DM is evenly distributed for gaps in the field and in the experiment, I would hesitate to declare it primed for one or for the other.

6.2.2.6 Priming of the uses of *however* across the IMRD sections

In the Introduction section, almost all of the discourse surrounding *however* is primed for the field while, in the Methods and Results sections, it is primed for gaps in the authors' experiment. Finally, in the Discussion section, there is a balance in the priming between the two gaps. The following table presents the frequency of the use of *however* across the four sections of IMRD.

Table 6.14: *However* across the IMRD sections of texts in the MRA corpus

Sections	Field	Experiment
Introduction	99	1
Methods	7	93
Results	10	90
Discussion	50	50

The present study supports Fraser's claim (2009a) that *however* is contrastive in nature, as well as that a DM can appear at a variety of locations within a sentence or a text. With regard to priming, each section has a different priming effect; this could be due to the local prosody of each section (see 5.3.1 for a more detailed explanation). By this I mean that each section of the IMRD framework has different requirements that are understood by the discourse community. In the Introduction section, *however* is primed for contrasting with the field, while in the Methods section, it is used primarily for contrasting and then choosing different materials or ways of interpreting data. In the Results section, *however* is used when discussing the contrast between data, outcomes, and procedures and, in the Discussion section, an additional divergence is also used for discussing limitations within the study and for making suggestions for future research.

6.2.3 *Moreover*

Moreover is the 904th most common word in the MRA corpus, and occurs 145 times. It is disproportionately found in the Discussion section (83), followed by the Introduction section (34) and the Results section (25), while it only appears 3 times in the Methods section. This section will mirror that for *however* by first describing ways in which others have treated this DM, followed by a close examination of where and how the word appears and is used.

Fraser (2009a, p.9) defines *moreover* as an EDM, as he felt it is used to signal that authors want to elaborate on S1 in S2. He further adds that *moreover* does not have a polysemous form; that is to say, there is no instance where *moreover* can be anything but a DM, unlike other DMs that can have an adverbial form. The example given is *on the other hand*, where the DM form and the adverbial form have vastly different meanings (Fraser 2009a, pp.10-11).

6.2.3.1 The uses of *moreover* in the Introduction section

Unlike *however* in the Introduction section, *moreover* only appears in the initial place in a sentence. There were only 34 instances of *moreover* found in the Introduction section, and the word can be found in any of its three parts, namely the first third (13), the middle third (9) and the final third (12). In the first third, it could be part of the Move *Present study situation* (see 4.3.1.1.1), as parts of the Moves *Framing the paper* and *Discussing the background*, and *moreover* could be used to supply additional information, for example:

> (15) L-tryptophan is an essential amino acid necessary for protein synthesis in mammalian cells. *Moreover*, tryptophan is the precursor for the neurotransmitter serotonin, for the hormone melatonin, and contributes to the synthesis of the coenzymes NADH and NADPH (Text 173).

In this example, *moreover* is signalling that the author has some additional information when defining the purpose of tryptophan. When it occurs in the middle third of the Introduction section, it may appear as part of Move Two, *Describe the*

problem (see 4.3.1.1.2), by adding additional information when describing the gap that the researcher is looking to fill. A typical example is:

> (16) Nonetheless, despite using similar strategies, the degree of functional overlap between the identified proteins in the different screens was very low. Importantly, these studies brought noteworthy knowledge on HIV-1/host interaction by identifying many cellular proteins that had not yet been related to HIV-1 infection. *Moreover*, the diversity of identified proteins suggests a vast complexity of host-virus interplay (Text 209).

In Example (16), *moreover* signals additional information when describing the failings of the previous studies. With regard to when *moreover* is used in the final third of the Introduction section, I expected that it would primarily be used in Move Three, *Way to the solution* (see 4.3.1.1.3), because it is providing supplemental description of the subjects in the experiment; however, all 12 instances were being used for Step B of the second Move, giving additional information concerning the gap. After examining all 34 instances of *moreover* in the Introduction section, all but two were used for providing additional explanation of findings or shortcomings in the field. This leads me to define *moreover* as being more primed for providing additional information regarding the field but, since in the majority of cases (24 out of 32) the author was expressing a view counter to one generally held in the field as in Example (16), I would argue that the word has a negative prosody in this section.

6.2.3.2 The uses of *moreover* in the Methods section

There are only three instances of *moreover* in the Methods section, which are as follows:

> (17) This means that people aged 45 or above are much less likely to have a tertiary (college/university) level education and less secondary (high school) education than people aged, 45 years old [45]. *Moreover*, in traditional families in China, a son (who lived with his parents after marriage) was usually more educationally- favoured over daughters (who moved to their in-laws' home on marriage) to ensure support for the parents in their old age, so males usually obtained more education than females [46] (Text 249).

(18) However, the sample size for males aged 18-44 was relative small (Table S2). *Moreover*, all the model variables were treated as categorical variables and we used the WLSMV method to estimate the model (Text 249).

(19) Continuous variables were categorized based on clinical reasoning. Bivariate analyses were carried out to assess the associations between two categorical variables by using the Chi-squared test. *Moreover,* in order to identify the predictors of the mental health, as measured by the K6, we carried out a multivariate logistic regression analysis, where clinically relevant factors were included in the model (Text 101).

Example (17) uses *moreover* to provide background on families in China, and give more detail about the subject, which is part of Move Five, *Describe what is being treated* (see 4.3.1.2.2). Example (18), which comes from the same text, was used to provide additional reasoning as to why a method was selected, possibly in order to counteract the shortcomings given in S1. This is similar to Example (19), in which *moreover* once again provides additional reasoning regarding why a certain method was employed in the analysis of the data. With such a small number of instances, it is not possible to offer an interpretation regarding priming or prosody.

6.2.3.3 The uses of *moreover* in the Results section

The DM *moreover* appears in the Results section 20 times, and in 19 of them it is the first word of the sentence. The one instance where it is not is as follows:

(20) All of the cancer samples exhibit significantly (P < 10-23) higher rates of novel SNPs than the normal sample; *moreover*, the ovarian tumor has a significantly (P < 10-39) higher rate of SNPs than the other cancer samples (Figure 6) (Text 141).

In Example (20), *moreover* is used to combine two results that were similar.

Every usage of *moreover* in the Results section focused on the results of the current experiment and did not add information to any segments regarding the field. As the usage, which typically looked like Example (21) below, described the current research, the prosody was positive because S2 was being used to elaborate and support S1.

(21) In the BI condition, subordinate mice showed a temporal profile for Immobility frequency characterized by very low levels, which increased only at the very end of the pain test. In contrast, the dominants' levels were almost reduced to zero through the whole experimental session. *Moreover*, BI subordinates were more likely to be immobile than dominants (Text 199).

6.2.3.4 The uses of *moreover* in the Discussion section

In 82 of 83 instances, the DM *moreover* was used to start a sentence; thus, while it can appear in other places in the sentence, it is primed to start one in MRAs. As in the previous section on *however*, I will separate the Discussion section into thirds, taking advantage of WordSmith Tools' (Scott 2008) ability to monitor where the word appears in the text. Table 6.15 will show the number of occurrences per third and the percentages thereof. Once again, it bears repeating that the section was only divided in this way to make the data more manageable, and should not be seen as commentary on the patterns being universally applicable.

Table 6.15: Frequency and percentages of occurrences across the three parts of the Discussion section

Parts of the Discussion section	Number of occurrences	Percentage
First third	29	36
Middle third	24	28
Final third	30	36

There is a fairly even distribution across the three parts of the Discussion section, regardless of whether the DM is being used to expand upon information from the field or the current experiment. In the first third, every usage is about the study, which may be because the first part of the section is generally used to describe the results and findings (see 4.3.1.3.1). An example of this can be seen below:

(22) In our study the rate of response was much higher, which we believe better represents the opinion of most of the professional groups. *Moreover*, our study population included a lower proportion of gynaecologists and midwives than the Swedish study, which may explain the lower percent of professionals in contact with cases (Text 234).

In Example (22), I would argue that through the use of *moreover*, the author is providing additional information and rationale about why the current study was better than the previous studies, which is Move Fourteen, *Compare with other studies* (see 4.3.1.4.4).

The middle third of the Discussion section reveals an even split between *moreover* being used for adding to information regarding the field and the current research. All 24 instances, however, are used to give additional information, either regarding the field or the experiment, which can support the findings of the current research.

> (23) However, given the growing body of evidence that neurons in primary and secondary visual cortex (V1 and V2) can perform some kind of higher-level processing and are sensitive to stimulus features in natural scenes (e.g., 78,79], we cannot rule out the possibility that some degree of higher-level sensitivity to stimulus orientation is already present in V1/V2. *Moreover*, recent evidence suggests the existence of even earlier neural encoding mechanisms of shape recognition already at the level of the retina [80]. (Text 245)

In Example (23), the authors give an example of why they chose to interpret their findings based on what others in the field had found. So, while *moreover* can be used to enhance the S1 of either the field or the current research, it is being done in a way that supports the rationale behind the authors' reasoning, which is similar to Move Thirteen, *Present a rationale* (see 4.3.1.4.3).

The final third of the Discussion section finds a two-to-one usage of *moreover* when discussing the field for the current experiment. However, much like the previous third, the authors are still using the field's experience as justification for their own results, as shown in Example (24):

> (24) Although we did not investigate the mechanism underlying the selective increase in T lymphocytes, previous studies indicated that BD T cells are partially protected against apoptosis [26]. *Moreover*, a recent report indicates that specific T-cell subsets are associated with sterile neutrophil-rich inflammation as observed in BD synovitis, which may explain the simultaneous increase of both T cells and PMN [27] (Text 128).

6.2.3.5 Priming of the uses of *moreover* across the IMRD sections

Moreover is defined by Fraser (2009) as an EDM, and nothing I have found in my corpus can dispute this. Relevance theorists have found that *moreover* is an adverbial that can occur in any part of an utterance (Rouchota 1996, p.8). In every instance in the MRA corpus, *moreover* signals that the authors want to relay further information pertinent to the information in S1. In the Introduction section, *moreover* is primarily used for adding evidence for why the authors are preparing to add reasons for their disagreement with findings from the field. There are too few instances in the Methods section to offer an analysis but, in the Results section, it seems that authors are using the DM to firm up the basis of their findings, as it is primarily focused on the current experiment. The Discussion section showed shifts in the priming of *moreover*, which may have been due to the Moves. When the authors were (re)stating their findings early in the section, they used *moreover* to shore up their claim, similar to the Results section, while later in the section they would use *moreover* to describe findings in the field that supported their own.

6.2.4 *Therefore*

There are 543 uses of *therefore* in the MRA corpus, making it the 206th most common word. It appears in the Introduction section 89 times, in the Methods section 59 times, in the Results 173 times and, in the Discussion section, it appears most frequently with 222 instances. This pattern is different from that of the previous two DMs, which generally had more instances in the Introduction sections than in the Results, but the reason for this may be due to what the authors want the readers to infer and partly just the semantics of the word.

Rouchota notes (1996, p.8) that *therefore*, like *moreover*, can occur initially, medially or finally in an utterance. According to Fraser's (2009) coding, *therefore* is an IDM, meaning its use is to signal that S2 is offering a potential inference to S1. Fraser (n.d. pp.1-3) separates IDMs into three types and categorises those into subclasses, paraphrased as follows:

Inferentials: SoA (S1) Causes the Result of the SoA of (S2)

A (*so*)

B (*then, in that case, under those conditions*)

C (*thus, therefore, hence, consequently*)

D (*as a result, for that reason, as a consequence, that's why*).

Telic: SoA (S1) Facilitates the Goal of (S2) (*so that, in order that, so as, for*).

Explanative: SoA (S1) Is a Result Caused by the SoA of (S2) (*because, since, for because of that, given that*).

He further claims that *therefore* is an inferential subclass that 'signals a conclusion which is "logically" or necessarily grounded in the message S1' (Fraser, n.d., p.86). He also defines *therefore* as being used to signal segments that are paratactic, meaning that there is a connection between two segments, but there is no conjunction connecting them. As such, I would expect to see *therefore* in the initial position of the sentence, without conjunctions.

6.2.4.1 The uses of *therefore* in the Introduction section

Therefore can occur at the beginning of a sentence:

> (25) Although turnover volume for catalase is very high, its affinity to H2O2 is relatively low. *Therefore*, when H2O2 accumulates in cells in even low concentration, it can cause oxidative damage to DNA, which then induces cancer or cell death (Nordberg & Arner, 2001) (Text 172).

in the middle of a sentence, where it frequently initiates a clause:

> (26) The formalin test lasts 30–50 minutes; *therefore*, allowing long enough observation of the animals to detect even a slight modulation of behaviour in both the subjects (Text 199).

and at the end of a sentence:

(27) Such geographic and racial-ethnic differences may influence the adaptation of prognostication as well as therapeutic protocols/algorithms shown to be effective elsewhere. An update and re-appraisal of Tanzanian ML diagnoses and characteristics possibly peculiar to this country is *therefore* needed (Text 12).

Table 6.16 shows the distribution of *therefore* across the Introduction section, showing its position within both the sentence and text.

Table 6.16: Distribution of *therefore* across sentences and texts of the Introduction section

Location	Sentence	Percentage	Text	Percentage
Beginning	68	75	15	16
Middle	12	13	26	29
End	9	12	48	55

In every instance, *therefore* signals a conclusion that should be clear from reading the issue described in S1. In Example (27), S1 outlines a problem, while S2 provides the solution, and *therefore* evidences the fact that S2 is the answer to S1. While no conjunctions are present in Example (26), there is a semicolon, which acts as the punctuation for a conjunction.[30]

The Introduction section presents a 2:1 ratio for *therefore* being used with information related to the field compared with information related to the current experiment. When it is used to describe the current experiment, it can be used to offer support for the reasoning of a choice made on the behalf of the authors:

(28) Under normoxic conditions facilitated diffusion of oxygen by oxy-myoglobin is not expected to play a significant role in oxygen transport in the myocardium [6, 17, 18]. *Therefore* we consider only the oxygen storage function of myoglobin in this computational study (Text 127).

When *therefore* is being used in conjunction with information from the field (see Examples 26 and 27), it can also be used to reinforce the rationale behind an authorial decision to act in some way and, in Example (27), as the basis of the entire experiment. Due to this possible priming, I looked at each of the 89 instances in the

---

[30] See online writing centre handbooks such as the one from University of Wisconsin, https://writing.wisc.edu/Handbook/Semicolons.html

Introduction section: 64 of them signalled support for any reason for any action undertaken by the authors. This suggests that the rhetorical effect of *therefore* is as a signal to persuade the reader that the authors took an action based on a theoretically sound idea.

6.2.4.2 The uses of *therefore* in the Methods section

In the Methods section, *therefore* can appear in different positions within a sentence and throughout the text, as seen in Table 6.17.

Table 6.17: Distribution of *therefore* in sentences and texts of the Methods section

Location	Sentence	Percentage	Text	Percentage
Beginning	38	65	15	25
Middle	6	10	18	30
End	15	25	26	45

However, Example (29) shows *therefore* occurring with a conjunction, which may contradict Fraser's (n.d., p.2) claim that *therefore* is paratactic:

> (29) The Chair of the Australian National University Human Research Ethics Committee advised that the evaluation work fell under the definition of quality assurance and *therefore* did not require formal approval by the Committee (Text 88).

Therefore can appear with conjunctions and does so nine times in this section.

As is the case with the usage in the Introduction section, *therefore* can be used to signal that S2 is focused on showing the consequence of S1, and is concerned with the authors' experiment. In this section, it is used only twelve times for describing the field, usually in terms of procedures (see Example (30)), with the remainder of the occurrences being focused solely on the experiment.

> (30) Since the myelin hydration layer (water layer) periodicity in the naive state roughly corresponds to 2.5 nm [54], then dehydration results in a reduction from roughly 16 nm to 13.5 nm [55–58]. *Therefore*, a lamellae periodicity of 16 nm and 13.5 nm for the naive and fixed axon respectively were used (Text 207).

Of the 59 occurrences in the Methods section, all but three were used to bolster the reasoning for performing a certain procedure or for choosing levels. This finding matches Step C, *Rationale*, of Move Five, *Describe what is treated* (see 4.3.1.2.2).

6.2.4.3 The uses of *therefore* in the Results section

There are 173 instances of *therefore* in the Results section and, as the table below shows, it can occur in different places within a sentence, as well as within a text.

Table 6.18: Distribution of *therefore* in sentences and texts of the Results section

Location	Sentence	Percentage	Text	Percentage
Beginning	143	83	48	28
Middle	12	7	63	36
End	18	10	62	36

While it is possible for *therefore* to occur throughout a sentence, it typically appears at the beginning of the sentence. There is a fairly balanced spread of *therefore* throughout the three thirds of the text. Much like its use in the previous sections, *therefore* is used sparingly (16 times) to describe the field although, even when *therefore* refers to the field, it is using the field in S1 to justify a decision for the experiment, as in:

> (31) Because Bdel2/ 3 currents inactivate so rapidly, their peak current could not be compared with modulation of Bdel2/ 2a currents (Vitko et al., 2008). *Therefore*, we took a pharmacological approach to determine whether the same slow pathway that inhibits CaV2.2/ 3 currents confers Bdel2/ 2a current inhibition (Text 158).

Generally, though, *therefore* is used by the authors to present additional strength or reasoning for their choices:

> (32) Our data demonstrates that the 14-3-3 gene is hyper-methylated in NHEM but hypomethylated in the highly metastatic melanoma cell line C8161.9. We *therefore* hypothesized that the hypermethylation status of 14-3-3 gene may be disrupted during melanoma progression as a result of genome-wide hypomethylation, which is a known characteristic of advanced-stage tumors (14) (Text 9).

In Example 32, the authors use the information in S1 from their data to make a decision in S2 but, considering the additional support from the field, I would argue that this is a device of persuasion that mirrors Step B, *Support with evidence*, which occurs in Move Seven, *Report the findings* (see 4.3.1.3.1). Every instance was used to explain or elaborate why the authors chose to interpret the data in a certain way.

6.2.4.4 The uses of *therefore* in the Discussion section

There are 222 instances of *therefore* in the Discussion section and, as Table 6.19 shows, it can occur in different places in a sentence, as well as in a text.

Table 6.19: Distribution of *therefore* in sentences and texts of the Discussion section

Location	Sentence	Percentage	Text	Percentage
Beginning	144	65	54	24
Middle	51	23	76	34
End	27	12	92	42

Therefore may appear at any point in the sentence or in the text in the Discussion section. In addition, the conjunction *and* appears to the immediate left of *therefore* 23 times, further illustrating how *therefore* may be used with conjunctions, as can be seen in Example (33):

> (33) The disruption of HP0256 and its effect on cell envelope architecture may modify the lipid profiles and/or membrane fluidity and *therefore* the function of the methyl-accepting chemotactic proteins (Text 95).

Example (33) shows a new use for *therefore* in this section. In the previous sections, *therefore* was a signal for the reasons that the authors chose to interpret or design their research in a certain way. Now, as can be seen in Example (33), it can also be used to provide a reason or explanation of why something might have happened. This signal matches Move Thirteen, *Present a rationale* (see 4.3.1.4.3). The phenomenon of using the information presented in S1 to justify a decision or interpretation remains:

> (34) To avoid the side effects we needed to understand the sensitivity of chemotherapeutic agents before the chemotherapy start, and let the treatment individualization. *Therefore*, before chemotherapy the drug sensitivity, to forecast it becomes necessary, especially. Most chemotherapeutic agents killed

tumor cells through inducing apoptosis, thus to investigate the regulatory factor in the procession of apoptosis will provide us an insight to know mechanism of the drug resistance (Text 60).

The following table shows the number of instances and percentages of how *therefore* is being used across field or experiment, and whether it is being used to provide justification or explanation.

Table 6.20: Use of *therefore* across the Discussion section

Use	Explanation	Percentage	Justification	Percentage
Field	4	2	52	23
Gap	34	15	132	59

When *therefore* is being used as an explanation, there is a sense of modality involved with the use of *may* or *could,* as shown in Example (35):

> (35) To put this in perspective, in the U.S. patients with mammographic lesions do not undergo (preoperative) biopsy if their risk of malignancy, radiologically, is considered to be less than 2% [25] (BI-RADS 3). *Therefore,* it could be argued that this low risk sub-group with non-bloody discharge may also be monitored safely (Text 105).

