
**Multidimensional Genre-Based Discourse Analysis of a
Corpus of English-Language Medical Research Articles**

Daniel Lees Fryer

A Thesis Presented to the
Department of Literature, Area Studies, and European Languages,
Faculty of Humanities, University of Oslo, Norway,
in Partial Fulfillment of the Requirements for the
English Language Master of Arts Degree, Fall Semester 2007

Abstract

Genre analysis studies concerning the medical research article are limited, and the few studies that do exist tend to focus exclusively on the textual aspects of the genre, with little consideration for the context and discourse community in which texts are produced. The objective of this study is thus: 1) to analyze and describe the genre of the medical research article by emphasizing the written medical discourse (text) and the activity of medical research (context); 2) to compare study findings with those in the literature; 3) to investigate possible changes in the genre over time; and 4) to examine the potential pedagogic implications of genre research.

A multidimensional model is developed based on elements of register and genre analysis as well as on an examination of the medical discourse community. This model is applied to a study corpus comprising 17 medical research articles published between 2004 and 2006, selected from the *British Medical Journal*, the *Journal of the American Medical Association*, the *Lancet*, and the *New England Journal of Medicine*.

Ten rhetorical moves comprising a total of 28 steps are identified, and a series of within- and cross-section patterns are observed for specific lexicogrammatical features. Comparison with the literature shows that there is variation between studies of the same genre, which may be explained by a number of variables including the choice of study material, the methodology, the anticipated audience, and potential changes in generic structure. Genre as a descriptive framework for understanding and producing text appears to be a valuable pedagogic tool, provided students are made aware of the potential pitfalls of following such guidelines.

Key words: text, discourse, genre, register, discourse community, medical research article, systemic-functional linguistics

Acknowledgments

This study was conducted at the Department of Literature, Area Studies, and European Languages (ILOS), Faculty of Humanities, University of Oslo, Norway, and at the Department of Linguistics, College of Humanities and Social Sciences, Macquarie University, Sydney, Australia. Support provided by supervisors Hilde Hasselgård (Oslo) and David Hall (Sydney), as well as the encouragement and advice of Paul Cheung, Barbro Bredesen Opset, and David Woodhouse, is gratefully acknowledged.

Multidimensional Genre-Based Discourse Analysis of a Corpus of English-Language Medical Research Articles

Table of Contents

1 Introduction	1
1.1 Background.....	1
1.2 Aims and Scope.....	2
1.3 Study Outline.....	2
2 Analyzing Text and Context.....	5
2.1 The Language of Medical Science.....	5
2.2 Systemic-Functional Linguistics.....	6
2.2.1 Ideational Metafunction.....	6
2.2.2 Interpersonal Metafunction.....	12
2.2.3 Textual Metafunction.....	16
2.3 Register.....	22
2.4 Genre.....	27
2.4.1 English for Specific Purposes.....	27
2.4.2 The Sydney School: Systemic-Functional Genrists.....	31
2.4.3 New Rhetoricians.....	33
2.4.4 Summarizing Genre.....	34
2.4.5 Text Types.....	35
2.5 Previous Genre-Based Research Concerning Medical Writing.....	35
2.6 Current Guidelines for the Writing of Medical Research Articles.....	37
2.7 Register and Genre in this Study: A Multidimensional Approach.....	39
3 Material and Methods.....	41
3.1 Collection of Study Material.....	41
3.2 Description of Study Corpus.....	43
3.3 Methodology.....	45
3.3.1 Contextual Analysis.....	46
3.3.2 Textual Analysis.....	48
4 Contextualizing the Medical Research Article.....	51
4.1 Context of Situation.....	51
4.1.1 Field.....	51
4.1.2 Tenor.....	52
4.1.3 Mode.....	54
4.2 Medical Discourse Community.....	55
4.2.1 Common Public Goals.....	55
4.2.2 Mechanisms of Intercommunication.....	56
4.2.3 Information and Feedback.....	56
4.2.4 Genre Utilized.....	57
4.2.5 Specific Lexis.....	57
4.2.6 Membership.....	59
5 Textual Analysis of the Medical Research Article.....	61
5.1 Structural Moves Analysis.....	61
5.2 Analysis of Introduction Section.....	64
5.2.1 Move 1.....	64