6.2.4.5 Priming of the uses of *therefore* across the IMRD sections

In the first three sections (Introduction, Methods and Results), *therefore* seems to use the information presented in S1 to justify the conclusion reached in S2, which is the basis for a decision on the part of the authors. In the Discussion section, it also is primed for justification, but can also be used to give a possible explanation of a result. The rhetorical principle behind the word seems to be that it signals to the reader that, based on the evidence, the author is interpreting the data or undertaking an action.

6.2.5 *So far*

Far, by itself, is not a DM. Instead, it is an adverb or adjective related to distance; however, when in a cluster, it can be used as a DM, as in *so far*. According to Biber, Conrad and Cortes (2004, p.372), a cluster is a group of repeated words that form a

'lexical bundle'. Bundles, however, can create additional problems for the analyst, who must identify the clusters and define their purpose as a DM. The following table shows the different clusters for *far* in the MRA corpus, using the following parameters: The word *far* had to collocate with at least one word and a maximum of five words; also, the cluster or bundle had to occur more than five times in my corpus in order to minimise random pairings or collocates.

Table 6.21: Bundles with *far* and number of occurrences in the MRA corpus

So far	39
Far from	16
Thus far	12
Far the	10
Far from the	8
Far more	8
As far	7
Are far	5
As far as	5
By far	5
The far	5
Far as	5

The cluster *so far* occurs most frequently, followed by *far from* and *thus far*, respectively. *So far* appears in my corpus 39 times. To the best of my knowledge, there have been no previous studies on *so far* as a potential DM. As *so far* highlights something that has not occurred up to this point in time and may be a negative projection of what has yet to occur, Fraser (2006) would have classed *so far* as being a TDM but, as of 2009, he no longer considers TDMs to be a class of DM, 'because it does not reflect a relationship between the interpretations of the two sentences, S1 and S2. However, Temporal may well be a type of DM; I just am unsure at the moment'. (2014, personal communication)

> (36) Nevertheless, most studies *so far* do not mention the time span over which patients reported their oral health-related quality of life, with a few exceptions (Text 111).

So far, as used here, shows that the authors believe that most other studies do not describe the time span of the experiment and that they should. WordSmith Tools (Scott 2008) was used for the examination of the position of *so far* to see if, or how, its position in a sentence could affect meaning.

Table 6.22: Distribution of *so far* in sentences and texts of the Introduction section

Location	Sentence	Percentage	Text	Percentage
Beginning	3	25	2	17
Middle	3	25	8	66
End	6	50	2	17

The cluster *so far* appeared on average mostly in the final third of a sentence, but it could also be used to start a sentence, as in:

> (37) *So far*, the prognostic value of plasma CT-proET-1 measurements has been proven only in patients during the acute phase of myocardial infarction [33, 34] (Text 219).

Here, *so far* also seems to have a slight negative tone concerning a failing in the past, in that proving the value of an experiment has been limited. *So far* may also appear at the end of a sentence, as shown here, where it signals a discourse boundary but may not be a discourse marker per se:

> (38) Although autologous conditioned serum (ACS, Orthokine®) proved slightly to moderately effective for alleviation of OA symptoms up to two years after treatment in human OA patients [15,16], many aspects of this therapy have remained unclear *so far* (Text 70).

There was only one instance of *so far* in the Methods section:

> (39) As an alternative, computing a (larger) SBM set based on all may allow detection of target sites that are recognised by a pairing structure different from those formed by the target sites known *so far*, which may be used to improve sensitivity. (Text 54)

However, in Example (39), *so far* is not a DM, but rather as part of a larger phrase, '*known so far'*, specifically concerning those target sites that had been found, although unlike Example (37), the discourse has not reached its boundary. Both instances suggest more investigation is needed, but Example (38) incorporates some rationale.

As for the Results section, there were only 8 instance of *so far*, while all of them were DMs, they could appear at all positions in a sentence or a text.

Table 6.23: Distribution of *so far* in sentences and texts of the Results section

Location	Sentence	Percentage	Text	Percentage
Beginning	5	63	3	37.5
Middle	2	25	3	37.5
End	1	12	2	25

In the Discussion section *so far* occur 18 times, with 17 of the 18 being DMs. As for sentence and textual positioning it can appear throughout.

Table 6.24: Distribution of *so far* in sentences and texts of the Discussion section

Location	Sentence	Percentage	Text	Percentage
Beginning	8	45	4	22
Middle	7	38	6	33
End	3	17	8	45

Once again, *so far* seems to signal a negative judgment because the therapy's benefits have remained unclear. No matter where it appeared in the sentence, as a DM, *so far* remained the same in content and force. I would argue that the prosody of the word is negative, as it is often partnered with words like *unclear, proven only, do not*, and so on. It seems to be used for comparison between the current study and previous studies.

However, when examining instances where the DM was used above sentence level, *so far* tended to occur in the middle of a text. This distribution is logical, as it follows the authors' apparent intent to signal something negative prior to its use. There was an imbalance within the IMRD sections of a text, with the Discussion sections containing nearly half (18) the usages. Frequencies of occurrence in the Results (8) and Introduction (12) sections were similar to each other, with nearly a quarter of the usage each, while the Methods section had only the one instance. Of the 39 instances of *so far* in the corpus, 37 were DMs.

6.2.5.1 Priming of the uses of *so far* across the IMRD sections

In the Introduction section, *so far* appears 12 times, and is used to indicate a difference from others' findings, thus assisting the authors in establishing a niche (Move One, Step B, *background work of others;* 4.3.1.1.1).

In the Results section, *so far*, with only eight instances, is used to indicate a gap in the current research, but one for which the authors leave room to fill later:

> (40) Our results *so far* analyzing the cross between TCR affinity and Itk expression suggest that higher affinity TCR could partially compensate for the absence of Itk for the development of CD4+ T cells (Text 208).

In Example (40), the authors claim that they do not yet have definitive results, but they believe that they are on the right track. In this section, *so far* is primed to discuss the current research, but the negativity found in the Introduction section is no longer present.

So far appears 18 times in the Discussion section, and is used to describe gaps in the field, as well as being generally used to set the research apart from that by others.

> (41) Interestingly, we observed a small early body inversion effect. *So far* only one EEG study investigated the effect of body inversion on the P1 component [59] but failed to find a significant effect (Text 245).

Here, the DM is used to highlight that there has not been sufficient work in the field in this regard, and would be placed in Move Fourteen, *Compare with other studies* (see 4.3.1.4.4).

Looking at the DM across all the sections, there is a negative prosody when it is used to describe the field but, when describing the current work, it seems to be mainly used to indicate a result that the authors are not able to prove conclusively but which they nonetheless still believe is valid.

6.2.6   *As mentioned*

In this section, I propose a different type of DM, one that breaks Fraser's (1990; 1999; 2006; 2009a) rule of S1 and S2 being congruent. I am describing a signal that an author may employ to remind the reader that the information in S2 has been discussed previously in the work and, since it is offering commentary on the more 'global' (Schiffrin 1987) aspect of the text, I have followed Teufel (1998) and termed this new class *Meta-discourse markers* (MDM). The example I offer is *as mentioned.* It occurs in my corpus 18 times. An example is:

> (42) The coefficients are found by solving a quadratic programming problem. *As mentioned* in the last section, the assignment of class weights and is critical in the implementation for highly unbalanced data. In this study, we assigned the parameter with following formula (Text 90).

Here, the authors signal that what they are using has already been discussed and bears repeating and in effect, are repeating it in order to reinforce the information. Teufel (1998) argues that authors need to persuade readers, and employ various rhetorical principles in order to do so. She shows that one of the ways authors may do this is by offering what she termed 'Meta-comments' (1998, p.46). These comments differ from other DMs, as they signal information as opposed to the content that other DMs connect. In an email from Fraser (2014, personal communication), he stated that he does not consider *as mentioned* a DM:

> The *as mentioned* is to be taken literally, and it refers back to something the speaker said but perhaps much earlier than the preceding sentence. In short, I don't think the term specifies a relationship between the interpretations of the two sentences.

I would disagree, as I feel there is interpretation involved in, for example, excerpts such as the following:

> (43) Our patients were extremely ill with low cardiac index and evidence of on going myocardial damage *as mentioned* above. It is possible that in such a state, intervening with inotropes may mitigate the overall neurohormonal activation (including inflammation) (Text 44).

Here, the authors are reminding the reader that they feel that this point is important and that it is because of this previously-given information that the results are as they are.

With only 18 instances and no previous studies, it would be difficult to draw conclusions, aside from the fact that MDMs are used to refer to information that is not in the adjacent position, but do signal something beyond the purely semantic relationship between S2 and S1.

6.3 Discussion and conclusion

This chapter examined five DMs to show how they might be used and what they might be used for. The following tables will review the possible priming of the DMs across the four IMRD sections, and the possible Moves that they signal.

Table 6.25: Discourse markers, priming and Moves in the Introduction section

Discourse Marker	Primed for	Moves
However	Contrasting the field	1. Present study situation
Moreover	Providing additional explanations to findings or shortcomings of the field	2. Describe the problem
Therefore	Signalling the reader that the authors took an action based on a theoretically sound idea	3. Describe way to the solution
So far	Signalling shortcomings of the field	1. Present study situation -or- 2. Describe the problem
As mentioned	Reinforcing the authors' solution to problem in field	2. Describe the problem

Table 6.26: Discourse markers, priming and Moves in the Methods section

Discourse Marker	Primed for	Moves
However	Showing how the authors' work will differ from previous studies	6. Describe experiment
Moreover	--	--
Therefore	Signalling a justification on the part of the author for making a choice for a procedure	6. Describe experiment
So far		
As mentioned	Signalling an important reason for making a choice that had previously been described	6. Describe experiment

Note: *moreover* and *so far* occurred too infrequently in this section to be discussed

Table 6.27: Discourse markers, priming and Moves in the Results section

Discourse Marker	Primed for	Moves
However	Showing contrast with the results from the current experiment and the expectations of the authors	7. Report the findings
Moreover	Adding information to the current experiment's	7. Report the findings

	results	-or- 9. Describe statistical tests and state findings
Therefore	Signalling the authors' justification for making a choice of procedure	7. Report the findings -or- 8. Review the experiment
So far	Indicating that, although there was a problem in the current research, the authors are leaving room to fill the gap later	8. Review the experiment
As mentioned	Signalling an important reason for making a choice that had been described previously.	8. Review the experiment

Note: There are too few instances of *as mentioned* to make any significant claims.

Table 6.28: Discourse markers, priming and Moves in the Discussion section

Discourse Marker	**Primed for**	**Moves**
However	Showing a contrast with the results from the current experiment and the expectations of the authors and also may be used to signal limitations or suggestions for future research	11. Discoveries from research -or- 13. Rationale for the solution -or- 14. Comparison to previous studies -or- 15. Limitations -or- 16. Suggestions for future research
Moreover	Adding information to the results of the current experiment	7. Report the findings -or- 9. Describe statistical tests and state findings
Therefore	Signalling the authors' justifications for making a choice of procedure and also may be used to explain why the result might have occurred	11. Discoveries from research -or- 13. Rationale for the solution -or- 14. Comparison to previous studies
So far	Showing failings in the field that need to be addressed	14. Comparison to previous studies -or- 15. Limitations -or- 16. Suggestions for future research
As mentioned	Signalling an important reason for making a choice that had been described previously.	11. Discoveries from research -or- 12. Treatment review -or- 13. Rationale for the solution

Note: There are too few instances of *as mentioned* to make any significant claims.

The DMs can be used to signal a wide variety of Moves, and their focus, the field or the experiment, can vary from one section to another, thus reinforcing the idea of the complexity of the effects of local prosody on how a word is primed. From the previous four tables, I would suggest that DMs in the Introduction section are primarily used to describe the field, while in Methods and Results, the focus swings to the current experiment, and the Discussion section can have both foci. This could be expected, given the functions of the four sections in themselves. The following table shows the five DMs this paper addresses, and how many times each occurs per section of IMRD.

Table 6.29: Frequency of discourse markers by section of text

Discourse Marker	Introduction	Methods	Results	Discussion	Total Number.
However	236	65	354	608	1263
Moreover	34	3	25	83	145
Therefore	89	59	173	222	543
So far	12	0	8	17	37
As mentioned	0	5	3	10	18
Total number	371	132	563	940	2006

From this table, it can be seen that the Discussion section has the greatest number of the five DMs that I examined, while the Methods section has the least. The Introduction section has the third highest total, but it is important to note that the Introduction (371 out of 153,257 words) is generally half the size of the Results section (563 out of 311,394 words), making the DM per word of the Introduction section second to that of the Discussion section. The Introduction and Discussion sections are extremely rich DM environments, which may be why so much research has gone into mapping the Rhetorical Moves of these sections (see 4.1.4).

The goal of this chapter was to study how authors of MRAs used DMs, as well as examining any possible pragmatic qualities that were not readily apparent, through the use of LP and SP. In addition, the current work has potentially identified a fifth type of DM, which has potential value in that MDMs may be used in MRAs to bolster claims, and may have a strong rhetorical value in assisting with the repetition of information, which can be seen as having a persuasive quality.

Chapter 7: Conclusion

7.1.   Introduction

This study began with a primary aim: to conduct a genre-based analysis of medical research articles (MRAs) in order to characterise the genre for non-native and novice writers. To do this, I developed a sample of members of the discourse community (DC), mapped out the Rhetorical Moves of the articles in the MRA corpus, examined stance through a study of 13 reporting verbs, and lastly, reviewed examples of four different types of discourse markers (DMs) and identified a possible fifth type. Since Swales' (1990) work on genre analysis (GA) appeared, undertaking studies based on genre has been especially popular in the English for Specific Purposes community. Since my impetus was in trying to help non-native speakers (NNS) and novice writers better understand any possible linguistic expectations of the genre of MRAs, GA was the most obvious choice with which to examine MRAs. However, I wanted to go beyond establishing that MRAs form a genre of their own, within the larger category of scientific research articles, by conducting a relatively broad-ranging study. Because of its utility for lexical investigations, I applied several techniques drawn from corpus analysis, such as semantic prosody and lexical priming, as tools to locate and study selected words and phrases and their possible colouring that could indicate stance on the part of the author, or other potential features of the genre.

In the next section I will review my research questions and incorporate brief summaries of the chapters, to indicate their contributions.

7.2   Review of research questions and findings

The first question was how to establish MRAs as a genre. To accomplish this, I followed Swales' (1990) socio-rhetorical framework for Moves, and located and consulted a small sample of members of the DC. Through a series of interviews with doctors who read and publish MRAs, as well as with freelance academic editors and editors of major journals, I was able to develop a definition of the genre that incorporated different aspects that the DC felt to be part of the genre, i.e. how the research is conducted and by whom:

> *A piece of original written research on a specific medical problem, either in previous treatments or as a new issue, carried out according to a design whose findings are measured and reported along with a discussion of the implications for the community.*

According to the majority of the participants, an MRA needs to be an original piece of work: this was a requirement mentioned by 11 of the 15 participants, in order to avoid (self) plagiarism. A second issue raised by the DC and addressed by this definition was that the MRA had to include measurable data and demonstrate importance to the field. While this definition works for MRAs, it raises the question of what can be done to help differentiate this genre from other research related genres, and this leads to the issue identified in the second research question.

The second research question was how best to identify the typical Rhetorical Moves of an article that belongs to this genre. MRAs have specific lexis and jargon centred in medicine, which alone helps distinguish them as belonging to a genre separate from other research articles. However, following Swales' (1990) work in GA, I also located and analysed the possible Rhetorical Moves found in such articles, which involved looking for shifts in the flow of information. There have been previous studies of MRAs and the Moves found within them; however, as one study found (Li & Ge 2009) Moves do undergo change over time. Accordingly, I wanted to update previous studies by providing an analysis of more recent articles, written since the turn of the twenty-first century. In addition, I hoped to derive findings from a larger corpus than

those previously studied, which typically either drew on a sample of 50 or fewer articles or which studied a single section, such as Results. Utilizing a corpus of 250 articles that I was able to compile from texts owned by the offices of the U.S. Institutes of Health (see Chapter 2 for a discussion on the compilation of the corpus), I systematically analysed the main sections of each article and noted similarities and differences in Moves I found with those of previous studies. I identified 17 Moves with four optional Moves (see Chapter 4), which is a greater number of Moves than previous studies had found, as well as some differences between these previous studies and my own, which may be due to the categorization of the different Moves. This work also benefited from having a larger corpus to draw from, as some of the earlier studies were done before the development of and/or easy access to PubMed, so that their compilers may not have had access to its large collection of texts. Having a larger corpus made it possible to seek to identify more Moves that might not have been located in the smaller ones. Unlike some of the previous studies, the current work has attempted to connect the Moves to rhetorical principles in an effort to more clearly define the use of a particular Move. A working definition of the genre and its Rhetorical Moves enabled me to do a bottom-up study focused on specific language patterns, such as how authors may use certain words and phrases to indicate stance as well as cohesion.

Reporting verbs are obviously important in an MRA, since their main function is to report on research conducted by the authors and their previous sources. The third research question was whether it might be possible to identify the author's stance and factiveness through an analysis of reporting verbs. I examined 13 different reporting verbs in terms of stance and factiveness, i.e., did the author agree with the information (factive), disagree (counter-factive), or show no feelings one way or another (non-factive)? I also utilized two concepts within corpus analysis, semantic prosody and lexical priming, in an attempt to categorise the 13 verbs (see Chapter 5). Semantic prosody is the theory that words are imbued with certain qualities through their collocates, such as factiveness, non-factiveness, and counter-factiveness, while lexical priming is the idea that all words are primed by a variety of factors like collocates, grammar and context. Using these lenses to review the 13 verbs throughout the corpus in their different structural forms (e.g. present tense and past tense) I also found that

reporting verbs in MRAs have different categorisations in regards to stance when compared to their definitions by Wyse (2009) (see 5.4).

Finally, research in response to the fourth research question investigated the characteristics of the different types of discourse markers (DMs) that authors may employ in writing an MRA and how they are being used, i.e. in agreement with the previous segment (the part of text being connected to the DM) or not. Following Fraser and Blakemore, I identified four types of DMs: contrastive markers (CDMs), elaborative markers (EDMs), inferential markers (IDMs), temporal markers (TDMs) and I also found a fifth type, meta-discourse markers (MDMs), which had not previously been described (see Chapter 6). An example of each type of DM was selected and examined throughout the corpus in terms of what segments of text were linked and whether there was any stance or factiveness being displayed. The four previously studied DMs were used much in the same way in the MRA corpus as in previous analyses of different genres; however it was my finding of the fifth type of DM that is new and potentially useful.

While traditional studies have found DMs to connect segments next to each other, the fifth DM is a meta-marker that cuts across the boundaries of having two segments in corresponding positions. The discovery of this fifth type of discourse marker, the meta-marker, exemplified by *as mentioned*, will allow researchers to look beyond certain constraints such as a need for contiguity when defining discourse markers. Its presence suggests the possibility of expanded research on how writers in general, and writers of medical research articles in particular, further substantiate their credibility and the strength of their research through a kind of repetition which creates the impact of self-attribution. The other major finding was that it is in the Discussion section of the article that the majority of these particular DMs are located.

Probably the most significant contribution of this study is the analysis of a large, specialised corpus of MRAs using a combination of techniques to conduct a genre analysis while incorporating elements of corpus analysis. In so doing, this work has not only identified a new type of discourse marker, and expanded our understanding of Rhetorical Moves in general as well as within a particular genre, it has also further exemplified the utility of semantic prosody and lexical priming for the examination of

authorial stance and features of authorial and textual persuasion present in medical research writing.

The findings from this study can be used pedagogically in that the Rhetorical Moves have been mapped out, allowing for the formation of a template for the MRA. The examination of the thirteen selected reporting verbs may help non-native speakers of English to better understand the range of meanings that such verbs can carry, reinforcing the fact that words may have different meanings than non-native speakers may expect.

7.3   Limitations

In the current work, the main limitations are that the sample of the discourse community was quite restricted in its coverage. In addition, we must consider this work to be only a snapshot, since the genre has and will continue to undergo changes over time. Future studies could follow similar steps to look at different genres in order to help differentiate them, and to improve the analysis of a genre. Also this work only really analysed articles that followed the problem-solution model and as such any findings within may only be applied to such articles.