5.2.2 Move 2.....	66
5.2.3 Move 3.....	68
5.2.4 Summary of Introduction Section.....	71
5.3 Analysis of Methods Section.....	72
5.3.1 Move 4.....	72
5.3.2 Move 5.....	81
5.3.3 Move 6.....	87
5.3.4 Additional Move.....	90
5.3.5 Summary of Methods Section.....	91
5.4 Analysis of Results Section.....	91
5.4.1 Move 7.....	92
5.4.2 Summary of Results Section.....	98
5.5 Analysis of Discussion Section.....	98
5.5.1 Move 8.....	98
5.5.2 Move 9.....	103
5.5.3 Move 10.....	104
5.5.4 Summary of Discussion Section.....	108
5.6 On the Use of Personals.....	109
6 Discussion of Study Results.....	111
6.1 Comparison with the Literature.....	111
6.1.1 Structural Moves Analysis.....	111
6.1.2 Lexicogrammar.....	115
6.1.3 Context.....	121
6.2 Implications.....	121
6.3 Limitations.....	124
6.4 Concluding Remarks.....	126
References.....	129
Appendix.....	137
Index.....	141
List of Tables and Figures	
Table 2.1. Processes, Participants, and Circumstances (adapted from Halliday & Matthiessen 2004: 260–263, tables 5(27) and 5(28)).....	9
Figure 2.1. Semiotic relationship of language, register, and genre (adapted from Martin & Rose 2003: 254, fig. 8.2)......	32
Table 2.2. Summarizing genre analysis.....	35
Table 3.1. Study article references and codes according to source journal.....	43
Table 3.2. Statistical description of the study corpus.....	44
Table 4.1. Summary of the context-of-situation analysis.....	55
Table 5.1. Frequency and stability of moves and steps identified in the corpus.....	62
Table 5.2. Moves summary for Introduction sections of corpus articles.....	72
Table 5.3. Moves summary for Methods sections of corpus articles.....	91
Table 5.4. Moves summary for Results sections of corpus articles.....	98
Table 5.5. Moves summary for Discussion sections of corpus articles.....	108
Table 5.6. Distribution of personals (<i>we/our</i>) in the study corpus.....	109
Figure 5.1. Distribution of personals in the study corpus.....	110
Table 6.1. Multi-source comparison of structural moves analyses of medical research articles.....	112

1 Introduction

1.1 Background

The analysis of distinct discourse types or genres provides valuable insight into how and why members of discourse communities communicate in the way they do (Bazerman 1988, Bhatia 2004, Swales 1990). Previous studies of register and genre have focused on the structural and grammatical features of texts (Biber 1986, Chafe & Danielewicz 1987, Swales 1981, 1990), as well as on their historical, socio-cultural, and ideological significance (Bazerman 1988, Bhatia 2004, Martin 2000, Miller 1984). The findings of some of these studies have provided the basis for descriptive frameworks that can be used in educational settings to assist learners in the interpretation and construction of genre-specific texts (e.g., Swales & Feak 2000). However, some researchers have warned against explicitly defining and teaching generic structures, stressing the importance of genre awareness over genre instruction and emphasizing the dynamic rather than static nature of genres (e.g., Bazerman 1988, Freedman 1994).

In terms of scientific discourse, much of the research concerning genre focuses on the reading and writing needs of non-native-speaking students at English-medium universities (see Holmes 1997, Hopkins & Dudley-Evans 1988, Kanoksilapatham 2005, Peng 1987). Due to the extensive use of English in medical education and medical research (Maher 1986, Swales 1990), there has been considerable interest in mapping and describing the generic characteristics of the medical research article. Some studies have looked at specific features of the genre (Adams Smith 1984, Gosden 1992, MacDonald 2002), while others have focused on parts of the medical research article in isolation (Bruce 1984, Dubois 1997). Only two papers have examined the medical research article as a whole (Nwogu 1997, Skelton 1994). However, neither of these studies considers the context and socio-cultural implications of the genre.