While the genre of medical research articles has been studied in some detail by previous researchers, a corpus comprising a sizeable number of full texts has not until now been compiled and examined, using the combination of genre analysis and corpus analysis in conjunction with semantic priming and lexical prosody. I followed Swales' (1990) framework to design and elicit a sample of the discourse community, to include 10 doctors representing several different countries and five editors of medical journals representing different sub-fields. I then drew on Swales' initial explanation of Moves in order to inventory what I found to be 17 Moves in the genre, four of which were optional. Because authors of MRAs attempt to persuade their readership that their research questions, methodology, and findings are valuable in that they represent good science, and are worth believing and even adopting in some instances, I then looked at authorial stance, and its expression through close examination of selected reporting verbs and discourse markers. I then looked at how

stance in MRAs may be expressed through reporting verbs and discourse markers. What I have found, however, can be expected to change.

New kinds of corpora will doubtless be developed, as corpus analysis will develop new techniques to undertake new applications. The world of medical research publishing is constantly evolving and with the emergence of e-readers and on-line storehouses (e.g. PubMed) of articles, the genre will continue to change. This work studies those articles that were open access and published as peer-reviewed articles in English between 2001 and 2011. It is not meant to be comprehensive of all articles in the genre, but future studies may well show comparable findings.

7.4   Final words

A genre-based discussion can assist novice and non-native writers who may not understand how best to arrange content in a particular genre in ways that meet expectations of its discourse community for diction, style, organization, and credibility. This work, a genre-based examination of a corpus of medical research articles, was undertaken with the aim of trying to derive a better understanding of how and why such articles are developed, hoping that this work will be of pedagogical use for those who develop materials for non-native speakers of English as well as novice writers who now find themselves entering the larger field of medical research and publication. It is the hope that this work will help in the development of rubrics or templates that can be used in the teaching of professional writers, either to non-native or novice writers.

References

Adams-Smith, D., 1984. Medical discourse: aspects of authors' comments. *The ESP Journal*, vol. 3, no. 1, pp.25-36.

Aijmer, K., 2002. *English discourse particles: evidence from a corpus*. Amsterdam: John Benjamins.

Aijmer, K., Foolen, A. & Simon-Vanderbergen, A., 2006. Discourse particles in the perspective of heteroglossia. In K. Fischer, ed. 2006. *Approaches to discourse particles*, Amsterdam: Elsevier, pp.101-114.

Aijmer, K., & Simon-Vandenbergen, A., 2011. Pragmatic markers. *Discursive Pragmatics*, vol. 8, pp.223-247.

Allen, D., 2009. Lexical bundles in learner writing: an analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education*, vol. 1, pp.105-127.

Allison, D., 2006. Review of JM Swales, 'Research Genres'. *Journal of English for Specific Purposes*, vol. 5, pp.244-246.

Altenberg, B. & Granger, S., 2001. Grammatical and lexical patterning of native and non-native student writing. *Applied Linguistics*, vol. 22, no. 2, pp.173-195.

Anthony, L., 2013. Developing AntConc for a new generation of corpus linguists. Proceedings of the Corpus Linguistics Conference (CL 2013), July 2226, 2013. Lancaster: Lancaster University, pp.14-16. Viewed 10 November 2014, http://www.laurenceanthony.net/research/20130722_26_cl_2013/cl_2013_paper_final.pdf.

Ariel, M., 1993. Pragmatic operators. In R. Asher & J. Simpson, eds. 1993 *Encyclopedia of languages and linguistics*, 1st ed. vol. 6, Oxford: Pergamon Press, pp.3250-3253.

Axelrod, R. & Cooper, C.R., 2001. *The St. Martin's guide to writing*. New York: St. Martin's Press.

Azevedo, L.F., Canario-Almeida, F., Almeida, F.J., Costa-Pereira, A., Winck, J.C. & Hespanhol, V., 2011. How to write a scientific paper: writing the methods section. *Revista Portuguesa Pneumologia,* vol. 17, no. 5, pp.232-238.

Bach, K., 2005. The top ten misconceptions about implicature. In B. Birner & G. Ward, eds. 2006. *Drawing the boundaries of meaning: neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*. Amsterdam: John Benjamins, pp.21-30.

Bach, K. & Harnish, R., 1979. *Linguistic communication and speech acts*. Cambridge, MA: MIT Press.

Badger, R., 2003. Legal and general: towards a genre analysis of newspaper law reports. *English for Specific Purposes*, vol. 22, no. 3, pp.249-263.

Baker, A., 1914. *A concordance to the poetical and dramatic works of Alfred, Lord Tennyson.* New York: Macmillan.

Baker, S.E. & Edwards, R., 2012. *How many qualitative interviews is enough?* UK: National Centre for Research Methods. Viewed 16 October 2014, http://eprints.ncrm.ac.uk/2273/4/how_many_interviews.pdf.

Barbara, L. & Scott, M., 2014. Homing in on a genre: invitations for bids. In F. Bargiela-Chiappini & C. Nickerson, eds. 2014. *Writing business: genres, media and discourses,* London: Longman, pp.227-254.

Bargiela-Chiappini, F. *&* Harris, S.J., 1997. *Managing language: the discourse of corporate meetings*. Amsterdam: John Benjamins.

Bawarshi, J. & Rieff, M., 2010. *Genre: an introduction to history*, *theory*, *research*, *and pedagogy.* West Lafayette, IN: Parlor Press.

Bazerman, C., 2000. *Shaping Written Knowledge: The Genre and Activity of the Experimenterimental Article in Science*. Fort Collins, CO: WAC Clearinghouse Landmark Publications in Writing. Viewed 7 August 2014, http://wac.colostate.edu/books/bazerman_shaping/.

Bazzanella, C., 1990. Phatic connectives as interactional cues in contemporary spoken Italian. *Journal of Pragmatics*, vol. 14, pp.629-647.

Beaufort, A., 1997. Operationalizing the concept of discourse community: a case study of one institutional site of composing. *Research in the Teaching of English*, vol. 31, pp.486-529.

Becker, J., 1975. *The phrasal lexicon.* Cambridge, MA: Bolt Beranek and Newman, Inc.

Bednarek, M., 2008. Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory*, vol. 4, no. 2, pp.119-139.

Besen, J. & Gan, S., 2014. A critical evaluation of clinical research study designs. *Journal of Investigative Dermatology*, 134, viewed at http://www.nature.com/jid/journal/v134/n3/full/jid2013545a.html

Bhatia, V., 1987. Language of the law. *Language Teaching*, vol. 20, no. 4, pp.227-234.

Bhatia, V., 1993. *Analysing genre: language use in professional settings*. London: Longman.

Bhatia, V., 1999. Integrating products, processes, purposes and participants in professional writing. In C.N. Candlin & K. Hyland, eds. 2014. *Writing: texts, processes and practices*. London: Longman, pp.21-39.

Bhatia, V., 2002. Applied genre analysis: a multi-perspective model. *Ibérica*, vol. 4, pp.3-19.

Biber, D., 1993. Representativeness in corpus design. *Literary and Linguistic Computing*, vol. 8, pp.243-257.

Biber, D., 1994. An analytical framework for register studies. In D. Biber & E. Finegan, eds. 1994. *Sociolinguistic perspectives on register*, Oxford: Oxford University Press, pp.31-56.

Biber, D., 2006. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, vol. 5, no. 2, pp.97-16.

Biber, D., 2010. What can a corpus tell us about registers and genres. In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*, New York: Routledge, pp.241-254.

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

Biber, D., Conrad, S. & Cortes, V., 2004. If you look at … : lexical bundles in university teaching and textbooks. *Applied Linguistics*, vol. 25, no. 3, pp.371-405.

Biber, D. & Finegan, E., 1988. Adverbial stance types in English. *Discourse Processes*, vol. 11, no. 1, pp.1-34.

Biber, D. & Finegan, E., 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text-interdisciplinary Journal for the Study of Discourse*, vol. 9, pp.93-124.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E., 1999. *Longman grammar of spoken and written English*, Harlow: Longman.

Blakemore, D., 1987. *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D., 1992. *Understanding utterances*. Oxford: Blackwell.

Blakemore, D., 1996. Are apposition markers discourse markers? *Journal of Linguistics,* vol. 32, no. 2, pp.325-347.

Blakemore, D., 2002. *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Blakemore, D., 2006. Divisions of labour: the analysis of parentheticals. *Lingua*, vol. 116, no. 10, pp.1670-1687.

Blecha, J., 2012. *Building specialised corpora*. MA. Masaryk University.

Bloch, J., 2009. The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, vol. 13, no. 1, pp.59-78.

Bloch, J., 2010. A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, vol. 2, no. 2, pp.219-244.

Boden, D., 1994. *The business of talk: organizations in action*. Cambridge: Polity.

Bondi, M., 2001. Small corpora and language variation: reflexivity across genres. In M. Ghadessy, A. Henry & R. Roseberry, eds. 2001. *Small corpus studies and ELT: theory and practice*. Amsterdam: John Benjamins, pp.135-173

Bowker, L. & Pearson, J., 2002. *Working with specialised language: a practical guide to using corpora*. London: Routledge.

Braun, S., 2006. *Corpora 4 Learning*. Viewed 27 September 2013, http:/www.corpora4learning.net/.

Brett, P., 1994. A genre analysis of the results section of sociology articles. *English for Specific Purposes*, vol. 13, no. 1, pp.47-59.

British National Corpus, 2007. *The British National Corpus*. version 3 (BNC XML Edition). Viewed December 17 2013, http://www.natcorp.ox.ac.uk/.

Bronckart, J., 2003. *Atividade de linguagem, textos e discursos: por um interacionismo sócio-discursivo* (*Language, text and discourse activity: towards a socio-discursive interactionism*). Translated by A. Machado and P. Cunha, 2nd ed., São Paulo: EDUC.

Bruce, I., 2008. Cognitive genre structures in methods sections of research articles: a corpus study. *Journal of English for Academic Purposes*, vol. 7, no. 1, pp.38-54.

Bruce, I., 2009. Results sections in sociology and organic chemistry articles: a genre analysis. *English for Specific Purposes*, vol. 28, no. 2, pp.105-124.

Burke, K., 1969. *A rhetoric of motives*. Los Angeles: University of California Press, Viewed 3 February 2014, http://jasonthamdotcom.files.wordpress.com/2013/10/a-rhetoric-of-motives-by-kenneth-burke.pdf.

Carter, R. & McCarthy, M., 2006. *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*. Cambridge: Cambridge University Press.

Celle, A. & Huart, R., 2007. *Connectives as discourse landmarks*. Amsterdam: John Benjamins.

Chafe, W.L. & Nichols, J., 1986. *Evidentiality: the linguistic coding of epistemology*. Norwood, NJ: Ablex Publishing Corporation.

Chang, C. & Kuo, C., 2011. A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes*, vol. 30, no. 3, pp.222-234.

Channell, J., 2000. Corpus-based analysis of evaluative lexis. In S. Hunston and G. Thompson, eds. 2000. *Evaluation in text: authorial stance and the construction of discourse*. Oxford: Oxford University Press, pp.38-55.

Chen, Q. & Ge, G., 2007. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes*, vol. 26, no. 4, pp.502-514.

Cobb, T., 2002. *Web Vocabprofile.* Viewed 15 January 2013 http://www.lextutor.ca/vp/.

Connor, U., Precht, K. & Upton, T., 2002. Business English: learner data from Belgium, Finland, and the US. In S. Granger, J. Hung & S. Precht, eds. 2002. *Computer learner corpora, second language acquisition, and foreign language teaching*. Amsterdam: John Benjamins, pp.175-194.

Cooper, C., 1985. *Aspects of article introductions in IEEE publications*. MSc, University of Aston.

Copage, J., 1999. *First certificate writing*. Harlow: Pearson Education Limited.

Corbett, J., 2006. Genre and genre analysis. In E. Brown & A. Anderson, eds. 2006. *Encyclopedia of Language & Linguistics*. Boston: Elsevier, pp.26-32

Coupland, J., 2009. Discourse, identity and change in mid-to-late life: interdisciplinary perspectives on language and ageing. *Ageing and Society*, vol. 29, no. 6, pp.849-861.

Coutinho, M. & Miranda, F., 2009. To describe textual genres: problems and strategies. In C, Bazerman, A. Bonini & D. Figueiredo, eds. 2009. *Genre in a changing world*. Fort Collins, CO: The WAC Clearinghouse and Parlor Press, pp.35-55.

Coxhead, A., 1998. *An academic word list*. School of Linguistics and Applied Language Studies, Victoria University of Wellington.

Crookes, G., 1986. Toward a validated analysis of scientific text structure. *Applied Linguistics*, vol. 7, no. 1, pp.57-70.

Davies, M., 2008. *The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present.* Viewed 15 January 2013, http://www.americancorpus.org.

De Groot, E., 2008. *English annual reports in Europe: a study on the identification and reception of genre characteristics in multimodal annual reports originating in the*

Netherlands and in the United Kingdom. PhD. University of Utrecht. Viewed 14 December 2013, http://www.lotpublications.nl/publish/issues/Groot_de/index.html.

Department of Biology, Bates College, 2002. *How to write a paper in scientific journal style and format.* Viewed 15 January 2014, http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWtoc.html.

Ding, H., 2007. Genre analysis of person statements. *English for Specific Purposes*, vol. 26, no. 3, pp.368-392.

Durrant, P. & Mathews-Aydınlı, J., 2011. A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes*, vol. 30, no. 1, pp.58-72.

Englebretson, R., 2007. *Stancetaking in discourse: subjectivity, evaluation, interaction.* Philadelphia: John Benjamins.

Erman, B., 1992. Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction. *Language Variation and Change*, vol. 4, no. 2, pp.217-234.

Field, M., 1997. The role of factive predicates in the indexicalization of stance: a discourse perspective. *Journal of Pragmatics*, vol. 27, no. 6, pp.799-814.

Fischer, K., 2000. *From cognitive semantics to lexical pragmatics.* New York: Mouton de Gruyter.

Fishelov, D., 1993. *Metaphors of genre.* University Park, PA: Pennsylvania State Press.

Flowerdew, L., 2005. An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies. *English for Specific Purposes*, vol. 24, no. 3, pp.321-332.

Flowerdew, L., 2013. Corpus-based discourse analysis. In J. Gee & M. Handford, eds. 2013. *The Routledge handbook of discourse analysis*, London: Routledge, pp.174-188

Foley, M. & Hall, D., 2003. *Advanced learner's grammar.* Harlow: Pearson Education Limited.

Foss, S., 2005. Theory of visual rhetoric. In K. Smith, S. Moriarty, S. Barbatsis & K. Kenney, eds. 2004. *Handbook of visual communication: theory, methods and media*, Mahwah, NJ: Lawrence Erlbaum, pp.141-152

Francis, W. & Kučera, H., 1979. *Brown Corpus manual: Brown University Department of Linguistics.* Viewed 13 October 2014, http://www.hit.uib.no/icame/brown/bcm.html.

Francis, W. & Kučera, H., 1982. *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Fraser, B., 1987. Pragmatic formatives. In J. Verschueren & M. Papi, eds. 1987. *The pragmatic perspective*, Amsterdam: John Benjamins, pp.179-194.

Fraser, B., 1988. Types of English discourse markers. *Acta Linguistica Hungarica*, vol. 38, pp.19-33.

Fraser, B., 1990. An approach to discourse markers. *Journal of Pragmatics*, vol.14, pp.383-395.

Fraser, B., 1996. Pragmatic markers. *Pragmatics*, vol. 6, pp.67-90.

Fraser, B., 1999. What are discourse markers? *Journal of Pragmatics*, vol. 31, pp.931-952.

Fraser, B., 2006. Towards a theory of discourse markers. In K. Fischer, ed. 2006. *Approaches to discourse particles*. Amsterdam: Elsevier, pp.189-204.

Fraser, B., 2009a. An account of discourse markers. *International Review of Pragmatics*, vol. 1, no. 2, pp.293-320.

Fraser, B., 2009b. Topic orientation markers. *Journal of Pragmatics*, vol. 41, no. 5, pp.892-898.

Fraser, B., n.d. Outline of discourse markers. Viewed 12 February 2014, personal website, http://people.bu.edu/bfraser/IDM%20Papers/Outline%20%20English%20IDMs.doc.

Freadman, A., 1988. Untitled: (on genre). *Cultural Studies*, vol. 2, no. 1, pp.67-99.

Frodesen, J. & Eyring, J., 2000. *Grammar dimensions*. Boston: Heinle & Heinle Publishers.

Frow, J., 2013. *Genre*. New York: Taylor & Francis.

Fryer, D., 2007. *Multidimensional genre-based discourse analysis of a corpus of English-language medical research articles*. MA Thesis, University of Oslo.

Fryer, D., 2012. Analysis of the generic discourse features of the English-language medical research article: a systemic-functional approach. *Functions of Language*, vol. 19, no. 1, pp.5-37.

Gillett, A., 1999. *Using English for academic purposes*. Viewed 16 March 2012, https://www.uefap.com/writingforapurpose/vocabulary/reporting_verbs.doc.

Gledhill, C., 2000. The discourse function of collocation in research article introductions. *English for Specific Purposes*, vol. 19, no. 2, pp.115-135.

Gledhill, C., 2009. Colligation and the cohesive function of present and past tense in the scientific research article. In D. Banks, ed. 2009. *Les temps et les textes de*

spécialité. Paris: L'Harmattan, pp.65-84. Viewed 2 February 2014, https://univ-paris-diderot.academia.edu/ChristopherGledhill/Papers.

Goffman, E., 1981. Footing. In E. Goffman, ed. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press, pp.124-159.

Grice, H.P., 1975. Logic and conversation. In P. Cole & J. Morgan, eds. 1975. *Syntax and semantics*, *Vol. 3, Speech Acts*. New York: Academic Press, pp.41-58.

Grote, B., 1998. Representing temporal discourse markers for generation purposes. In *Coling/ACL workshop on discourse relations and discourse markers*, pp.22-28. Viewed 8 October 2013, http://www.aclweb.org/anthology/W98-0304.

Guthrie, D., Guthrie, L. & Wilks, Y., 2008. An unsupervised probabilistic approach for the detection of outliers in corpora. *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08),* pp.3409-3413. Viewed 4 February 2014, http://www.lrec-conf.org/proceedings/lrec2008/pdf/866_paper.pdf.

Halliday, M.A.K. & Hasan, R., 1976. *Cohesion in English*. London: Longman.

Halliday, M.A.K. & Matthiessen, C., 2013. *An introduction to functional grammar*. 3rd ed. London: Arnold.

Handford, M., 2010a. What can a corpus tell us about specialist genres. In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. New York: Routledge, pp.255-269

Handford, M., 2010b. *The language of business meetings*. Cambridge: Cambridge University Press.

Harewood, G. & Hendrick, L., 2010. Prospective, controlled assessment of the impact of formal evidence-based medicine teaching workshop on ability to appraise the medical literature. *Irish Journal of Medical Science,* vol. 17, no. 1, pp.91-94.

Henry, A. & Roseberry, R., 2001. A narrow-angled corpus analysis of moves and strategies of the genre: letter of application. *English for Specific Purposes*, vol. 20, no. 2, pp.153-167.

Hoey, M., 2005. *Lexical priming: a new theory of words and language*. London: Routledge.

Holmes, R., 1997. Genre analysis, and the social sciences: an investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes,* vol. 16, no. 4, pp.321-337.

Hooper, J.B., 1974. *On assertive predicates*. Indiana University Linguistics Club. Viewed 13 August 2013, http://openlibrary.org/books/OL19765296M.

Hopkins, A., 1985. *An investigation into the organizing and organizational features of published conference papers*. MA. University of Birmingham.

Hopkins, A. & Dudley-Evans, T., 1988. A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes,* vol. 7, no. 2, pp.113-121.

Hunston, S., 2000. Evaluation and the planes of discourse. In S. Hunston & G. Thompson, eds. 2000. *Evaluation in text*. Oxford: Oxford University Press, pp.176-207

Hunston, S., 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S., 2007. Semantic prosody revisited. *International Journal of Corpus Linguistics*, vol. 12, no. 2, pp.249-268.

Hunston, S. & Thompson, G., 2000. *Evaluation in text: authorial stance and the construction of discourse*. Oxford: Oxford University Press.

Hyland, K., 1998. *Hedging in scientific research articles*. Amsterdam: John Benjamins.

Hyland, K., 2002. *Teaching and researching writing*. London: Longman.

Hyland, K., 2005. Stance and engagement: a model of interaction in academic discourse. *Discourse Studies*, vol. 6, no. 2, pp.173-191.

Hyland, K., 2008. As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, vol. 27, no. 1, pp.4-21.

Hyland, K., 2009. *Academic Discourse*. London: Continuum.

Hyland, K., 2012. Genre and discourse analysis in language for specific purposes. In C. Chapelle, ed. 2013. *The encyclopedia of applied linguistics*. London: Blackwell Publishing, pp.2281-2288.

Hyland, K., 2014. Dialogue, persuasion and community in research writing, In M. Luz Gil-Salom & C. Soler-Monreal, eds. 2014. *Dialogicity in written specialised genres*, Amsterdam: John Benjamins, pp.1-20.

Hyland, K. & Guinda, C.S. eds., 2012. *Stance and voice in written academic genres*. New York: Palgrave Macmillan.

Hyland, K. & Tse, P., 2004. Metadiscourse in academic writing: A reappraisal. *Applied linguistics* 25, no. 2, pp.156-177.

Hymes, D., 2003. *Foundations of sociolinguistics: an ethnographic approach*. Philadelphia, PA: University of Pennsylvania Press.

Hyon, S., 1996. Genre in three traditions: implications for ESL. *TESOL Quarterly*, vol. 30, pp.693-722.

International Council of Medical Journal Editors (ICMJE), 2014a. *Protection of research participants*. Viewed 4 July 2013, http://www.icmje.org/recommendations/browse/roles-and-responsibilities/protection-of-research-participants.html.

International Council of Medical Journal Editors (ICMJE), 2014b. *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journal*. Viewed 8 December 2014, http://www.icmje.org/icmje-recommendations.pdf

Ioannidis, J., 2005. Why most published research findings are false. *PLoS Medicine.* Viewed 11 December 2013, http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0020124.

Jaffe, A., ed. 2009. *Stance: sociolinguistic perspectives*. Oxford: Oxford University Press.

Jirapanakorn, N., 2012. How doctors report: a corpus-based contrastive analysis of reporting verbs in research article introductions published in international and Thai medical journals. *The Bangkok Medical Journal*, vol. 4, pp.39-46.

Johnstone, B., 1987. Introduction: perspectives on repetition. *Text*, vol.7, no. 3, pp.205-214.

Jordan, J., 2005. Dell Hymes, Kenneth Burke's 'Identification,' and the birth of sociolinguistics. *Rhetoric Review*, vol. 24, no. 3, pp.264-279.

Jordan, M., 1984. *Rhetoric of everyday English texts*. London: George Allen and Unwin.

Junqueira, L. & Cortes, V., 2014. Metadiscourse in book reviews in English and Brazilian Portuguese: A corpus-based analysis. *Rhetoric, Professional Communication, and Globalization*, vol. 6, no. 1, pp.88-109.

Kanoksilapatham, B., 2005. Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, vol. 24, no. 3, pp.269-292.

Kay, P., 1995. Construction grammar. In J. Verschueren, J. Ostmann & J. Blommaert, eds. 1995. *Handbook of pragmatics*, Amsterdam: John Benjamins, pp.171-177.

Kecskes, I., 2014. *Intercultural pragmatics*. Oxford: Oxford University Press.

Kirkman, J., 1975. That pernicious passive voice. *Physics in Technology*, vol. 6, no. 5, pp.197-200.

Kleinert, S., Benham, D., Coolingridge, D., Summerskill, W. & Horton , R., 2014. Further emphasis on research in context. *The Lancet*, vol. 381, no. 9961, pp.2176-2177.

Knott, A. & Dale, R., 1994. Using linguistic phenomena to motivate coherence relations. *Discourse Processes*, vol. 18, no.1, pp.5-62.

Koester, A., 2010. Building small specialised corpora. In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. New York: Routledge, pp.66-79.

Labov, W., 1984. Intensity. In D. Schiffrin, ed. 1984. *Meaning, form, and use in context: linguistic applications*. Washington DC: Georgetown University Press, pp.43-70.

Laufer, B. & Nation, I., 1995. Lexical richness in L2 written production: can it be measured. *Applied Linguistics,* vol. 16, no. 3, pp.307-322.

Lee, D. & Swales, J., 2006. A corpus-based EAP course for NNS doctoral students: moving from available specialised corpora to self-compiled corpora. *English for Specific Purposes*, vol. 25, no. 1, pp.56-75.

Li, L. & Ge, G. 2009. Genre analysis: structural and linguistic evolution of the English-medium medical research article (1985–2004). *English for Specific Purposes*, vol. 28, no. 2, pp.93-104.

Lim, J., 2006. Method sections of management research articles: a pedagogically motivated qualitative study. *English for Specific Purposes*, vol. 25, no. 3, pp.282-309.

Lindeberg, A., 1994. Rhetorical conventions in the discussion/conclusion sections of research articles in finance, management and marketing. In M. Brekke, O. Anderson, T. Dahl & J. Myking, eds. 1994. *Applications and implications of current LSP research: proceedings of the 9th European LSP Symposium*, Bergen, Norway, Fagbokforlaget, pp.647-656.

Louw, W., 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis & E. T Bonelli, eds. 1993. *Text and technology: in honour of John Sinclair*, Amsterdam: John Benjamins, pp.157-176.

Maher, J., 1986. English for medical purposes. *Language Teaching*, vol.19, no. 2, pp.112-145.

Mann, W., Matthiessen, C. & Thompson, S., 1989. *Rhetorical structure and text analysis*. Marina del Rey, CA: Technical Report, Information Sciences Institute. Viewed 11 March 2013, https://www.researchgate.net/publication/259272404_Mann_William_C._Christian_M.I.M._Matthiessen__Sandra_A._Thompson._1992._Rhetorical_Structure_Theory_and_Text_Analysis._USCISI_Report._Also_in_William_C._Mann__Sandra_A._Thompson_%28eds.%29_Discourse_Description_Diverse_Linguistic_Analyses_of_a_Fund_Raising_Text._Amsterdam_Benjamins._39-78.

Mann, W. & Thompson, S., 1988. Rhetorical structure theory: towards a functional theory of text organization. *Text*, vol. 8, no. 3, pp.243-281.

Marco, M., 2000. Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes*, vol. 19, no. 1, pp.63-86.

Marius, R., 1990. A review of *Genre analysis: English in academic and research settings* by John M. Swales.

Martin, J., 2000. Beyond exchange: appraisal systems in English. In S. Hunston & G. Thompson, eds. 2000. *Evaluation in text*. Oxford: Oxford University Press, pp.142-175.

Martin, J. & White, P., 2005. *The language of evaluation: appraisal in English*. Basingstoke: Palgrave.

McCarthy, M. & O'Keeffe, A., 2010. Historical perspective: what are corpora and how have they evolved. In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. New York: Routledge, pp.3-13.

McEnery, T. & Wilson, A., 2010. *Corpus linguistics.* 2nd ed. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R. & Tono, Y., 2006. *Corpus-based language studies*. London: Routledge.

Meyer, C. ed., 2002. *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.

Milagros del Saz Rubio, M., 2011. A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes*, vol. 30, no. 4, pp.258-271.

Millar, N., Budgell, B. & Fuller, K., 2013. Use the active voice whenever possible: the impact of style guidelines in medical journals. *Applied Linguistics*, vol. 34, no. 4, pp.393-414.

Moon, R., 1998. *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Clarendon Press.

Moon, R., 2010. What can a corpus tell us about lexis? In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. London: Routledge, pp.197-211.

Moore, N., 2003. Review of 'Small corpus studies and ELT: theory and practice' by M. Ghadessy, A. Henry and R. Roseberry (eds). *System*, vol. 31, pp.299-308

Morley, J. & Partington, A., 2009. A few Frequently Asked Questions about semantic or evaluative prosody. *International Journal of Corpus Linguistics*, vol. 14, no. 2, pp.139-158.

Morris, L. & Cobb, T., 2004. Vocabulary profiles as predictors of the academic performance of Teaching English as a Second Language trainees. *System*, vol. 32, no. 1, pp.75-87.

Mosby, Inc., 2012. *Mosby's dictionary of medicine, nursing, and health professions*. 9th ed. St. Louis, MO: Elsevier.

Myers, G., 1991. Lexical cohesion and specialised knowledge in science and popular science texts. *Discourse Processes*, vol. 14, no. 1, pp.1-26.

Ng, K. & Peh, W., 2008. Writing the results. *Singapore Medical Journal*, vol. 49 pp.967-969.

Nwogu, K., 1997. The medical research paper: structure and functions. *English for Specific Purposes*, vol. 16, no. 2, pp.119-138.

O'Connor, M., Cooter, M., & Ufnalska, S.B., 2013. 1.1: Copy-editing scientific papers. In P. Smart, H. Maisonneuve & A. Polderman, eds. 2013 *European association of science editors' handbook*. Viewed 21 January 2014, http://www.ease.org.uk/publications/science-editors-handbook

O'Keeffe, A., McCarthy, M.J. & Carter, R.A., 2007. *From corpus to classroom: language use and language teaching*. Cambridge University Press, Cambridge.

Oakey, D., 2002. Formulaic language in English academic writing. In R. Reppen, S. Fitzmaurice & D. Biber, eds. *Using corpora to explore linguistic variation*, Philadelphia: John Benjamins, pp.111-129.

Ochs, E., 1989. Introduction. *Text-Interdisciplinary Journal for the Study of Discourse*, vol. 9, pp.1-6.

Östman, J.O., 1995. Pragmatic particles twenty years after. In B. Wårvik, S. Tanskanen & R. Hiltunen, eds. 1995. *Proceedings from the Turku conference on organization in discourse, Anglicana Turkuensia*, vol. 14, pp.95-108.

Ozturk, I., 2007. The textual organization of research article introductions in applied linguistics: variability within a single discipline. *English for Specific Purposes*, vol.26, no. 1, pp.25-38.

Paltridge, B., 1994. Genre analysis and the identification of textual boundaries. *Applied Linguistics*, vol. 15, no. 3, pp.288-299.

Peacock, M., 2002. Communicative moves in the discussion section of research articles. *System*, vol. 30, no. 4, pp.479-97.

Pearson, J., 1998. *Terms in context*. John Benjamins, Amsterdam.

Polanyi, L. & Scha, R., 1983. The syntax of discourse. *Text: Interdisciplinary Journal for the Study of Discourse*, vol. 3, no. 3, pp.261-270.

Poncini, G., 2004. *Discursive strategies in multicultural business meetings*. Bern: Peter Lang.

Posteguillo, S., 1999. The schematic structure of computer science research articles. *English for Specific Purposes*, vol.18, no. 2, pp.139-160.

Pridham, F., 2001. *The language of conversation*. London: Routledge.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J., 1985. *A comprehensive grammar of the English language*. London: Longman.

Redeker, G., 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics,* vol. 14, no. 3, pp.367-81.

Redeker, G., 1991. Linguistic markers of discourse structure. *Linguistics*, vol. 29, no. 6, pp.1139-1172.

Reid, J., 1996. U.S. academic readers, ESL writers, and second sentences. *Journal of Second Language Writing*, vol. 5, no. 2, pp.129-161.

Reppen, R., 2010. Building a corpus: what are the key considerations. In A. O'Keeffe & M. McCarthy, eds. 2010. *The Routledge handbook of corpus linguistics*. New York: Routledge, pp.31-37.

Röhrig, B., du Prel J., & Blettner, M., 2009. Study design in medical research. *Deutsches Ärzteblatt International*, vol. 106, pp.184-189.

Röhrig, B., du Prel J., Wachtlin, D., & Blettner, M., 2009. Types of study in medical research. *Deutsches Ärzteblatt International*, vol. 106, pp.262-268.

Rouchota, V., 1996. Discourse connectives: what do they link? In J. Harris & P. Backley, eds. 1996. *UCL working papers in linguistics*, vol. 8, pp.199-212.

Salager-Meyer, F., 1994. Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, vol. 13, no. 2, pp.149-170.

Salkie, R., 1995. *Text and discourse analysis*. London: Routledge.

Samraj, B., 2002. Introductions in research articles: variations across disciplines. *English for Specific Purposes*, vol. 21, no. 1, pp.1-17.

Schiffrin, D., 1987. *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D., 2001. Discourse markers: language, meaning, and context. In D. Schiffrin, D. Tannen & H. Hamilton, eds. 2001. *The handbook of discourse analysis*. Oxford: Blackwell, pp.54-75.

Schiffrin, D., 2006. Discourse marker research and theory: revisiting *and*. In K. Fischer. ed. 2006. *Approaches to discourse particles*. Amsterdam: Elsevier, pp.315-338.

Schourup, L., 1985. *Common discourse particles in English conversation: like, well, y'know*. New York: Garland.

Schourup, L., 2011. The discourse marker *now*: a relevance-theoretic approach. *Journal of Pragmatics*, vol. 43, no. 8, pp.2110-2129.

Scott, M., 2008. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

Sheen, A.P., 1982. *Breathing life into medical writing: a handbook.* St. Louis, MO: Mosby.

Sheldon, E., 2013. *The research article: a rhetorical and functional comparison of the native and non-native English writers and native Spanish writers.* PhD. University of New South Wales. Viewed 21 January 2014, http://www.isfla.org/Systemics/Print/Theses/Sheldon2013.pdf.

Šimčikaitė, A., 2012. Spoken discourse markers in learner academic writing. *Studies About Languages*, vol. 20, pp.27-34.

Simpson, P., 2003. *Language, ideology and point of view.* London: Routledge.

Sinclair, J., 1985. On the integration of linguistic description. In T.A. van Dijk, ed. 1985. *Handbook of discourse analysis*, vol. 2, London: Academic Press, pp.13-28.

Sinclair, J., 1987. *Looking up*. HarperCollins, London.

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

Sinclair, J., 1996. Corpus to corpus: a study of translation equivalence. *International Journal of Lexicography*, vol. 9, pp.179-196.

Sinclair, J., 1998. The lexical item. In E. Weigand, ed. 1998. *Contrastive lexical semantics*. Amsterdam: John Benjamins, pp.1-24.

Sinclair, J., 2004a. *How to use corpora for language teaching*. Amsterdam: John Benjamins.

Sinclair, J., 2004b. *Trust the text: language, corpus and discourse*. London: Routledge.

Skelton, J., 1994. Analysis of the structure of original research papers: an aid to writing original papers for publication. *British Journal of General Practice*, vol. 44, no. 387, pp.455-459.

Small, H., 2010. Referencing through history: how the analysis of landmark scholarly texts can inform citation theory. *Research Evaluation*, vol. 19, no. 3, pp.185-193.

Smith, N., and Hardie, A., 2009. Keyword analysis. Viewed 11 December 2010, http://www.lancaster.ac.uk/fss/courses/ling/corpus/blue/l03_2.htm.

Sperber, D. & Wilson, D., 1986. *Relevance: communication and cognition*. Oxford: Blackwell.

Stephens, M., 1999. *Proficiency writing*. Essex: Pearson Education Limited.

Stewart, D., 2010. *Semantic prosody: a critical evaluation*. London: Routledge.

Stubbs, M., 1983. *Discourse analysis: the sociolinguistic analysis of natural language*. Chicago: University of Chicago Press.

Stubbs, M.,1995. Corpus evidence for norms of lexical collocation. In G. Cook and B. Seidlhofer, eds. 1995. *Principles and practice in applied linguistics: studies in honour of H G Widdowson*. Oxford: Oxford University Press, pp.243-256.

Stubbs, M., 2001. *Words and phrases: corpus studies of lexical semantics*, Oxford: Blackwell.

Stubbs, M., 2008. 4 language corpora. In A. Davies & C. Elder, eds. 2008. *The handbook of applied linguistics*. New York: Wiley, pp.106-132.

Swales, J., 1981. *Aspects of article introductions*. Language Studies Unit, University of Aston, Birmingham.

Swales, J., 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Swales, J., 2004. *Research genres: explorations and applications*. Cambridge: Cambridge University Press.

Swales, J., 2012. *From context to text: investigating structures, functions and forms in research articles*. Slide presentation for CNRS Lattice, March 2012, Paris. Viewed 21 March 2013, http://www.lattice.cnrs.fr/Conference-John-Swales-From?lang=fr.

Swales, J., 2014. Variation in citational practice in a corpus of student biology papers from parenthetical plonking to intertextual storytelling. *Written Communication*, vol. 31, no. 1, pp.118-141.

Swales, J. & Feak, C., 2004. *Academic writing for graduate students*. 2nd ed. Ann Arbor, MI: University of Michigan Press.

Taboada, M., 2006. Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, vol. 38, no. 4, pp.567-592.

Taboada, M. & de los Ángeles Gómez-González, M., 2012. Discourse markers and coherence relations: Comparison across markers, languages and modalities. *Linguistics and the Human Sciences*, vol. 6, no. 1-3, pp.17-41.

Taboada, M. & Mann, W., 2006. Applications of rhetorical structure theory. *Discourse Studies*, vol. 8, no. 4, pp.567-588.

Tadros, A., 1985. *Prediction in text,* (No. 10.). English Language Research, University of Birmingham Discourse Monographs.

Tadros, A., 1993. The pragmatics of text averral and attribution in academic texts. In Hoey, M., ed. 1993. *Data, description, discourse: papers on the English language in honour of John McH Sinclair on his sixtieth birthday*, London: Harper Collins, pp.98-114.

Teufel, S., 1998. Meta-discourse markers and problem-structuring in scientific articles. In *Proceedings of the Workshop on Discourse Relations and Discourse Markers at the 17th International Conference on Computational Linguistics*, pp.43-49. Viewed 3 March 2014, http://www.cl.cam.ac.uk/~sht25/papers/discourse.pdf.

Teufel, S. & Moens, M., 2002. Summarizing scientific articles: experiments with relevance and rhetorical status. *Computational Linguistics*, vol. 28, no. 4, pp.409-445.

Thomas, S. & Hawes, T., 1994. Reporting verbs in medical journal articles. *English for Specific Purposes*, vol.13, no. 2, pp.129-148.

Thompson, D., 1993. Arguing for experimental facts in science: a study of research article results sections in biochemistry. *Written Communication*, vol.10, no. 1, pp.106-128.

Thompson, G. & Ye, Y., 1991. Evaluation in the reporting verbs used in academic papers. *Applied Linguistics*, vol. 12, no. 4, pp.365-382.

Thompson, P., 2000. Citation practices in PhD theses. In L. Burnard & T. McEnery, eds. 2000. *Rethinking language pedagogy from a corpus perspective*, Berlin: Peter Lang, pp.91-101.

Thompson, P., 2002. Manifesting intertextuality in the PhD thesis. *Revista Canaria de Estudios Ingleses*, vol. 44, pp.97-114.

Thompson, P., 2005. Aspects of identification and position in intertextual reference in PhD theses. In E. Tognini-Bonelli & G. Camiciotti, eds. 2005. *Strategies in academic discourse*, Philadelphia: John Benjamins, pp.31-50.

Thompson, P. & Tribble, C., 2001. Looking at citations: using corpora in English for academic purposes. *Language Learning and Technology*, vol. 5, no. 3, pp.91-105.

Todorov, T., 1976. The origin of genres. *New Literary History*, vol. 8, pp.159-170.

Tribble, C., 2000. Genres, keywords, teaching: towards a pedagogic account of the language of project proposals. In L. Burnard & T. McEnery, eds. 2000. *Rethinking language pedagogy from a corpus perspective,* Frankfurt: Peter Lang, pp.75-90.

Tribble, C., 2002. Corpora and corpus analysis: new windows on academic writing. In J. Flowerdew, ed. 2013, *Academic discourse*, New York: Routledge, pp.131-149.

Tribble, C., 2012. Teaching and language corpora: Quo Vadis. In *10th Teaching and Language Corpora Conference (TALC)*, Warsaw, 11-14 July 2012. Viewed 3 May 2013, http://talc10.ils.uw.edu.pl/files/2011/09/Tribble.pdf.

Tseronis, A., 2009. *Qualifying standpoints: stance adverbs as a presentational device for managing the burden of proof.* Utrecht: Netherlands Graduate School of Linguistics.

Tufte, E., 2001. *The visual display of quantitative information.* 2nd ed. New York: Graphics Press.