1.2 Aims and Scope

In order to address this apparent disparity between textual and contextual studies of genre, the primary objective of this study is to analyze and describe the structure and rhetorical function of the medical research article, using a combination of theoretical approaches. This includes register and genre analysis as well as a description of the medical discourse community in order to emphasize the relationship between written medical discourse (text) and the activity of medical research (context).

A secondary objective is to compare the study findings with those in the literature: 1) to assess the validity of the primary findings, 2) to evaluate whether a change in the genre has occurred over time, and 3) to examine the potential pedagogic implications of genre research.

1.3 Study Outline

This study is organized into six parts. The following section, section 2, begins by considering the nature of scientific and medical language, and what makes it distinct from other “styles” of language. The examples used to illustrate this form the basis for the ensuing discussion of the theoretical and practical means of describing and distinguishing different types of texts. Also included in this section is a presentation and discussion of the conflicting ideologies related to the application of genre research. A summary of previous studies concerning medical discourse and the medical research article as well as currently available guidelines for writers of medical research articles is also presented.

In section 3, the selection of study material is described, and details of the selected study articles and their source publications are provided. This is followed by a description of the methodology, in which the study’s multidimensional model of analysis, based on the theories and models of section 2, is presented. A rationale for this methodological framework is also provided.

Sections 4 and 5 present the results and interpretations of the contextual and textual analyses. The findings are presented according to the methodological procedures of the previous section, and functional interpretations are given along with examples from the study articles and other material related to the medical

discourse community. Tables and figures are used throughout to illustrate and summarize the main findings.

In the final section, section 6, the results of the previous sections are discussed in light of the available literature. Comparisons are made, and reasons for the observed similarities and differences are explored. The implications of these findings are then discussed in terms of their potential applicability to education, and possible study limitations and improvements are considered. The section concludes with a summary of the main findings and suggestions for further research.

In addition to these six sections, a list of references, an appendix, and a full index are provided at the end of the thesis.

2 Analyzing Text and Context

2.1 *The Language of Medical Science*

What *is* medical or indeed scientific language, and what, if anything, makes it distinct from other types of language? Consider examples (1)–(3) below. The first might be recognizable as scientific or medical English, but could the same be said of examples (2) and (3)? Probably not. There appears to be something in the vocabulary or the grammar that tells the reader that (1) is an example of medical English while the other two are not, even if all three share similarities in terms of their content.

- (1) In highly comminuted fractures where arthrosis commonly occurs, a primary surgical arthrodesis could increase the chance of a good patient outcome.
- (2) When you or a loved one has a minor injury, you want to help it heal as quickly as possible.
- (3) Jack fell down and broke his crown

All of the above examples have of course been removed from their original contexts, i.e., an article in a medical journal, an advertisement for bandages, and a children’s nursery rhyme.¹ Nevertheless, many readers, both native and non-native speakers of English, may be familiar with these different “styles of language,” and even if the specific excerpts or full texts from which they are taken are unfamiliar, they may still be recognized as belonging to a particular set of texts that have certain shared characteristics.² Recognition of such text types is important in understanding how a reader categorizes and, more importantly, derives meaning from texts (Halliday & Hasan 1985: 38–39).

The following subsections discuss how meaning is construed in language and how text types can be defined. The models described therein provide a

¹ (1) from Allmacher, D. H., Galles, K. S., & Marsh, J. L. (2006). Intra-articular calcaneal fractures treated nonoperatively and followed sequentially for 2 decades. *Journal of Orthopaedic Trauma* 20: 464–469; (2) from the Band-Aid website www.bandaid.com (accessed November 16, 2006); and (3) traditional, from the children’s nursery rhyme “Jack and Jill.”