Upton, T., 2002. Understanding direct mail letters as a genre. *International Journal of Corpus Linguistic*s, vol. 7, no. 1, pp.65-85.

Upton, T. & Cohen, M., 2009. An approach to corpus-based discourse analysis: the move analysis as example. *Discourse Studies,* vol. 11, no. 5, pp.585-605.

Upton, T. & Connor, U., 2001. Using computerized corpus analysis to investigate the text linguistic discourse moves of a genre. *English for Specific Purposes*, vol. 20, no. 4, pp.313-329.

van Dijk, T., 1979. Pragmatic connectives. *Journal of Pragmatics*, vol. 3, no. 5, pp.447-456.

van Dijk, T., 1980. *Macrostructures*. New Jersey: Lawrence Erlbaum Associates, Inc.

van Langenhove, L. & Harré, R., 1999. Introducing positioning theory. In R. Harré & L. van Langenhove, eds. 1999. *Positioning theory: moral contexts of intentional action*, Oxford: Blackwell, pp.14-31.

Vlăduțescu, Ș., 2014. Twelve communicational principles of propaganda. *International Letters of Social and Humanistic Sciences*, vol. 23, pp.71-80.

Vološinov, V.N., 1973. *Marxism and the philosophy of language.* Translated from Russian by L. Matejka & I. R. Titunik, Amherst: Massachusetts University Press, Amherst.

Wang, J., Liang, S. & Ge, G., 2008. Establishment of a medical academic word list. *English for Specific Purposes*, vol. 27, no. 4, pp.442-458.

Webber, B., 2004. D-LTAG: Extending lexicalized TAG to Discourse. *Cognitive Science*, vol. 28, no. 5, pp.751-779.

White, H., 2004. Citation analysis and discourse analysis revisited. *Applied Linguistics*, vol. 25, no. 1, pp.89-116.

White, P., 2003. Beyond modality and hedging: a dialogic view of the language of intersubjective stance. *TEXT*, vol. 23, no. 2, pp.259-284.

Whitsitt, S., 2005. A critique of the concept of semantic prosody. *International Journal of Corpus Linguistics*, vol. 10, no. 3, pp.283-305.

Widdowson, H., 2000. On the limitations of linguistics applied. *Applied Linguistics*, vol. 21, no. 1, pp.3-25.

Williams, H., 2004. Lexical frames and reported speech. *ELT Journal*, vol. 58, no. 3, pp.247-257.

Williams, I., 1996. A contextual study of lexical verbs in two types of medical research report: clinical and experimental. *English for Specific Purposes*, vol. 15, no. 3, pp.175-197.

Wilson, D. & Sperber, D., 2002. Relevance theory. In L. Horn & G. Ward, eds. 2008. *Handbook of pragmatics*, Oxford: Blackwell, pp.607-632.

World Medical Association. 1964. *Helsinki Guidelines (Ethical principles for medical research involving human subjects)*. Viewed 4 July 2013, http://www.wma.net/en/30publications/10policies/b3/.

Wray, A., 2000. Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics*, vol. 21, no. 4, pp.463-489.

Wray, A., 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. & Perkins, M., 2000. The functions of formulaic language: an integrated model. *Language & Communication*, vol. 20, no. 1, pp.1-28.

Writing Center at UNC 2014, *Passive voice*. Viewed 5 May 2014, http://writingcenter.unc.edu/handouts/passive-voie/.

Wyse, B., 2009. *Factive/non-factive predicate recognition within question generation systems*. London: Open University.

Yang, R. & Allison, D., 2004a. Research articles in applied linguistics: moving from results to conclusions. *English for Specific Purposes*, vol. 22, no. 4, pp.365-385.

Yang, R. & Allison, D., 2004b. Research articles in applied linguistics: structures from a functional perspective. *English for Specific Purposes*, vol. 23, no. 3, pp.264-279.

Yang, S., 2012. Discourse markers? An area of confusion. In F. Rodriguez Muñoz, ed. *Los marcadores del discurso en español y en chino mandarín*, Universidad de

Almería, Almería, pp.37-44. Viewed 4 April 2014, http://www.ual.es/revistas/PhilUr/pdf/PhilUr07.4.YangShanru.pdf.

Yates, J. & Orlikowski, W., 1992. Genres of organizational communication: A structurational approach to studying communication and media. *Academy of Management Review*, vol. 17, no. 2, pp.299-326.

Yu, N., Hassan, K., Long, X. & Wang, X., 2013. Young doctors and the pressure of publication. *The Lancet*, vol. 381, 2 February 2013. Viewed 5 August 2014, http://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(13)60933-2.pdf.

Yuan, H., Xu, W. & Hu, H. 2013. Young Chinese doctors and the pressure of publication. *The Lancet*, vol. 381, 2 February 2013. Viewed 5 August 2014, http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)60174-9/fulltext

Appendix 1: Articles in the medical research article corpus

2001   Effler, P. et al., Factors contributing to the emergence of Escherichia coli O157 in Africa *Emerging Infectious Diseases*

2002   Germinario, R. et al., Different forms of Vanadate on sugar transport in insulin target and nontarget cells, *Journal of Biomedicine and Biotechnology*

2002   Kitchener, B. et al,. Mental health first aid training for the public: evaluation of effects on knowledge, attitudes and helping behavior, *BMC Psychiatry*

2002   Samel, S. et al., Microscopy of bacterial translocation during small bowel obstruction and ischemia in vivo – a new animal model, *BMC Surgery*

2002   Sharif, W. et al., Distinct functions of S. pombe Rec12 (Spo11) protein and Rec12-dependent crossover recombination (chiasmata) in meiosis I; and a requirement for Rec12 in meiosis II, *Cell & Chromosome*

2002   Takasu, J. et al., Aged garlic extract therapy for sickle cell anemia patients, *BMC Blood Disorders*

2002   Yamboliev, I. et al., Canine mesenteric artery and vein convey no difference in the content of major contractile proteins, *BMC Physiology*

2003   Bailie, R. et al., Preventive medical care in remote Aboriginal communities in the Northern Territory: a follow-up study of the impact of clinical guidelines, computerised recall and reminder systems, and audit and feedback, *BMC Health Services Research*

2003   Franklin, J. et al., Revision and complication rates in 654 Exeter total hip replacements, with a maximum follow-up of 20 years, *BMC Musculoskeletal Disorders*

2003   Olsson, K. et al., Change of diurnal heart rate patterns during pregnancy and lactation in dogs (Canis familiaris), *Acta Veterinaria Scandinavica*

2003   Palmer, R. et al., Analysis of circulating insulin-like growth factor-1 (IGF-1) and IGF binding protein-3 (IGFBP-3) in tobacco smokers and non-smokers, *Tobacco Induced Diseases*

2003   Reis, B. et al., Time series modeling for syndromic surveillance, *BMC Medical Informatics and Decision Making*

2003   Shidham, V. et al., Colon biopsies for evaluation of acute graft-versus-host disease (A-GVHD) in allogeneic bone marrow transplant patients, *BMC Gastroenterology*

2004   Budak, Y. et al., Erytrocyte membrane anionic charge in type 2 diabetic patients with retinopathy, *BMC Opthamology*

2004   Fay, J. et al., Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae, *Genome Biology*

2004   Joseph, J. et al., Serum interleukin-5 levels are elevated in mild and moderate persistent asthma irrespective of regular inhaled glucocorticoid therapy, *BMC Pulmonary Medicine*

2004   Lempp, H. et al., A qualitative study of the perceptions and experiences of Pre-Registration House Officers on teamwork and support, *BMC Medical Education*

2004   Nishtar, S. et al., Prevention of non-communicable diseases in Pakistan: an integrated partnership-based model, *Health Research Policy and Systems*

2004   Tong, W. et al., Assessment of prediction confidence and domain extrapolation of two structure–activity relationship models for predicting estrogen receptor binding activity, *Environmental Health Perspectives*

2005   Bhardwaj, N. et al., Kernel-based machine learning protocol for predicting DNA-binding proteins, *Nucleic Acids Research*

2005   Brameid, L. et al., The use of end-quintile comparisons to identify under-servicing of the poor and over-servicing of the rich: A longitudinal study describing the effect of socioeconomic status on healthcare, *BMC Health Services Research*

2005   Henneken, M. et al., Differential expression of chemokine receptors on peripheral blood B cells from patients with rheumatoid arthritis and systemic lupus erythematosus, *Arthritis Research & Therapy*

2005   Janno, S. et al., Validity of Simpson-Angus Scale (SAS) in a naturalistic schizophrenia population, *BMC Neurology*

2005   Lee, R. et al., Adherence to oral hypoglycemic agents in Hawaii, *Preventing Chronic Disease*

2005   Mhawech, P. et al., 14-3-3 proteins—an update, *Cell Research*

2005   Sverdlov, A. et al., Conservation versus parallel gains in intron evolution, *Nucleic Acids Research*

2005   Wang, X. et al., Predominant constitutive CFTR conductance in small airways, *Respiratory Research*

2005   Wang, Z. et al., Cryptic promoter activity in the DNA sequence corresponding to the pim-1 5-UTR, *Nucleic Acids Research*

2005   Wong, K. et al., Yersinia pseudotuberculosis spatially controls activation and misregulation of host cell Rac1, *PLoS Pathogens*

2006   Clark, S. et al., Attitudes of US obstetricians toward a combined tetanus-diphtheria-acellular pertussis vaccine for adults, *Preventing Chronic Disease*

2006   Dauer, L. et al., Evaluating the effectiveness of a radiation safety training intervention for oncology nurses: a pretest – intervention – posttest study, *BMC Medical Education*

2006   Dillon, M. et al., The role of major duct excision and microdochectomy in the detection of breast carcinoma, *BMC Cancer*

2006   Hannon, C. et al., Play across Boston: a community initiative to reduce disparities in access to after-school physical activity programs for inner-city youths, *Preventing Chronic Disease*

2006   Homanics, G. et al., Production and characterization of murine models of classic and intermediate maple syrup urine disease, *BMC Medical Genetics*

2006   Kim, D. et al., Clinical significance of platelet count at day +60 after allogeneic peripheral blood stem cell transplantation, *Journal of Korean Medical Science*

2006   Lin, H. et al., Cyclin D1 and p16 expression in recurrent nasopharyngeal carcinoma, *World Journal of Surgical Oncology*

2006   Russell, C. et al., Spatial control of rabies on heterogeneous landscapes, *PLoS One*

2006   Shulman, J. et al., Neurovirulent vaccine-derived polio viruses in sewage from highly immune populations, *PLoS One*

2006   Uddin, R. et al., Breakpoint Associated with a novel 2.3 Mb deletion in the VCFS region of 22q11 and the role of Alu (SINE) in recurring microdeletions, *BMC Medical Genetics*

2006   Vordermark, D. et al., Plasma osteopontin levels in patients with head and neck cancer and cervix cancer are critically dependent on the choice of ELISA system, *BMC Cancer*

2006   Wahle, M. et al., Failure of catecholamines to shift T-cell cytokine responses toward a Th2 profile in patients with rheumatoid arthritis, *Arthritis Research & Therapy*

2007   Abdurahman, S. et al., Characterization of the invariable residue 51 mutations of human immunodeficiency virus type 1 capsid protein on in vitro CA assembly and infectivity, *Retrovirology*

2007   Baïsse, B. et al., Evolutionary conservation of P-selectin glycoprotein ligand-1 primary structure and function, *BMC Evolutionary Biology*

2007   Blau , V. et al., The face-specific N170 component is modulated by emotional facial expression, *Behavioral and Brain Functions*

2007   Corthorn, J. et al., Spatio-temporal expression of MMP-2, MMP-9 and tissue kallikrein in uteroplacental units of the pregnant guinea-pig (Cavia porcellus), *Reproductive Biology and Endocrinology*

2007   Curtis, D. Comparison of artificial neural network analysis with other multimarker methods for detecting genetic association, *BMC Genetics*

2007   de Torres, J. et al., COPD heterogeneity: gender differences in the multidimensional BODE index, *International Journal of Chronic Obstructive Pulmonary Disease*

2007   Gaston, K. et al., C to U editing at position 32 of the anticodon loop precedes tRNA 5_leader removal in trypanosomatids, *Nucleic Acids Research*

2007   Hamard, P. et al., Sumoylation delays the ATF7 transcription factor subcellular localization and inhibits its transcriptional activity, *Nucleic Acids Research*

2007   Hongzhej, L. et al., SirT1 modulates the estrogen–insulin-like growth factor-1 signaling for postnatal development of mammary gland in mice, *Breast Cancer Research*

2007   Kane, T. et al., Thoracoscopic-assisted esophagectomy and laparoscopic gastric pull-up for lye injury, *Journal of the Society of Laparoendoscopic Surgeons*

2007   Kaper, T. et al., Nanosensor detection of an immunoregulatory tryptophan influx/kynurenine efflux cycle, *PLoS Biology*

2007   Kawano, H. et al., A new therapy for highly effective tumor eradication using HVJ-E combined with chemotherapy, *BMC Medicine*

2007   Makareeva, E. et al., Procollagen triple helix assembly: an unconventional chaperone-assisted folding paradigm, *PLoS One*

2007   Nakamura, R. et al., Identification of two novel activities of the Wnt signaling regulator Dickkopf 3 and characterization of its expression in the mouse retina, *BMC Cell Biology*

2007   Nisabagasani, X. et al., Home-based management of fever in rural Uganda: community perceptions and provider opinions, *Malaria Journal*

2007   O'Malley, D. et al., Leptin promotes rapid dynamic changes in hippocampal dendritic morphology, *Molecular and Cellular Neurosciences*

2007   Ono, K. et al., Autologous transplantation of conjunctival epithelial cells cultured on amniotic membrane in a rabbit model, *Molecular Vision*

2007   Oşan, R. et al., Subspace projection approaches to classification and visualization of neural network-level encoding patterns, *PLoS One*

2007   Ping, A. et al., Regulatory polymorphisms in the Cyclophilin A gene, PPIA, accelerate progression to AIDS, *PLoS Pathogens*

2007   Rosell, R. et al., BRCA1: a novel prognostic factor in resected non-small-cell lung cancer, *PLoS One*

2007   Sasaki, M. et al., A novel strategy in production of oligosaccharides in digestive tract: prevention of postprandial hyperglycemia and hyperinsulinemia, *Journal of Clinical Biochemistry and Nutrition*

2007   Schulz, T. Variable expression of cre recombinase transgenes precludes reliable prediction of tissue-specific gene disruption by tail-biopsyg genotyping, *PLoS One*

2007   Tahwinder, U. et al., Recent technological developments: in situ histopathological interrogation of surgical tissues and resection margins, *Head & Face Medicine*

2007   Tenaillon, O. et al., Quantifying organismal complexity using a population genetic approach, *PLoS One*

2007   Thaipadungpanit, J. et al., A dominant clone of leptospira interrogans associated with an outbreak of human leptospirosis in Thailand, *PLoS Neglected Tropical Diseases*

2007   Weaver, D. et al., Computational and transcriptional evidence for microRNAs in the honey bee genome, *Genome Biology*

2007   Yang, M. et al., Social approach behaviors are similar on conventional versus reverse lighting cycles, and in replications across cohorts, in BTBR T+ tf/J, C57BL/6J, and vasopressin receptor 1B mutant mice, *Frontiers in Behavioral Neuroscience*

2008   Alten, R. et al., The human anti-IL-1_ monoclonal antibody ACZ885 is effective in joint inflammation models in mice and in a proof-of-concept study in patients with rheumatoid arthritis, *Arthritis Research & Therapy*

2008   Bonzheim, I. Hamano K et al., Analysis of single nucleotide polymorphisms in the FAS and CTLA-4 genes of peripheral T-cell lymphomas, *Journal of Hemopathology*

2008   Cauchi, S. et al., Post genome-wide association studies of novel genes associated with type 2 diabetes show gene-gene interaction and high predictive value, *PLoS One*

2008   Dahmen, N. et al., Elevated peripheral visfatin levels in narcoleptic patients, *PLoS One*

2008   Darwish, M. et al., Dose proportionality of Fentanyl buccal tablet in healthy Japanese volunteers, *Archives of Drug Information*

2008   Denny, S. et al., Hand-held Internet tablets for school-based data collection, *BMC Research Notes*

2008   Farba, K. et al., Release assay to detect latent tuberculosis infection is retained in HIV-infected patients but dependent on HIV/AIDS progression, *PLoS One*

2008   Ferreira, M. et al., Degradation of 4-fluorophenol by arthrobacter sp. strain IF1, *Applied and Environmental Microbiology*

2008   Hamano, K. et al. Associations of renal vascular resistance with albuminuria and other macroangiopathy in Type 2 diabetic patients, *Diabetes Care*

2008   Jones, D. et al., Increasing physical activity and decreasing sedentary activity in adolescent girls –the Incorporating More Physical Activity and Calcium in Teens (IMPACT) study, *The International Journal of Behavioral Nutrition and Physical Activity*

2008   Kulkarni, H. et al., CCL3L1-CCR5 genotype improves the assessment of AIDS risk in HIV-1-infected individuals, *PLoS One*

2008   Larsen, K. et al., Bicaudal D actively regulates microtubule motor activity in lipid droplet transport, *PLoS One*

2008   Lee, E. et al., Splice variants of the forkhead box protein AFX exhibit dominant negative activity and inhibit AFXa-mediated tumor Cell apoptosis, *PLoS One*

2008   Lema, S. et al., Dietary exposure to 2,2,4,4-Tetrabromodiphenyl ether (PBDE-47) alters thyroid status and thyroid hormone–regulated gene transcription in the pituitary and brain, *Environmental Health Perspectives*

2008   Lesterlin, C. et al., Asymmetry of chromosome replichores renders the DNA translocase activity of ftsK essential for cell division and cell shape maintenance in Escherichia coli, *PLoS Genetics*

2008   Lindemose, S. et al., Dissecting direct and indirect readout of cAMP receptor protein DNA binding using an inosine and 2,6-diaminopurine in vitro selection system, *Nucleic Acids Research*

2008   Llewellyn, R. et al., Endometrial expression of the insulin-like growth factor system during uterine involution in the postpartum dairy cow, *Domestic Animal Endocrinology*

2008   Lope, V. et al., Municipal distribution of ovarian cancer mortality in Spain, *BMC Cancer*

2008   Ma, X. et al., Mapping genetically compensatory pathways from synthetic lethal interactions in yeast, *PLoS One*

2008   Maruyama, R. et al., Cytoplasmic RASSF2A is a proapoptotic mediator whose expression is epigenetically silenced in gastric cancer, *Carcinogenesis*

2008   McFarlane-Anderson, N. et al., Cervical dysplasia and cancer and the use of hormonal contraceptives in Jamaican women, *BMC Women's Health*

2008   Meeren, H. et al., Early category-specific cortical activation revealed by visual stimulus inversion, *PLoS One*

2008   Menendez, C. et al., A randomized placebo-controlled trial of intermittent preventive treatment in pregnant women in the context of insecticide treated nets delivered through the antenatal clinic, *PLoS One*

2008   Moxon, S. et al., A scoring matrix approach to detecting miRNA target sites, *Algorithms for molecular biology*

2008   Overbeek, L. et al., Most patients with colorectal tumors at young age do not visit a cancer genetics clinic, *Diseases of the Colon and Rectum*

2008   Press, J. et al., Ovarian carcinomas with genetic and epigenetic BRCA1 loss have distinct molecular abnormalities, *BMC Cancer*

2008   Raphael, B. et al., A sequence-based survey of the complex structural organization of tumor genomes, *Genome Biology*

2008   Sandgren, E. et al., HIV/AIDS awareness and risk behaviour among pregnant women in Semey, Kazakhstan, 2007, *BMC Public Health*

2008   Schoeman, C. et al., The relative influence of competition and prey defenses on the phenotypic structure of insectivorous bat ensembles in Southern Africa, *PLoS One*

2008   Shakhiev, R. et al., Patterns signaling via toll-like receptors 2 and 5 contribute to epithelial repair, growth and survival TLR and epithelial homeostasis, *PLoS One*

2008   Singh, A. et al., An information-theoretic analysis of genetics, gender and age in cancer patients, *PLoS One*

2008   Suzuki, T. et al., Identification of kaonashi mutants showing abnormal pollen exine structure in Arabidopsis thaliana, *Plant and Cell Physiology*

2008   Tarendeau, F. et al., Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit, *PLoS Pathogens*

2008   Tarres, J. et al., Data transformation for rank reduction in multi-trait MACE model for international bull comparison, *Genetics, Selection, Evolution*

2008   Teske, P. et al., Oceanic dispersal barriers, adaptation and larval retention: an interdisciplinary assessment of potential factors maintaining a phylogeographic break between sister lineages of an African prawn, *BMC Evolutionary Biology*

2008   van der Meulen, M. et al., The Dutch version of the Oral Health Impact Profile (OHIP-NL): translation, reliability and construct validity, *BMC Oral Health*

2008   Velasco-Villa, A. et al., Enzootic rabies elimination from dogs and reemergence in wild terrestrial carnivores, United States, *Emerging Infectious Diseases*

2008   Vythilingam, I. et al., Plasmodium knowlesi in humans, macaques and mosquitoes in peninsular Malaysia, *Parasites and Vectors*

2008   Wang, R. et al., Health related quality of life measured by SF-36: a population-based study in Shanghai, China, *BMC Public Health*

2008   Weerasinghe, M. et al., Safe storage of pesticides in Sri Lanka – Identifying important design, *BMC Public Health*

2008   Wild, J. et al., Rescripting early memories linked to negative images in social phobia: a pilot, *Behavior Therapy*

2008   Zhang, W. et al., Term-tissue specific models for prediction of gene ontology biological processes using transcriptional profiles of aging in drosophila melanogaster, *BMC Bioinformatics*

2009   Amonsin, A. et al., Comparative analysis of complete nucleotide sequence of porcine reproductive and respiratory syndrome virus (PRRSV) isolates in Thailand (US and EU genotypes), *Virology Journal*

2009   Barutcuoglu, Z. et al., Aneuploidy prediction and tumor classification with heterogeneous hidden conditional random fields, *Bioinformatics*

2009   Barwell, N. et al., Individual responsiveness to exercise-induced fat loss is associated with change in resting substrate utilization, *Metabolism*

2009   Battacharya, K. et al., Titanium dioxide nanoparticles induce oxidative stress and DNA-adduct formation but not DNA-breakage in human lung cells, *Particle and Fibre Toxicology*

2009   Bauman, A. The International Prevalence Study on physical activity: results from 20 countries, *International Journal of Behavioral and Physical Activity*

2009   Bortolotti, G. et al., Physiological stress mediates the honesty of social signals, *PLoS One*

2009   Brown, R. et al., Sequence skill acquisition and off-line learning in normal aging skill acquisition, *PLoS One*

2009   Bystroem, P. et al., Serum vitamin B12 and folate status among patients with chemotherapy treatment for advanced colorectal cancer, *Upsala Journal of Medical Sciences*

2009   Cañete, J. et al., Distinct synovial immunopathology in Behçet disease and psoriatic arthritis, *Arthritis Research & Therapy*

2009   Chandriani, S. et al., A core MYC gene expression signature is prominent in basal-like breast cancer but only partially pverlaps the core serum response, *PLoS One*

2009   Charest, I. et al., Electrophysiological evidence for an early processing of human voices, *BMC Neuroscience*

2009   Chini, F. et al., Road traffic injuries in one local health unit in the Lazio region: results of a surveillance system integrating police and health data, *International Journal of Health Geographies*

2009   de Grassi, A. et al., Tandem repeats modify the structure of human genes hosted in segmental duplications, *Genome Biology*

2009   Dobbins, M. et al., A description of a knowledge broker role implemented as part of a randomized controlled trial evaluating three knowledge translation strategies *Implementation Science*

2009   Du, Y. et al., Thermochemotherapy effect of nanosized As2O3/Fe3O4 complex on experimental mouse tumors and its influence on the expression of CD44v6, VEGF-C and MMP-9, *BMC Biotechnology*

2009   Dussan, K. et al., Effects of a refugee elective on medical student perceptions, *BMC Medical Education*

2009   Eijinden, R. et al., Compulsive internet use among adolescents: bidirectional parent–child relationships, *Journal of Abnormal Child Psychology*

2009   Gioiosa, L. et al., A trouble shared is a trouble halved: social context and status affect pain in mouse dyads social modulation of pain, *PLoS One*

2009   Harrington C et al., Use of bioanalyzer electropherograms for quality control and target evaluation in microarray expression profiling studies of ocular tissues, *Journal of Ocular Biology, Diseases, and Informatics*

2009   Hensler, J. et al., What is the optimal value of the g-ratio for myelinated fibers in the rat CNS? A theoretical approaching-ratio optimization, *PLoS One*

2009   Holmes, E. et al., Can playing the computer game "Tetris" reduce the build-up of flashbacks for trauma? a proposal from cognitive science, *PLoS One*

2009   Hong, T. et al., Effects of retinoic acid isomers on apoptosis and enzymatic antioxidant system in human breast cancer cells, *Nutrition Research and Practice*

2009   Hughes, B. et al., Mclk1+/- mice are not resistant to the development of atherosclerosis, *Lipids in Health and Disease*

2009   Jensen, A. et al., Use of fertility drugs and risk of ovarian cancer: Danish population based cohort study, *BMJ (Clinical Research)*

2009   Jing, Y. et al., A novel UBIAD1 mutation identified in a Chinese family with Schnyder crystalline corneal dystrophy, *Journal of Abnormal Child Psychology*

2009   Juengel, E. et al., Combining the receptor tyrosine kinase inhibitor AEE788 and the mammalian target of rapamycin (mTOR) inhibitor RAD001 strongly inhibits adhesion and growth of renal cell carcinoma cells, *BMC Cancer*

2009   Kaplan-Marcusan, A. et al., Perception of primary health professionals about female genital mutilation: from healthcare to intercultural competence, *BMC Health Services Research*

2009   Koc, B. et al., Lower leptin levels in young non-obese male smokers than non-smokers, *Upsala Journal of Medical Sciences*

2009   Lanfear, D. et al., Short term effects of milrinone on biomarkers of necrosis, apoptosis, and inflammation in patients with severe heart failure, *Journal of Translational Medicine*

2009   Langton, T. et al., Fragments of the key flowering gene GIGANTEA are associated with helitron-type sequences in the pooideae grass Lolium perenne, *BMC Plant Biology*

2009   Lenoir, T. et al., Tracking the current rise of Chinese pharmaceutical bionanotechnology, *Journal of Biomedical Discovery and Collaboration*

2009   Li, X. et al., Macro role(s) of microRNAs in fragile X syndrome? *Neuromolecular Medicine*

2009   Lowery, A. et al., MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer, *Breast Cancer Research*

2009   Mahjoub, F. et al., Mast cell gastritis: children complaining of chronic abdominal pain with histologically normal gastric mucosal biopsies except for increase in mast cells, proposing a new entity, *Diagnostic Pathology*

2009   Marins, J. et al., Biopure MTAD induces DNA damage but not cellular death: an in vitro study, *European Journal of Dentistry*

2009   Maumus, F. et al., Potential impact of stress activated retrotransposons on genome evolution in a marine diatom, *BMC Genomics*

2009   McNeill, C., Colorectal cancer prevention for low-income, sociodemographically-diverse adults in public housing: baseline findings of a randomized controlled trial, *BMC Public Health*

2009   Mfinanga, G. et al., Health facility-based active management of the third stage of labor: findings from a national survey in Tanzania, *Health Research Policy and Systems*

2009   Mhlanga, M. et al., In Vivo colocalisation of oskar mRNA and trans-acting proteins revealed by quantitative imaging of the drosophila oocyte, *PLoS One*

2009   Mitra-Ganguli, T. et al., Orientation of palmitoylated CaV_2a relative to CaV2.2 is critical for slow pathway modulation of N-type Ca2+ current by tachykinin receptor activation, *Journal of Biomedical Discovery and Collaboration*

2009   Moon, H. et al., Lactational coumestrol exposure increases ovarian apoptosis, *Archives of Toxicology*

2009   Mwai, L. et al., Chloroquine resistance before and after its withdrawal in Kenya, *Malaria Journal*

2009   Oman, K. et al., Specialist training in Fiji: Why do graduates migrate, and why do they remain? a qualitative study, *Human Resources for Health*

2009   Pernet, C. et al., Brain classification reveals the right cerebellum as the best biomarker of dyslexia, *BMC Neuroscience*

2009   Pike, N. et al., Antibiotic treatment leads to the elimination of Wolbachia endosymbionts and sterility in the diplodiploid collembolan Folsomia candida, *BMC Biology*

2009   Qiu, B. et al., High-throughput assay for modulators of mitochondrial membrane potential identifies a novel compound with beneficial effects on db/db mice, *Diabetes*

2009   Rojas, E. et al., DNA-AP sites generation by Etoposide in whole blood cells, *BMC Cancer*

2009   Sloan, R. et al., Associations between cardiorespiratory fitness and health-related quality of life, *Health and Quality of Life Outcomes*

2009   Soumahoro, M. et al., Impact of Chikungunya virus infection on health status and quality of life: a retrospective cohort study, *PLoS One*

2009   Traylor, R. et al., Microdeletion of 6q16.1 encompassing EPHA7 in a child with mild neurological abnormalities and dysmorphic features: case report, *Molecular Cytogenetics*

2009   Vestby, L. et al., Biofilm forming abilities of Salmonella are correlated with persistence in fish meal and feed factories, *BMC Veterinary Research*

2009   Volandes, A. et al., Video decision support tool for advance care planning in dementia: randomised controlled trial, *BMJ (Clinical Research)*

2009   Wagner, K. et al., Single-molecule analysis reveals two separate DNA-binding domains in the Escherichia coli UvrA dimer, *Nucleic Acids Research*

2009   Wehmeier, P. et al., Change in the direct cost of treatment for children and adolescents with hyperkinetic disorder in Germany over a period of four years, *Child and Adolescent Psychiatry and Mental Health*

2009   Williams, A. et al., Experimentression profiling inmild asthmatic human airways and effect of corticosteroid therapy, *PLoS One*

2009   Wolthuis, E. et al., Mechanical ventilation using non-injurious ventilation settings causes lung injury in the absence of pre-existing lung injury in healthy mice, *Critical Care*

2009   Wu, F. et al., Roles of the creatine kinase system and myoglobin in maintaining energetic state in the working heart, *BMC Systems Biology*

2009   Yang, C. et al., The SmpB-tmRNA tagging system plays important roles in Streptomyces coelicolor growth and development, *PLoS One*

2009   Yen, C. et al., Effects of hip joint position and intra-capsular volume on hip joint intra-capsular pressure: a human cadaveric model, *Journal of Orthopedic Surgery and Research*

2009   Young, L. et al., Bringing home the trash: do colony-based differences in foraging distribution lead to increased plastic ingestion in laysan albatrosses? *PLoS One*

2009   Yu, X., Socioeconomic disparities in breast cancer survival: relation to stage at diagnosis, treatment and race, *BMC Cancer*

2009   Zhang, X. et al., Amphiphilic blockers punch through a mutant CLC-0 pore, *The Journal of General Physiology*

2009   Zhao, J. et al., Cloning, identification, and characterization of the rpoS-like sigma factor rpoX from Vibrio alginolyticus, *Journal of Biomedicine and Biotechnology*

2010   Allender, C. et al., Origins of the amphiploid species Brassica napus L. investigated by chloroplast and nuclear molecular markers, *BMC Plant Biology*

2010   Beriat, N. et al., Effect of different polymerization methods on the cytotoxicity of dental composites, *European Journal of Dentistry*

2010   Bielecka-Dabrowa, A. et al., Influence of co-existing atrial fibrillation on the efficacy of atorvastatin treatment in patients with dilated cardiomyopathy, *Lipids in Health and Disease*

2010   Brune, T. et al., Quality, stability, and safety data of packed red cells and plasma processed by gravity separation using a new fully integrated hollow-fibre filter device, *Advances in Hematology*

2010   Cameirão, A. et al., Neurorehabilitation using the virtual reality based Rehabilitation Gaming System: methodology, design, psychometrics, usability and validation, *Journal of Neuroengineering and Rehabilitation*

2010   Chen, M. et al., A PPAR_ promoter variant impairs ERR-dependent transactivation and decreases mortality after acute coronary ischemia in patients with diabetes, *PLoS One*

2010   Chodroff, R. et al., Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes, *Genome Biology*

2010   Chui, R. et al., Regulation of CCL2 and CCL3 expression in human brain endothelial cells by cytokines and lipopolysaccharide, *Journal of Neuroinflammation*

2010   Clifford, D. et al., Novel O-palmitolylated beta-E1 subunit of pyruvate dehydrogenase is phosphorylated during ischemia/reperfusion injury, *Proteome Science*

2010   Dana, N. et al., Production objectives and trait preferences of village poultry producers of Ethiopia: implications for designing breeding schemes utilizing indigenous chicken genetic resources, *Tropical Animal Health and Production*

2010   deGreeff, S. et al., Seroprevalence of pertussis in the Netherlands: evidence for increased circulation of bordetella pertussis, *PLoS One*

2010   Dorobantu, M. et al., Prevalence, awareness, treatment, and control of hypertension in Romania: results of the SEPHAR Study, International, *Journal of Hypertension*

2010   Douglas, R. et al., Improving trauma care in India: a recommendation for the implementation of ATLS training for emergency department medical officers, *International Journal of Emergency Medicine*

2010   Douillard, F. et al., The HP0256 gene product is involved in motility and cell envelope architecture of Helicobacter pylori, *BMC Microbiology*

2010   Douw, L. et al., Functional connectivity is a sensitive predictor of epilepsy diagnosis after the first seizure, *PLoS One*

2010   Fehri, L. et al., Helicobacter pylori Induces miR-155 in T cells in a cAMP-Foxp3-dependent manner, *PLoS One*

2010   Feisner, G. et al., The physical and functional borders of transit peptide-like sequences in secondary endosymbionts, *BMC Plant Biology*

2010   Firth, L. et al., Cell cycle arrest by a gradient of Dpp signaling during Drosophila eye development, *BMC Developmental Biology*

2010   Földes, G. et al., Innate immunity in human embryonic stem cells: comparison with adult human endothelial cells, *PLoS One*

2010   Garrida-Estepa, M .et al., Evaluation of mammographic density patterns: reproducibility and concordance among scales, *BMC Cancer*

2010   Gumbo, F. et al., Risk factors of HIV vertical transmission in a cohort of women under a PMTCT program at three peri-urban clinics in a resource-poor setting, *Journal of Perinatology*

2010   Gupta, R. et al., The state of pediatric asthma in Chicago's Humboldt Park: a community-based study in two local elementary schools, *BMC Pediatrics*

2010   Hauabold, M. et al., Bone morphogenetic protein 4 (BMP4) signaling in retinoblastoma cells, *International Journal of the Biological Sciences*

2010   Huopaniemi, I. et al., Multivariate multi-way analysis of multi-source data, *Bioinformatics*

2010   Ion, G. et al., Maternal consumption of canola oil suppressed mammary gland tumorigenesis in C3(1) TAg mice offspring, *BMC Cancer*

2010   Jeevaratnam, K. et al., Periodic assessment of plasma sFlt-1 and PlGF concentrations and its association with placental morphometry in gestational hypertension (GH) - a prospective follow-up study, *BMC Pregnancy and Childbirth*

2010   Jianfang, H. et al., Itk derived signals regulate the expression of Th-POK and controls the development of CD4+ T cells, *PLoS One*

2010   Jima, D. et al., Malaria indicator survey 2007, Ethiopia: coverage and use of major malaria prevention and control interventions, *Malaria Journal*

2010   Johnsen, E. et al., Effectiveness of second-generation antipsychotics: a naturalistic, randomized comparison of olanzapine, quetiapine, risperidone, and ziprasidone, *BMD Psychiatry*

2010   Kammerer, R. et al., Coevolution of activating and inhibitory receptors within mammalian carcinoembryonic antigen families, *BMC Biology*

2010   Kaur, G. et al., Calmodulin-dependent nuclear import of HMG-box family nuclear factors: importance of the role of SRY in sex reversal, *The Biochemical Journal*

2010   Kim, B. et al., Relationship between new osteoporotic vertebral fracture and instrumented lumbar arthrodesis, *Asian Spine Journal*

2010   Kim, Y. et al., Midkine, heparin-binding growth factor, blocks kainic acid-induced seizure and neuronal cell death in mouse hippocampus, *BMC Neuroscience*

2010   Kollmorgen, S. et al., Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention, *PLoS Computational Biology*

2010   Komatsu, H. et al., Cellular immunity in children with successful immunoprophylactic treatment for mother-to-child transmission of hepatitis B virusMC, *Infectious Diseases*

2010   Kuwahara, H. et al., Temperature control of fimbriation circuit switch in uropathogenic escherichia coli: quantitative analysis via automated model abstraction, *PLoS Computational Biology*

2010   Liao, Q. et al., Situational awareness and health protective responses to pandemic influenza A (H1N1) in Hong Kong: a cross-sectional study, *PLoS One*

2010   Liu, S. et al., Androgen excess produces systemic oxidative stress and predisposes to cell failure in female mice, *PLoS One*

2010   Liu, X. et al., Functional analysis of GbAGL1, a D-lineage gene from cotton (Gossypium barbadense), *Journal of Experimental Botany*

2010   Marchini, C. et al., Is there a nonadherent subtype of hypertensive patient? a latent class analysis approach, *Patient Preference and Adherence*

2010   Markham, J. et al., Characterization of the cognitive impairments induced by prenatal exposure to stress in the rat, *Frontiers in Behavioral Neuroscience*

2010   Makrs, T., Mesenchymal/stromal gene expression signature relates to basal-like breast cancers, identifies bone metastasis and predicts resistance to therapies, *International Journal of Hypertension*

2010   Meybohm, P., Mild hypothermia alone or in combination with anesthetic post-conditioning reduces expression of inflammatory cytokines in the cerebral cortex of pigs after cardiopulmonary resuscitation, *Critical Care*

2010   Miles, S. et al., Statistics teaching in medical school: opinions of practising doctors, *BMC Medical Education*

2010   Mwakigonja, A. et al., Tanzanian malignant lymphomas: WHO classification, presentation, ploidy, proliferation and HIV/EBV association, *PLosOne*

2010   Nakamura, M. et al., TRPM1 mutations are associated with the complete form of congenital stationary night blindness, *Molecular Vision*

2010   O'Grady, S. et al., Aberrant water homeostasis detected by stable isotope analysis, *PloS One*

2010   Palinichamy, M .et al., Potential pitfalls in MitoChip detected tumor-specific somatic mutations: a call for caution when interpreting patient data, *BMC Cancer*

2010   Rassweller, A. et al., Triggers and maintenance of multiple shifts in the state of a natural community, *Oecologia*

2010   Rato, S. et al., Novel HIV-1 knockdown targets identified by an enriched kinases/ phosphatases shRNA library using a long-term iterative screen in Jurkat T-cells, *PLoS One*

2010   Riep, B. et al., Tumor necrosis factor and norepinephrine lower the levels of human neutrophil peptides 1-3 secretion by mixed synovial tissue cultures in osteoarthritis and rheumatoid arthritis, *Arthritis Research & Therapy*

2010   Rurup, M. et al., The use of opioids at the end of life: the knowledge level of Dutch physicians as a potential barrier to effective pain management, *BMC Palliative Care*

2010   Rutgers, M. et al., Cytokine profile of autologous conditioned serum for treatment of osteoarthritis, in vitro effects on cartilage metabolism and intra-articular levels after injection, *Asian Spine Journal*

2010   Sartori, T. et al., Influence of demographic and metabolic variables on forearm blood flow and vascular conductance in individuals without overt heart disease, *Vascular Health and Risk Management*

2010   Schütz, M. et al., The human deafness-associated connexin 30 T5M mutation causes mild hearing loss and reduces biochemical coupling among cochlear non-sensory cells in knock-in mice, *Human Molecular Genetics*

2010   Chen, M. et al., Serological response in RT-PCR confirmed H1N1-2009 influenza A by hemagglutination inhibition and virus neutralization assays: an observational study, *PLoS One*

2010   Sheppard, S. et al., LIM-domain proteins TRIP6 and LPP associate with shelterin to mediate telomere protection, *Aging*

2010   Sloan, R. et al., Expression of Nef from unintegrated HIV-1 DNA downregulates cell surface CXCR4 and CCR5 on T-lymphocytes, *Retrovirology*