² The term **text** refers to “any instance of language, in any medium, that makes sense to someone who knows the language” (Halliday & Matthiessen 2004: 3), although this paper is primarily concerned with text as an instance of *written* language.

theoretical framework for the analysis of medical research articles such as those of this study corpus. Each text can be analyzed and compared, and the similarities and differences revealed can be used to make predictions about other related texts.

These models and the subsequent multidimensional model proposed in this study provide an overall method of analysis that will be used to respond to the study objectives presented in section 1.

2.2 Systemic-Functional Linguistics

A useful methodological framework for studying language in context, whether it be the context of science, advertising, children's nursery rhymes, or indeed any other **context of situation**, is that offered by systemic-functional linguistics (SFL).³ SFL describes "language in use" rather than language as a "set of generalized rules detached from any particular context of use" (Thompson 2004: 1). Language and context are inextricably linked according to SFL theory, and such a model is therefore useful in identifying and describing text types.

The SFL approach divides language into three semantic functions, or **metafunctions**, which together represent meaning as it is construed in language.⁴ These metafunctions correspond to **ideational**, **interpersonal**, and **textual** meanings, and a brief description of each is provided below.⁵

2.2.1 Ideational Metafunction

The ideational metafunction relates to the way in which human experience is construed (or "language as reflection"; Halliday & Matthiessen 2004: 29–30), and is subdivided into **experiential** and **logical** meanings.

³ **Context of situation** is a term first used by Malinowski (1923, in Halliday & Hasan 1985: 6) to refer to the environment of a text. The term **context**, as it is used in this paper, refers to the situation or environment in which a given text is produced.

⁴ In SFL, **metafunction** is preferred to the more general term **function**. Function in linguistics is generally used to describe the "purpose or way of using a language," while metafunction is intended to emphasize that "functionality is intrinsic to language" and therefore an integral component in analyzing language (Halliday & Matthiessen 2004: 30–31).

⁵ A full description of these metafunctions is not possible within the scope of this study. The reader is referred to Halliday & Matthiessen (2004) for a more comprehensive overview.

2.2.1.1 *Experiential Meaning*

Experiential meanings encode experience, and include the way concrete and abstract events, things, and qualities are categorized. These meanings are represented in the clause by the functional constituents **Process**, **Participant**, and **Circumstance**, and can be summarized rather succinctly in terms of a question: “Who (Participant) does what (Process) to whom (Participant) under what circumstances (Circumstance)?” (adapted from Butt et al. 2000: 46).⁶

The Process, realized by a verb or verbal group, is pivotal in experiential meaning. In traditional English grammar, every clause requires a verb, and the same applies in the systemic-functional model of grammar. Every clause requires a Process, and the type of Process determines the meaning of the clause in which it appears. Participants, which are realized by nouns, nominal groups, and prepositional phrases, “interact” with the Process in a clause, and the roles of Participants reflect the Process type. These two functional constituents display an interdependency that Halliday & Matthiessen (2004: 176) term the “experiential centre of the clause.” Circumstances, on the other hand, which are realized by adverbial groups, prepositional phrases, and nominal groups, “augment this [experiential] centre [...] but their status is more peripheral, and unlike participants they are not directly involved in the process” (ibid.).

Processes can be classified into six types: material, mental, relational, behavioral, verbal, and existential. Material Processes typically construe the act of doing or happening. They include verbs such as *run*, *build*, and *fall*, and relate to the experience of outside-world events, as in *Jack fell down* or *Jill pushed Jack*. In contrast, mental Processes represent the inner experience of perception, consciousness, and emotion, and include verbs such as *sense*, *feel*, *think*, *want*, and *enjoy*, e.g., *Jack felt he was going to fall*.