2010   Steinberg, B. et al., A cation counterflux supports lysosomal acidification, *Journal of Cell Biology*

2010   Sulaiman, N. et al., The prevalence and correlates of depression and anxiety in a sample of diabetic patients in Sharjah, United Arab Emirates, *BMC Family Practice*

2010   Tomioka, I. et al., Generating induced pluripotent stem cells from common marmoset (Callithrix jacchus) fetal liver cells using defined factors, including Lin28, *Genes to Cells*

2010   Wandel, S. et al., Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis, *BMJ (Clinical Research)*

2010   Weth, O. et al., Modular insulators: genome wide search for composite CTCF/thyroid hormone receptor binding sites, *PLoS One*

2010   Wu, W. et al., Biocompatibility of Fe3O4/DNR magnetic, nanoparticles in the treatment of hematologic, nanoparticles in the treatment of hematologic malignancies, *International Journal of Nanomedicine*

2010   Yang, Y. et al., The tricarboxylic acid cycle in Shewanella oneidensis is independent of fur and RyhB control, *BMC Microbiology*

2010   Yu, B. et al., Expression of the apoptosis-related genes BCL-2 and BAD in human breast carcinoma and their associated relationship with chemosensitivity, *Journal of Experimental and Clincal Cancer Research*

2010   Yusein, S. et al., Functional consequences of mutations in the Drosophila histamine receptor, HCLB, *Journal of Insect Physiology*

2010   Zong, J. et al., Coping flexibility in college students with depressive symptoms *Health and Quality of Life Outcomes*

2010   Zwart, J. et al., Obstetric intensive care unit admission: a 2-year nationwide population-based cohort study, *Intensive Care Medicine*

2011   Chang, Y. et al., Pathogenic VCP/TER94 alleles are dominant actives and contribute to neurodegeneration by altering cellular ATP level in a drosophila IBMPFD model, *PLoS Genetics*

2011   Crosby, M. et al., Serum vascular endothelial growth factor (VEGF) levels correlate with number and location of micrometastases in a murine model of uveal melanoma, *British Journal of Opthamology*

2011   Demirag, G. et al., Expression of plakophilins (PKP1, PKP2, and PKP3) in gastric cancers, *Diagnostic Pathology*

2011   Jankowska, E. et al., Identification of chronic heart failure patients with a high 12-month mortality risk using biomarkers including Plasma C-Terminal Pro-Endothelin-1, *PLoS One*

2011   Payzn-LeNestour, E. et al., Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings, *PLoS Computational Biology*

2011   Weinstock-Guttman, L. et al., Chronic cerebrospinal vascular insufficiency is not sssociated with HLA DRB1*1501 status in Multiple Sclerosis patients, *PLoS One*

2011   Winter, J. et al., Blood vessel density in basal cell carcinomas and benign trichogenic tumors as a marker for differential diagnosis in dermatopathology, *Journal of Skin Cancer*

2011   Zhao, R. et al., An autosomal dominant cataract locus mapped to 19q13-qter in a Chinese family, *Molecular Vision*

Appendix 2: **Journals in the medical research article corpus** (n=250)

Numbers in parentheses indicate number of times a journal is represented

Acta Veterinaria Scandinavica
Advances in Hematology
Aging
Algorithms for Molecular Biology
Applied and Environmental Microbiology
Archives of Drug Information
Archives of Toxicology
Arthritis Research & Therapy (6)
Asian Spine Journal (2)
Behavior Therapy
Behavioral and Brain Functions
Bioinformatics (2)
BMC Bioinformatics
BMC Biology (2)
BMC Biotechnology
BMC Blood Disorders
BMC Cancer (10)
BMC Cell Biology
BMC Developmental Biology
BMC Evolutionary Biology (2)
BMC Family Practice
BMC Gastroenterology
BMC Genetics
BMC Genomics
BMC Health Services Research (3)
BMC Infectious Diseases
BMC Medical Education (4)
BMC Medical Genetics (2)
BMC Medical Informatics and Decision Making
BMC Medicine
BMC Microbiology (2)
BMC Musculoskeletal Disorders
BMC Neurology
BMC Neuroscience (3)
BMC Opthamology
BMC Oral Health
BMC Palliative Care
BMC Pediatrics
BMC Physiology

BMC Plant Biology (3)

BMC Pregnancy and Childbirth

BMC Psychiatry

BMC Public Health (4)

BMC Pulmonary Medicine

BMC Research Notes

BMC Surgery

BMC Systems Biology

BMC Veterinary Research

BMC Women's Health

BMD Psychiatry

BMJ (Clinical Research) (3)

Breast Cancer Research (2)

British Journal of Opthamology

Carcinogenesis

Cell & Chromosome

Cell Research

Child and Adolescent Psychiatry and Mental Health

Critical Care (2)

Diabetes

Diabetes Care

Diagnostic Pathology (2)

Diseases of the Colon and Rectum

Domestic Animal Endocrinology

Emerging Infectious Diseases (2)

Environmental Health Perspectives (2)

European Journal of Dentistry (2)

Frontiers in Behavioral Neuroscience (2)

Genes to Cells

Genetics, Selection, Evolution

Genome Biology (5)

Head & Face Medicine

Health and Quality of Life Outcomes (2)

Health Research Policy and Systems (3)

Human Molecular Genetics

Human Resources for Health

Implementation Science

Intensive Care Medicine

International Journal of the Biological Sciences

International Journal of Behavioral and Physical Activity

International Journal of Chronic Obstructive Pulmonary Disease

International Journal of Emergency Medicine

International Journal of Health Geographies

International Journal of Hypertension (2)

The Biochemical Journal
The International Journal of Behavioral Nutrition and Physical Activity
The Journal of General Physiology
Tobacco Induced Diseases
Tropical Animal Health and Production
Upsala Journal of Medical Sciences (2)
Vascular Health and Risk Management

Appendix 3: Discourse community sample interview

Interview questions for doctors: What does the discourse community of doctors consider to be the process of writing and publishing scientific/medical articles?

Name_____

Area of Medicine _____Years Practicing_____

How many journal articles do you read a month? _____

What areas are those journals in?

_____

How many articles have you published in peer reviewed journals?_____

When was the first time you published? _____ When was the last time? _____

What is your first language? _____

Part I: Some background questions

1. What do you use journals for?

2. What is your own definition of a research article?

3. How do you feel about journals that are online only vis-a-vis printed journals?

4. Is there a difference between a research article and a review article? If so, please explain.

Now I would like you to talk about writing an article.

5. Please tell me, step by step, how you write an article, and conduct research to do it.

6. What section is the hardest for you to write and why?

7. What, in your opinion, belongs in the introduction?

8. What belongs in the methods?

9. What belongs in the results?

10. What belongs in the discussion/conclusion?

11. How do you feel about the statement 'publish or perish'?

12. Have you ever taken classes on how to write for publishing, or have you read any specific books on the topic?

13. What do you do about proofreading, do you, for example, go to an editor or friend for help?

Now I would like to ask about your process of reading an article

14. When you read an article, do you read straight through or jump around and why?

15. What is the hardest part of reading research articles?

16. Have articles changed in your time as a doctor and if so, has the change been for the better or worse?

Let's talk about the field

17. Are articles too data-driven or not enough?

18. Have you ever noticed any difference between American doctors' papers and non-American or foreign doctors' work?

19. Does anything need to be done to improve the medical humanities in relation to the global village?

20. Are there any suggestions that have helped you when it comes to writing an article?

21. What would you tell a fellow doctor to help them get published?

Part II: Components of Articles

Fill in the chart with the title of the section in which you would expect to find the following items and then in a few words, what they are for. Sections are: Introduction, Methods, Results, Discussion, Abstract and Elsewhere. Items may appear in more than one section, but place them where you expect to see them. An example is done for you.

Items	Section	Used for
Acknowledgements	Elsewhere	Identifying who helped or where funding was from.
Diagrams		
Percentages		
Statement of impartiality		
Lists of materials used		
Bio-data of subjects		
The history of the field		
Limitations of the experiment		
The treatment		
Definition of problem that is to be examined		

Appendix 4: Classification of Reporting Verbs

In the following, tables for the 13 Reporting Verbs being examined have been separated into 5 categories.

1. *Class*, which refers to the class of word and has been categorized as follows:

Reporting Verb – a verb being used to report something: *John claims success*.

Action Verb – when the verb is an action other than reporting: *John sang today*.

Infinitive – the unmarked or base form of the verb: *John wants **to sing***.

Presupposition – a belief assumed by speaker/writer and hearer/reader: the statement that *John is happy* presupposes the existence of John

Noun – a word that identifies persons, places, things, and abstract entities: in *John claims success*, John is a proper noun and success is a common noun

Noun Phrase – a phrase headed by a noun, which typically includes one or more modifiers, although it can be a single word: *the red-headed woodpecker*

Verb Phrase – a phrase consisting of a verb and its auxiliary or modal particles, typically forming the predicate: ate seventeen mealy worms

Adjective – a word that modifies a noun: *red-headed* in *red-headed woodpecker*

Idiom – a multi-word construction that cannot be decomposed; its meaning can not be derived from its parts: *kick the bucket* does not mean a literal action

Participle – a word derived from a verb, usually using –ing/-ed, to be used as an adjective: shining in *the shining sands*

2. *Actor* refers to the person or thing that is stating the information being reported

Authors refers to those who wrote the text in question

Others refers to other authors or other studies

Data is used for any raw findings

Results are used for finalized findings.

Field refers to the particular field of medicine being reported and typically represents a consensus viewpoint from *Others* working in that area

Study Participants/participants refers to those people being tested

Algorithm is a formula or equation that is used for finding the solution

3. *Semantic prosody* presents words found in the span of the node, that are 'colouring' the node

4. *Level* refers to whether the information is describing the current *experiment* or the *field* of medicine (see 5.4.2)

5. *Location* is the section in which the occurrence appears (i.e. Introduction (I), Methods (M), Results (R), Discussion (D)).

Table 1A: Argue

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others	one can	Field	I
2.	RV	Others	attempted	Field	I
3.	Presupposition	Others	one can	Field	I
4.	PS	Others	some	Field	I
5.	RV	Others	try	Field	I
6.	PS	Authors	one can	Field	I
7.	RV	Authors	our findings	Experiment	I
8.	RV	Others	Some try	Field	I
9.	RV	Authors	Further	Field	R
10.	PS	Results	The results	Experiment	R
11.	RV	Authors	results	Experiment	D
12.	RV	Others	some authors	Field	D
13.	RV	Others	these data	Experiment	D
14.	RV	Others	indeed	Field	D
15.	RV	Results	the results	Experiment	D
16.	RV	Results	present findings	Experiment	D
17.	RV	Others	seems	Field	D
18.	RV	Results	our findings	Experiment	D
19.	PS	Others	some	Experiment	D
20.	PS	Others	others	Experiment	D
21.	RV	Authors	our data	Experiment	D
22.	RV	Others	taken together	Experiment	D
23.	RV	Results	but	Experiment	D
24.	PS	Others	one could	Experiment	D
25.	RV	Authors	seem sensible	Experiment	D
26.	RV	Results	might	Field	D
27.	RV	Authors	strongly	Field	D
28.	PS	Others	one could	Experiment	D

	Class	Actor	Semantic prosody	Level	Location
29.	RV	Others	however	Experiment	D
30.	RV	Others	could	Field	D
31.	PS	Others	one could	Experiment	D
32.	PS	Others	some	Experiment	D
33.	RV	Authors	although	Experiment	D

Table 1B: Argued

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Authors	do	Field	I
2.	RV	Others	could	Field	I
3.	RV	Others	may	Field	I
4.	RV	Others	but	Field	I
5.	PS	Results	could	Experiment	R
6.	RV	Others	however	Field	R
7.	RV	Others	confidently	Field	R
8.	RV	Others	might	Experiment	D
9.	RV	Others		Field	D
10.	PS	Others	some	Field	D
11.	RV	Others		Field	D
12.	RV	Authors		Experiment	D
13.	RV	Others		Field	D
14.	PS	Results	it could	Experiment	D
15.	PS	Others	come	Field	D
16.	RV	Authors	might	Field	D
17.	PS	Results	it could	Experiment	D
18.	PS	Others	even though	Field	D
19.	PS	Others	however	Experiment	D

Table 1C: Argues

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others		Field	I
2.	RV	Others	has since	Field	I
3.	RV	Others		Field	R
4.	RV	Others	but	Field	D
5.	RV	Authors	none can	Experiment	D
6.	RV	Others	truly	Field	D

Table 2A: Claim

	Class	Actor	Semantic prosody	Level	Location
1.	Noun (N)	Results	seems to support	Experiment	M
2.	N	Others	separate	Experiment	M
3.	N	Others	previous	Experiment	M
4.	N	Authors		Experiment	M
5.	RV	Others	difficult	Field	M
6.	RV	Others	flawed	Field	R
7.	RV	Authors	not	Experiment	D
8.	RV	Others	uncertainty	Field	D
9.	RV	Authors	although	Experiment	D
10.	RV	Others	might	Experiment	D
11.	N	Results	supports	Experiment	D
12.	N	Authors		Experiment	D
13.	RV	Authors	In spite	Field	D

Table 2B: Claimed

	Class	Actors	Semantic prosody	Level	Location
1.	RV	Others	controversially	Field	I
2.	RV	Others	found evidence for	Field	I
3.	RV	Others	despite	Field	I
4.	RV	Survey participants	but, only	Experiment	R
5.	RV	Survey participants		Experiment	R
6.	RV	Survey participants		Experiment	R
7.	RV	Survey participants		Experiment	R
8.	RV	Study participants		Experiment	R
9.	RV	Others		Field	D
10.	RV	Survey participants	sharp contrast	Experiment	D
11.	RV	Survey participants		Experiment	D

Table 2C: Claims

	Class	Actor	Semantic prosody	Level	Location
1.	N	Results	appear		I
2.	N	Results	seems		M
3.	N	Results	isolated		M
4.	N	Results			M
5.	N	Results	isolated		M
6.	N	Others	claims and claims		M
7.	N	Results	conflicts	Field	R
8.	N	Results	seems	Field	R
9.	N	Authors	significant	Experiment	R
10.	N	Others			D
11.	N	Others	can		D
12.	RV	Others	recent	Experiment	D
13.	N	Others			D

Table 3A: Discover

	Class	Actor	Semantic prosody	Level	Location
1.	Infinitive (INF)	Others		Field	I
2.	INF	Previous studies	misconceptions	Field	I
3.	INF/RV	Authors		Experiment	I
4.	Action verb (AV)	Authors	new	Experiment	M
5.	INF	Participants		Experiment	R
6.	INF	Participants	novel	Experiment	R
7.	INF	Everyone		Field	D
8.	INF/RV	Authors	first	Experiment	D
9.	INF	Participants	initially	Experiment	D

Table 3B: Discovered

	Class	Actor	Semantic prosody	Level	Location
1.	AV	Others		Field	I
2.	AV	Others		Field	I
3.	AV	Others	however	Field	I

	Class	Actor	Semantic prosody	Level	Location
4.	AV	Others	recently	Field	I
5.	AV	Others		Field	I
6.	RV	Authors	unexpected	Experiment	I
7.	AV	Others	currently	Field	I
8.	AV	Authors		Experiment	M
9.	AV	Authors		Experiment	M
10.	N	Authors	yet to be	Experiment	M
11.	RV	Authors		Experiment	R
12.	RV	Authors	yet	Experiment	R
13.	RV	Authors	newly	Experiment	R
14.	AV	Authors	recently	Experiment	R
15.	AV	Authors	recently	Experiment	R
16.	RV	Authors	notably	Experiment	R
17.	AV	Authors	recently	Experiment	R
18.	RV	Authors	however	Experiment	R
19.	RV	Authors	might	Experiment	R
20.	RV	Authors	increasingly	Experiment	R
21.	RV	Authors		Experiment	R
22.	RV	Authors	unexpected	Experiment	D
23.	AV	Authors	significant	Field	D
24.	RV	Authors	necessary	Experiment	D
25.	RV	Authors	in fact	Experiment	D
26.	N	Field		Field	D
27.	RV	Authors	striking feature	Experiment	D
28.	AV	Others		Field	D
29.	AV	Others		Field	D
30.	N	Others		Field	D

Table 4A: Find

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Authors	previously/now	Experiment	I
2.	AV	Others		Field	I
3.	INF	Others	emphasize	Field	I
4.	AV	Others	surprisingly	Field	I
5.	RV	Authors	estimate	Experiment	M
6.	INF	Authors	consequently	Experiment	M
7.	INF	Authors		Experiment	M
8.	AV	Authors	can/goal	Experiment	M
9.	RV	Authors	Unlike	Experiment	R
10.	RV	Authors	If/then contrastive	Experiment	R
11.	AV	Authors	could no	Experiment	R
12.	AV	Authors	did not	Experiment	R
13.	AV	Authors	can	Experiment	R
14.	RV	Authors	despite	Experiment	R
15.	INF	Participants		Experiment	R
16.	RV	Authors	possible exception	Experiment	R
17.	AV	Participant		Experiment	R
18.	RV	Authors	did not	Experiment	D
19.	RV	Authors	in contrast	Experiment	D
20.	RV	Authors	surprised	Experiment	D
21.	RV	Authors	at least in this case	Experiment	D
22.	RV	Authors	did not find this to be true	Experiment	D
23.	RV	Authors	in contrast	Experiment	D
24.	INF	Anyone/		Field	D

			suggestion		
25.	RV	Authors	in this study/ contrastive	Experiment	D
26.	RV	Authors	pleased	Experiment	D
27.	RV	Authors	first	Experiment	D
28.	INF	Authors	the reasons	Experiment	D
29.	RV	Authors	specifically	Experiment	D
30.	AV	Authors	couldn't	Experiment	D
31.	AV	Authors	not able	Experiment	D
32.	AV	Others	did not	Field	D
33.	INF	Authors	surprisingly hard	Experiment	D

Table 4B: Found

	Class	Actor	Semantic prosody	Level	Location
1.	AV	Others	did	Field	I
2.	RV	Authors	on the other hand	Field	I
3.	RV	Others	did not	Field	I
4.	RV	Others	have	Field	I
5.	RV	Machine	detected	Experiment	M
6.	RV	Authors		Experiment	M
7.	RV	Authors	significant	Experiment	R
8.	RV	Authors		Experiment	R
9.	RV	Authors	than	Experiment	R
10.	RV	Authors		Experiment	R
11.	RV	Authors	however	Experiment	R
12.	RV	Authors		Experiment	R
13.	RV	Authors		Experiment	R
14.	RV	Participants	between	Experiment	R
15.	RV	Authors		Experiment	R
16.	RV	Authors		Experiment	R
17.	RV	Authors		Experiment	R
18.	RV	Authors	contrary	Experiment	D
19.	RV	Authors		Experiment	D
20.	RV	Others	have	Field	D
21.	RV	Authors	well beyond	Experiment	D
22.	RV	Authors	although	Experiment	D
23.	RV	Authors	however	Experiment	D
24.	RV	Others	previously	Field	D
25.	RV	Authors	however	Experiment	D
26.	RV	Others		Field	D
27.	RV	Authors	as compared	Both	D
28.	RV	Authors	but…not	Experiment	D
29.	RV	Authors	although	Experiment	D
30.	RV	Authors		Experiment	D
31.	RV	Others	previously	Field	D
32.	RV	Authors	significant	Experiment	D
33.	RV	Authors	correlation	Experiment	D

Table 4C: Finds

	Class	Actor	Semantic prosody	Level	Location
1.	AV	Algorithm		Experiment	I
2.	AV	Review		Experiment	I
3.	AV	Method		Experiment	I
4.	AV	Method		Experiment	I

5.	AV	The tool		Experiment	M
6.	AV	The method		Experiment	R
7.	AV	The model		Experiment	R
8.	AV	The method		Experiment	R
9.	AV	The method		Experiment	R
10.	AV	The method	however	Experiment	R
11.	AV	RSS feed		Field	R