Relational Processes “encode relationships of being and having between two Participants” (Butt et al. 2000: 51), and typically function to identify and characterize Participants. The most common relational Process is the verb *be*, but other identifying and attributive verbs such as *seem*, *indicate*, and *represent* serve a

⁶ Note that the terminology in this paper follows the conventions of Halliday & Matthiessen (2004). Terms written with an initial capital letter indicate function labels (e.g., Participant), while those written in lower-case letters (e.g., noun) are reserved for terms of more general linguistic terminology that refer to lexical and grammatical class.

similar **intensive** relational function, e.g., *Jack seems to be [\approx is] accident prone.*⁷ Relational Processes can also be considered in terms of their **possessive** and **circumstantial** properties. Possessive relational Processes are typically realized by the verb *have*, e.g., *Jack has a bruise on his head*, and indicate “ownership.” In circumstantial relational Processes, the relationship is one of “time, place, manner, cause, accompaniment, role, matter or angle” (Halliday & Matthiessen 2004: 240). In *The well was on top of the hill*, for example, the verb *was* (a circumstantial relational Process) functions to relate *The well* to a specific place.

Processes that construe physiological and psychological behavior are known as behavioral Processes, and can be represented by verbs such as *dance*, *smile*, and *dream*. These Processes can sometimes be difficult to identify, as they may also be interpreted as material or mental Processes. For example, *Jack cried when he broke his crown* could be interpreted as either behavioral (representing the physiological process of crying) or material (signifying the act of crying). Similarly, a behavioral Process like *dream* (e.g., *Jack dreamed he was falling*) could also be considered a mental Process (the “consciousness” of dreaming).

Verbs such as *speak*, *ask*, and *report* commonly represent verbal Processes. They “encode the experiences of bringing the inner world outside by speaking” (Butt et al. 2000: 51), and they project the words of what has been or is being said or written, e.g., *Jill said that Jack fell down*.

Existential Processes function to present or highlight the “existence” of a Participant. These Processes are most commonly represented by the verb *be* accompanied by existential *there*, e.g., *There was a boy called Jack*. In this example, as in existential clauses in general, the Process introduces a central Participant (*a boy called Jack*) whose role may be further emphasized in subsequent clauses, e.g., *who went up the hill to fetch a pail of water*.

Participant roles vary depending on the Process type and are determined by the nature of the Participant’s involvement with the Process. For example, the Participant in the material clause *Jack fell down* (Participant: Actor) is different to that of the existential clause *There was a boy called Jack* (Participant: Existent). Participant roles in relation to Process type are summarized below in Table 2.1.

⁷ Relational Processes that ascribe a particular attribute or quality to a Participant are known as “intensive.”

As mentioned above, in contrast to Participants, Circumstances occur freely in all types of clause and are not dependent on Process type. This can be illustrated by adding different Circumstances to the same Participant–Process combination, as in *Jack fell yesterday* (Circumstance: location: temporal) and *Suddenly Jack fell* (Circumstance: manner: quality). A list of Circumstance types is provided in Table 2.1.

Table 2.1. Processes, Participants, and Circumstances (adapted from Halliday & Matthiessen 2004: 260–263, tables 5(27) and 5(28))

Process type	Clause meaning	Participants, directly involved	Participants, indirectly involved	Circumstance type
material	“doing” “happening”	Actor, Goal	Recipient, Client; Scope; Initiator; Attribute	extent, location, manner, cause, contingency, accompaniment, role, matter, angle
mental	“sensing”	Senser, Phenomenon		
relational	“being” “having”	Carrier, Attribute Identified, Identifier; Token, Value	Attributor, Beneficiary, Assigner	
behavioral	“behaving”	Behaver	Behavior	
verbal	“saying”	Sayer, Target	Receiver; Verbiage	
existential	“existing”	Existent		

The functional constituents of experiential meaning can be illustrated more clearly by reconsidering examples (1)–(3). In the analyses below, the functions in parentheses refer to the words/phrases that follow them.⁸

- (1) (Circumstance: location: spatial) In highly comminuted fractures *where* (Participant: Actor) arthrosis (Process: material: happening) commonly occurs, (Participant: Actor) a primary surgical arthrodesis (Process: material: amount) could increase (Participant: Goal) the chance of a good patient outcome.
- (2) *When* (Participant: Carrier) you or a loved one (Process: relational: attributive) has (Participant: Attribute) a minor injury, (Participant: Senser) you (Process: mental: desiderative) want (Participant: Phenomenon) to help it heal (Circumstance: manner: quality) as quickly as possible.

⁸ The conjunctions *where* (1) and *When* (2) have no experiential meaning, and are considered part of the textual metafunction (see section 2.2.3).

- (3) (Participant: Actor) Jack (Process: material: motion) fell down / and (Process: material: make-up) broke (Participant: Goal) his crown

Each of the above contains one or more Processes and Participants. Examples (1) and (3) consist of material Processes, which indicate that the clauses are concerned with the experience of “doing” or “happening,” while (2) uses relational and mental Processes to characterize a physical state and to describe an emotional response or desire. The Participants in these three examples include people (in (2) and (3)), medical conditions and procedures (in (1) and (2)), and abstractions (in (1) and (2)). Examples (1) and (2) also include Circumstances, which provide further experiential information concerning “where” and “how” the Processes occur.

2.2.1.2 Logical Meaning

Logical meanings are encoded in clause complexes, in which “clusters” of clauses and their experiential meanings combine to form elaborate interclause messages. Clauses can be described in terms of their interdependency (**taxis**), as either **paratactic** (equal status) or **hypotactic** (unequal status). They can also be defined in terms of their logico-semantic relationship (Halliday & Matthiessen 2004: 376), as **expansion** or **projection**.

For example, in (3) below, the two clauses are paratactic, i.e., of equal status, and the second clause “expands” on the first. More specifically, in this instance, the expansion of the first clause by the second can be described as “enhancing” the information provided in the first clause by qualifying what happened after Jack fell. According to SFL, there are three ways of expanding a clause: by **enhancement** (see (3) below), by **extension** (see (4) below), and by **elaboration** (see (5) below).⁹

- | | | |
|-----|---|---------------------------|
| (3) | Jack fell down and broke his crown | (paratactic, enhancing) |
| (4) | Jack fell down, and Jill came tumbling after | (paratactic, extending) |
| (5) | Jack fell down; well, what did you expect? | (paratactic, elaborating) |

⁹ Triple vertical lines (|||) mark the beginning and end of clause complexes, and double vertical lines (||) separate individual clauses. Double brackets indicate the start ([[) and end (]] of embedded clauses (see below).

Examples of hypotaxis, in which the second clause is dependent on the dominant first clause, are illustrated in (6)–(8).

- (6) ||| Jack fell down, || because he slipped ||| (hypotactic, enhancing)
 (7) ||| Jack fell down, || while Jill continued up the hill ||| (hypotactic, extending)
 (8) ||| Jack fell down, || which is no surprise ||| (hypotactic, elaborating)

Projecting clauses are realized by verbal and mental Processes like *say*, *report*, *believe*, and *predict*. These clauses can be used for quoting and reporting, as for example in speech, which can be direct (paratactic clauses) or indirect (hypotactic clauses), as demonstrated in (9) and (10). An example of reporting using a mental Process is given in (11).

- (9) ||| Jill said, || “Jack fell” ||| (paratactic, projecting – quoting)
 (10) ||| Jill said || that Jack had fallen ||| (hypotactic, projecting – reporting)
 (11) ||| Jill thought || that Jack had fallen ||| (hypotactic, projecting – reporting)

As can be seen from the examples above, clauses are frequently bound together by conjunctions, and these can have either a coordinating (paratactic) or a subordinating (hypotactic) effect on the clause to which they belong. The choice of conjunction indicates the type of expansion or projection, and predicts the logical relationship between clauses. For example, in (6), the conjunction *because* introduces a reason for Jack’s falling, giving the clause a “causal: reason” function. If this conjunction were replaced with, for example, the conjunction *whenever*, the clause would still be enhancing but with a “temporal” function indicating the frequency or regularity with which Jack falls.