Table 5A: Indicate

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Studies		Field	I
2.	RV	Studies		Field	I
3.	RV	Results		Experiment	I
4.	INF	Data		Experiment	I
5.	INF	Participants		Experiment	M
6.	RV	Authors		Experiment	M
7.	RV	Participants		Experiment	M
8.	AV	Authors		Experiment	M
9.	RV	Results		Experiment	R
10.	RV	Results	may	Experiment	R
11.	AV	Authors		Experiment	R
12.	RV	Data		Experiment	R
13.	RV	Results	clearly	Experiment	R
14.	RV	Observations	despite	Experiment	R
15.	RV	Results	might	Experiment	R
16.	RV	Data		Experiment	R
17.	INF	Others	seemingly	Field	R
18.	RV	Results	seems	Experiment	R
19.	RV	Data		Experiment	D
20.	RV	Results		Field	D
21.	RV	Data		Experiment	D
22.	RV	Results		Experiment	D
23.	RV	Data	suggesting/ may	Field	D
24.	INF	Others		Experiment	D
25.	RV	Data		Experiment	D
26.	RV	Data		Experiment	D
27.	RV	Results	appears to	Experiment	D
28.	RV	Others	suggest	Field	D
29.	RV	Data		Experiment	D
30.	RV	Results	suggesting	Experiment	D
31.	RV	Results		Experiment	D
32.	RV	Results	strongly	Experiment	D
33.	RV	Results	seems	Experiment	D

Table 5B: Indicated

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Data		Experiment	I
2.	RV	Follow-up evaluation		Field	I
3.	RV	Data		Experiment	M
4.	RV	Others		Experiment	M
5.	Noun Phrase (NP)	Others		Field	M
6.	RV	Others		Experiment	M

7.	AV	Data		Experiment	M
8.	NP	Data		Experiment	M
9.	RV	Results		Experiment	M
10.	NP	Data		Experiment	M
11.	RV	Data		Experiment	R
12.	RV	Results		Experiment	R
13.	RV	Data		Experiment	R
14.	RV	Others	might	Field	R
15.	RV	Analysis		Experiment	R
16.	RV	Results		Experiment	R
17.	RV	Data		Experiment	R
18.	RV	Data		Experiment	R
19.	RV	Data		Experiment	R
20.	RV	Data		Experiment	R
21.	RV	Results		Experiment	R
22.	RV	Data		Experiment	D
23.	RV	Results		Experiment	D
24.	RV	Results		Experiment	D
25.	RV	Others		Field	D
26.	RV	Data		Experiment	D
27.	RV	Results		Experiment	D
28.	RV	Others	may	Field	D
29.	RV	Others		Field	D
30.	RV	Participants		Field	D
31.	RV	Participants		Experiment	D
32.	RV	Results		Experiment	D
33.	RV	Others	differ	Experiment	D

Table 5C: Indicates

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others		Field	I
2.	RV	Others		Field	I
3.	RV	Others		Field	I
4.	RV	Data		Experiment	M
5.	RV	Data		Experiment	M
6.	RV	Results	despite	Experiment	R
7.	RV	Results	not only	Experiment	R
8.	RV	Results	more than	Experiment	R
9.	RV	Data		Experiment	R
10.	RV	Data	may	Experiment	R
11.	RV	Results		Experiment	R
12.	RV	Results		Experiment	R
13.	RV	Others		Field	R
14.	RV	Results		Experiment	R
15.	RV	Data		Experiment	R
16.	RV	Data		Experiment	R
17.	RV	Results		Experiment	D
18.	RV	Data	the lack of	Experiment	D
19.	RV	Data		Experiment	D
20.	RV	Data		Experiment	D

	Class	Actor	Semantic prosody	Level	Location
21.	RV	Data		Experiment	D
22.	RV	Results		Field	D
23.	RV	Results		Experiment	D
24.	RV	Results		Experiment	D
25.	RV	Results		Experiment	D
26.	RV	Results	the need for	Experiment	D
27.	RV	Results		Field	D
28.	RV	Others		Field	D
29.	RV	Data	the possibility	Experiment	D
30.	RV	Results		Experiment	D
31.	RV	Data		Experiment	D
32.	RV	Results		Experiment	D
33.	RV	Results		Experiment	D

Table 6A: Observe

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Authors	first time	Experiment	I
2.	INF	Authors	set out	Experiment	I
3.	INF	Authors		Experiment	I
4.	INF	Authors		Experiment	M
5.	INF	Authors		Experiment	M
6.	INF	Authors		Experiment	R
7.	RV	Authors	even at	Experiment	R
8.	AV	Authors		Experiment	R
9.	AV	Authors	significant	Experiment	R
10.	RV	Authors		Experiment	R
11.	RV	Authors		Experiment	R
12.	AV	Authors	should	Experiment	R
13.	RV	Authors		Experiment	R
14.	RV	Authors	can	Experiment	R
15.	RV	Authors		Experiment	R
16.	INF	Authors	not possible	Experiment	R
17.	RV	Authors		Experiment	R
18.	RV	Authors		Experiment	R
19.	RV	Authors		Experiment	R
20.	RV	Authors		Experiment	D
21.	RV	Authors		Experiment	D
22.	RV	Others		Field	D
23.	INF	Authors		Experiment	D
24.	RV	Authors	not significant	Experiment	D
25.	RV	Authors		Experiment	D
26.	INF	Authors		Experiment	D
27.	INF	Authors		Experiment	D
28.	RV	Authors	did not	Experiment	D
29.	RV	Others	also	Field	D
30.	RV	Authors	did not	Experiment	D
31.	INF	Authors		Experiment	D
32.	RV	Authors		Experiment	D
33.	INF	Authors		Experiment	D

Table 6B: Observed

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others	since	Field	I
2.	INF	Participants		Experiment	I
3.	NP	Authors		Experiment	M
4.	NP	Model		Experiment	M
5.	AV	Authors		Experiment	M
6.	RV	Authors	in particular	Experiment	R
7.	RV	Authors	than	Experiment	R
8.	RV	Authors		Experiment	R
9.	RV	Authors		Experiment	R
10.	NP	Authors		Experiment	R
11.	RV	Authors	similar results	Experiment	R
12.	RV	Authors		Experiment	R
13.	RV	Authors		Experiment	R
14.	RV	Authors		Experiment	R
15.	RV	Authors		Experiment	R
16.	RV	Authors	in contrast	Experiment	R
17.	RV	Authors		Experiment	R
18.	RV	Authors		Experiment	R
19.	AV	Authors		Experiment	R
20.	RV	Authors		Experiment	R
21.	RV	Others		Field	D
22.	RV	Authors		Experiment	D
23.	N	Authors		Experiment	D
24.	RV	Authors		Experiment	D
25.	RV	Authors	although	Experiment/field	D
26.	AV	Others		Field	D
27.	NP	Authors	No other	Experiment	D
28.	RV	Other		Field	D
29.	RV	Authors		Experiment	D
30.	RV	Authors	only	Experiment	D
31.	RV	Others		Field	D
32.	RV	Authors		Experiment	D
33.	RV	Authors		Experiment	D

Table 7A: Pointed out

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others	later questioned	Field	I
2.	RV	Reviews	nevertheless	Field	I
3.	RV	Others	do not strictly follow	Field	I
4.	RV	Others		Field	R
5.	RV	Results	the discrepancy	Experiment	D
6.	RV	Others		Experiment	D
7.	RV	Data		Experiment	D
8.	RV	Others		Field	D

Table 8: Prove

	Class	Actor	Semantic prosody	Level	Location

1.	INF	Data		Experiment	I
2.	Verb Phrase (VP)	Information	may		I
3.	VP	Data	may		D
4.	VP	Data	may		D
5.	VP	Results	may		D
6.	VP	Results	might		D
7.	VP	Results	should		D
8.	VP	Future studies	may		D
9.	VP	Results	may		D

Table 9A: Report

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Participants		Field	I
2.	N	Others		Field	I
3.	RV	Authors		Field	I
4.	INF	Authors		Experiment	I
5.	N	Others		Field	I
6.	RV	Authors		Experiment	I
7.	N	Others		Field	I
8.	RV	Authors		Experiment	M
9.	RV	Authors		Experiment	M
10.	RV	Participants		Experiment	M
11.	N	Results		Experiment	M
12.	N	Others		Experiment	M
13.	N	Others		Experiment	M
14.	N	Results		Experiment	M
15.	N	Others		Field	R
16.	N	Footnotes		Field	D
17.	N	Others		Field	D
18.	RV	Authors	first to	Experiment	D
19.	RV	Participants		Experiment	D
20.	N	Others	not detailed enough	Field	D
21.	RV	Authors		Experiment	D
22.	N	Authors		Experiment	D
23.	N	Authors		Experiment	D
24.	RV	Authors		Experiment	D
25.	N	Authors		Experiment	D
26.	N	Others		Field	D
27.	N	Others		Experiment	D
28.	RV	Authors		Experiment	D
29.	RV	Others		Field	D
30.	N	Authors		Experiment	D
31.	RV	Authors		Experiment	D
32.	N	Authors		Experiment	D
33.	RV	Others		Experiment	D

Table 9B: Reported

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others		Field	I

2.	RV	Participants		Field	I
3.	RV	Results		Field	I
4.	RV	Others		Field	I
5.	RV	Others		Field	I
6.	RV	Authors		Experiment	M
7.	RV	Data		Field	M
8.	RV	Data		Experiment	M
9.	RV	Results		Experiment	M
10.	RV	Data		Experiment	R
11.	RV	Participants		Experiment	R
12.	RV	Others		Experiment	R
13.	RV	Data		Field	R
14.	Adjective (adj)	Others		Field	R
15.	RV	Others		Experiment	R
16.	RV	Participants		Experiment	R
17.	RV	Participants		Experiment	R
18.	RV	Others	recent	Field	R
19.	RV	Results	may suggest	Experiment	R
20.	RV	Participants			R
21.	RV	Others		Field	D
22.	RV	Others	however	Field	D
23.	RV	Results		Field	D
24.	RV	Others		Field	D
25.	RV	Others		Field	D
26.	RV	Results		Experiment	D
27.	RV	Others		Field	D
28.	RV	Others		Field	D
29.	RV	Others		Field	D
30.	RV	Authors	previously	Field	D
31.	RV	Results	best of our knowledge	Field	D
32.	RV	Others	very few	Field	D
33.	RV	Others	majority	Field	D

Table 9C: Reports

	Class	Actor	Semantic prosody	Level	Location
1.	N	Others		Field	I
2.	N	Others		Field	I
3.	N	Others		Field	I
4.	N	Others	previous	Field	I
5.	N	Others		Field	I
6.	N	Others		Experiment	M
7.	N	Others	in light of	Field	M
8.	N	Others	previous	Field	R
9.	N	Others	numerous	Field	R
10.	RV	Authors		Experiment	R
11.	N	Others		Field	R
12.	N	Others		Field	R
13.	N	Others		Field	D
14.	N	Others		Field	D

15.	N	Others		Field	D
16.	N	Others		Field	D
17.	N	Others	despite	Field	D
18.	N	Others		Field	D
19.	N	Others	controversies	Field	D
20.	N	Others	common to most	Field	D
21.	N	Others	previous	Field	D
22.	N	Others		Field	D
23.	N	Others	confirm previous	Field	D
24.	N	Others	in line with	Field	D
25.	RV	Others	similar ideas	Field	D
26.	N	Others	conversely	Field	D
27.	N	Others	comparable to other	Field	D
28.	N	Others	concordant	Field	D
29.	N	Others		Field	D
30.	N	Others		Field	D
31.	N	Authors		Experiment	D
32.	N	Others	or not	Field	D
33.	N	Others		Field	D

Table 10A: Say

	Class	Actor	Semantic prosody	Level	Location
1.	Idiom	Authors		Field	I
2.	Idiom	Authors		Experiment	M
3.	Idiom	Authors		Experiment	M
4.	RV	Authors		Experiment	M
5.	RV	Questionnaire		Experiment	M
6.	RV	Experiment		Experiment	R
7.	INF			Experiment	R
8.	INF			Experiment	R
9.	Idiom	Authors		Experiment	D
10.	RV	Authors		Experiment	D
11.	RV	Authors		Experiment	D
12.	Adj	Chinese people		Field	D

Table 10B: Said

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Participants		Experiment	R
2.	RV	Participants		Experiment	R
3.	RV	Participants		Experiment	R
4.	RV	Participants		Experiment	R
5.	RV	Participants		Experiment	R
6.	RV	Participants		Experiment	R
7.	RV	Participants		Experiment	R
8.	RV	Participants		Experiment	R
9.	RV	Participants		Experiment	R
10.	RV	Participants		Experiment	R
11.	RV	Participants		Experiment	R
12.	RV	Participants		Experiment	R
13.	RV	Participants		Experiment	R

14.	RV	Participants		Experiment	R
15.	RV	Participants		Experiment	R
16.	RV	Participants		Experiment	R
17.	RV	Participants		Experiment	R
18.	RV	Participants		Experiment	M
19.	RV	Participants		Experiment	D
20.	RV	Participants		Experiment	D
21.	RV	Participants		Experiment	D
22.	RV	Authors		Experiment	D
23.	RV	Authors		Experiment	D
24.	RV	Observers		Field	D
25.	RV	Authors		Field	D
26.	RV	Participants		Experiment	D
27.	RV	Participants		Experiment	D
28.	RV	Participants		Experiment	D

Table 11: Stress

	Class	Actor	Semantic prosody*	Level	Location
1.	N	Patients		Experiment	I
2.	N	Subjects		Experiment	I
3.	N	Subjects		Experiment	I
4.	N	Participants		Experiment	I
5.	N	Participants		Experiment	I
6.	N	Patients		Experiment	I
7.	N	Subjects		Experiment	I
8.	N	Participants		Experiment	I
9.	N	Subjects		Experiment	M
10.	N	Subjects		Experiment	M
11.	N	Patients		Experiment	R
12.	N	Others		Experiment	R
13.	N	Tests		Experiment	R
14.	N	Procedure		Experiment	R
15.	N	Patients		Experiment	R
16.	N	Subjects		Experiment	R
17.	N	Participants		Experiment	D
18.	N	Participants		Experiment	D
19.	N	Patients		Experiment	D
20.	N	Subjects		Experiment	D
21.	N	Participants		Experiment	D
22.	N	Subjects		Experiment	D
23.	N	Subjects		Experiment	D
24.	N	Patients		Experiment	D
25.	N	Participants		Experiment	D
26.	N	Participants		Experiment	D
27.	N	Patients		Experiment	D
28.	N	Subjects		Experiment	D
29.	N	Participants		Experiment	D
30.	N	Subjects		Experiment	D

31.	N	Subjects		Experiment	D
32.	N	Patients		Experiment	D
33.	N	Others		Experiment	D

*I would claim *stress* has a negative prosody as it is an ailment, but doctors may not view it as negative.

Table 12A: Suggest

	Class	Actor	Semantic prosody	Level	Location
1.	AV	Others	attempted	Field	I
2.	AV	Others		Field	I
3.	RV	Data	strongly	Experiment	I
4.	AV	Others	possibly	Experiment	I
5.	RV	Others	additional complications	Experiment	I
6.	AV	Others		Experiment	M
7.	RV	Results	so far	Experiment	R
8.	RV	Data	may	Experiment	R
9.	RV	Data	despite, significantly	Experiment	R
10.	RV	Others	also	Field	R
11.	RV	Results		Field	R
12.	RV	Results		Experiment	R
13.	RV	Results	tentatively/ could at least play a possible role	Experiment	D
14.	RV	Results	support	Experiment	D
15.	AV	Authors	could	Experiment	D
16.	RV	Others	but there is scant available data	Experiment	D
17.	AV	Results	may	Experiment	D
18.	RV	Data	may	Experiment	D
19.	RV	Data		Field	D
20.	RV	Others	seemingly	Experiment	D
21.	RV	Results	may not	Experiment	D
22.	RV	Participants	may	Experiment	D
23.	AV	Authors	might	Experiment	D
24.	RV	Others	are possible	Field	D
25.	RV	Results		Experiment	D
26.	RV	Others	can	Field	D
27.	RV	Authors	for now/ may	Experiment	D
28.	RV	Data	may in fact	Experiment	D
29.	RV	Results	correct	Experiment	D
30.	RV	Results		Experiment	D
31.	RV	Authors	in support	Experiment	D
32.	RV	Results		Field	D
33.	RV	Results	may	Field	D

Table 12B: Suggested

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Results	appears to have	Field	I
2.	RV	Others		Field	I
3.	RV	Data		Field	I
4.	RV	Others	possible	Field	I
5.	RV	Others	nonetheless, generalizing should	Field	I

	Class	Actor	Semantic prosody	Level	Location
			be cautious		
6.	RV	Others		Field	I
7.	RV	Others	however	Field	I
8.	AV	Authors		Experiment	M
9.	AV	Authors		Experiment	M
10.	RV	Results		Experiment	R
11.	RV	Results	similar to others	Experiment	R
12.	RV	Data	our, not	Experiment	R
13.	RV	Participant		Field	R
14.	RV	Authors	could not imagine	Experiment	D
15.	RV	Others		Field	D
16.	RV	Others	in accordance with, support	Field	D
17.	RV	Others	may	Field	D
18.	Participle			Field	D
19.	RV	Others		Field	D
20.	RV	Others	although	Field	D
21.	AV	Others	although	Field	D
22.	RV	Others	however	Field	D
23.	RV	Others		Field	D
24.	RV	Others	but	Field	D
25.	RV	Others		Field	D
26.	RV	Authors	might	Experiment	D
27.	RV	Others		Field	D
28.	RV	Others		Field	D
29.	RV	Others	may	Field	D
30.	RV	Data		Field	D
31.	RV	Others	similar	Field	D
32.	AV	Others		Field	D
33.	RV	Others		Field	D

Table 12C: Suggests

	Class	Actor	Semantic prosody	Level	Location
1.	RV	Others	common paradigm shift	field	I
2.	RV	Data		Field	I
3.	RV	Others		Field	I
4.	RV	Data		Field	I
5.	RV	Results	might	Experiment	R
6.	RV	Results		Experiment	R
7.	RV	Authors		Experiment	R
8.	RV	Results		Experiment	R
9.	RV	Results		Experiment	R
10.	RV	Data		Experiment	R
11.	RV	Results		Experiment	R
12.	RV	Data		Experiment	R
13.	RV	Others		Field	D
14.	RV	Results		Experiment	D
15.	RV	Others		Field	D
16.	RV	Data		Field	D
17.	RV	Others		Field	D
18.	RV	Data	may, could	Field	D
19.	RV	Results		Field	D
20.	RV	Results	might	Experiment	D
21.	RV	Data	may, however	Field	D
22.	RV	Results	despite	Experiment	D

23.	RV	Data		Field	D
24.	RV	Results	may or may not	Field	D
25.	RV	Data		Field	D
26.	RV	Result		Field	D
27.	RV	Results		Experiment	D
28.	RV	Data	but	Experiment	D
29.	RV	Data	may	Field	D
30.	RV	Data		Field	D
31.	RV	Results	very likely	Experiment	D
32.	RV	Results	may	Field	D
33.	RV	Data		Experiment	D

Table 13: Think

	Class	Actor	Semantic prosody	Level	Location
1.	INF	Others		Field	I
2.	RV	Participants		Field	I
3.	RV	Others		Field	I
4.	RV	Authors		Experiment	M
5.	RV	Participants		Experiment	M
6.	INF	Others		Field	M
7.	RV	Participants	differ	Experiment	M
8.	RV	Questionnaire		Experiment	M
9.	RV	Authors	unlikely	Experiment	R
10.	RV	Participants		Experiment	R
11.	RV	Participants		Experiment	R
12.	RV	Participants		Experiment	R
13.	RV	Participants		Experiment	R
14.	RV	Participants		Experiment	R
15.	RV	Participants		Experiment	R
16.	RV	Participants		Experiment	R
17.	RV	Participants		Experiment	R
18.	RV	Participants		Experiment	R
19.	RV	Participants		Experiment	R
20.	INF	Authors			D
21.	RV	Authors	highly unlikely	Experiment	D
22.	RV	Authors	may limit	Experiment	D
23.	RV	Authors	nevertheless	Experiment	D
24.	RV	Others	may	Experiment	D
25.	RV	Authors	could	Experiment	D
26.	RV	Others	although	Field	D
27.	RV	Others	may	Experiment	D
28.	RV	Others	nevertheless	Experiment	D
29.	RV	Others	however	Experiment	D
30.	RV	Authors	however	Experiment	D
31.	RV	Authors	may	Field	D
32.	RV	Questionnaire		Experiment	D
33.	RV	Participants	traditionally	Field	D