Clause positioning within clause complexes also has an important role in encoding logical meaning. Functions can be emphasized or de-emphasized by placing a specific clause at the start or end of a clause complex, provided the grammar permits it. This may have certain rhetorical implications. For example, in (12) below, the positioning of the conditional clause *If Jack fell* before the dominant clause *It was probably Jill’s fault* could be intended to emphasize suspicion concerning Jack’s fall (\approx “If indeed Jack did fall” or “Jack wouldn’t normally fall”). In (13), however, less emphasis on the conditional clause, by

positioning it after the dominant clause, may lend the clause complex a more accusatory tone, by implicating Jill in Jack's fall.¹⁰

- (12) ||| If Jack fell || it was probably Jill's fault |||
(13) ||| It was probably Jill's fault || if Jack fell |||

Texts (1)–(3) can thus be reinterpreted in terms of their logical meaning:

- (1) || (finite independent clause) In highly comminuted fractures [[(finite embedded clause) where arthrosis commonly occurs,]] a primary surgical arthrodesis could increase the chance of a good patient outcome. ||
(2) ||| (finite hypotactic enhancing clause, conditional) When you or a loved one has a minor injury, || (finite independent dominant clause) you want [[(non-finite embedded relative clause, purposive) to help it heal as quickly as possible.]] |||
(3) ||| (finite independent main clause) Jack fell down / || (paratactic enhancing clause) and broke his crown |||

Embedded clauses such as that in example (1) do not form clause complexes, since they function as clause constituents, and therefore cannot be categorized in terms of taxis or logico-semantic meaning (see Thompson 2004: 201). The clause complexes of texts (2) and (3), however, demonstrate hypotactic and paratactic enhancement, respectively. In (2), the conditional enhancement clause appears before the dominant clause, perhaps to emphasize the importance of the message in the enhancing clause.

2.2.2 *Interpersonal Metafunction*

In contrast to the ideational metafunction, the interpersonal metafunction represents “language as action” (Halliday & Matthiessen 2004: 30) and the way human relationships are enacted. In written language, interpersonal meanings refer to the “interaction” between writer and reader, and include the description of **mood** and **modality**.

In the English language, there are three basic types of interpersonal clause structure—the **declarative**, the **interrogative**, and the **imperative**—collectively

¹⁰ See also section 2.2.3 regarding the relevance of “thematic” and “rhematic” dependent clauses.

known as mood.¹¹ The declarative and interrogative can be identified according to the **Subject** and **Finite** operator of a clause. In (3), for example, the mood element of the first clause is *Jack fell* (Subject ^ Finite), while the rest of the clause, in this case *down*, is termed the **Residue**.¹² In this particular example, the Subject–Finite combination, or “Mood block” of the clause (Butt et al. 2000: 91), functions to provide the reader with information, i.e., that Jack fell.¹³ The same applies to the second clause in (3), i.e., that Jack broke something. In the latter clause, however, the Subject has been omitted, but can be readily determined from the previous clause (added below for clarity).¹⁴ Note that the Residue—*down* (Adjunct) in the first clause, *his crown* (Complement) in the second—has no effect on the mood (or Mood) of either clause; both are declarative.

- (3) Jack fell down and [Jack/he] broke his crown (Mood: providing info, declarative)

Clauses that request information can be considered in terms of two categories. One, known as the **yes/no interrogative**, demands a polarized response, as in (14); the other, the **WH- interrogative**, seeks content-based information, as in (15).¹⁵ The yes/no interrogative can also be used to offer goods and services (G&S), as in (16).

- (14) Did Jack fall down? (Mood: requesting info, yes/no interrogative)
 (15) Why did Jack fall? (Mood: requesting info, WH- interrogative)
 (16) Shall I fetch some water for you? (Mood: offering G&S, yes/no interrogative)

The distinction between declarative and interrogative clauses lies in the ordering of the Subject and Finite operator. In (3), the Subject appears before the Finite (*Jack + fell*, [*Jack*] + *broke*), while in (14), (15), and (16) the order is

¹¹ Another less frequently used clause type, which is not discussed further in this study, is the **exclamative** (see, for example, Thompson 2004: 55–56).

¹² The Residue consists of one or a combination of three functional elements: **Predicator**, **Complement**, and **Adjunct**. The Predicator represents the rest of the verbal group in a clause; the Complement, any additional nominal groups; and the Adjunct, adverbial groups, nominal groups, and prepositional phrases that would otherwise represent Circumstance in the experiential meaning of a clause (see section 2.2.1.1). A caret (^) denotes “is followed by.”

¹³ Note that mood (lower-case “m”) refers to the choice of clause type, while Mood (upper-case “M”) refers to the combined function of the Subject and Finite.

¹⁴ See also section 2.2.3 regarding “ellipsis.”

¹⁵ Note that these are idealized responses. Actual responses to yes/no interrogatives, for example, do not necessarily have to be polarized, e.g., “Did Jack fall down?” “I don’t know.”

reversed, i.e., Finite ^ Subject.¹⁶ However, there are exceptions. For example, in WH- interrogative clauses in which the WH- element functions as Subject, the Mood block is Subject ^ Finite rather than Finite ^ Subject, e.g., (Subject) *Who* (Finite) *pushed* (Residue: Complement) *Jack?*. This can also apply to yes/no interrogatives, e.g. (Subject) *Jack* (Finite) *fell* (Residue: Adjunct) *down?*. In such instances, the yes/no interrogative functions in a similar way to the declarative, in that it may represent a statement of exclamation (e.g., disbelief, as in “I don’t believe you!”) as well as a request for confirmation (e.g., “is it true?”).¹⁷ Similarly, the typical Mood block of declaratives (Subject ^ Finite) can also be reversed, as in the second clause in “*It wasn’t me,*” (Finite) *protested* (Subject) *Jill* (Residue: Adjunct) *furiously*.

The declarative and interrogative represent an exchange of information collectively known as the **indicative**, but information and G&S can also be “demanded” by using the imperative. The imperative differs from the indicative in that the clause does not necessarily require a Mood block. In (17), for example, there is no Mood block, only a Predicator, a Complement, and an Adjunct. Examples including parts of the Mood block (Subject or Finite) or the entire Mood block (Subject + Finite) are provided in (18)–(20).

- (17) (Predicator) Fetch (Complement) some water (Adjunct) now!
- (18) (Subject) You (Predicator) be (Complement) careful!
- (19) (Finite) Don’t (Predicator) fall!
- (20) (Finite) Don’t (Subject) you (Predicator) go and (Predicator) fall!

In the same way as the Process is pivotal in experiential meaning, so too is the Finite operator in interpersonal meaning. In addition to the role of the Finite in determining the mood of a clause (see above), it can also encode speaker/writer opinion by expressing modality. Verbal operators such as *can*, *may*, *could*, *should*, and *must* are known as **modal Finites**, and function to express probability or obligation (see examples below). They may also reflect varying degrees of

¹⁶ In (15), *Why* is classed as part of the Residue (a WH- Adjunct). The Mood block is hence the same as that of (14), i.e., *did* + *Jack*.

¹⁷ This mood is also termed the **queclarative** (Thompson 2004: 80) to illustrate the interrogative–declarative function of the clause. It shares the same standard Mood block as the declarative (Subject ^ Finite).

probability or obligation. For example, in (23) a lower degree of obligation is expressed than that in (24).

- | | | |
|------|---------------------------------------|---------------|
| (21) | Jack <u>could</u> have fallen. | (probability) |
| (22) | Jill <u>ought to</u> have warned him. | (obligation) |
| (23) | Jack <u>should</u> wear better shoes. | (obligation) |
| (24) | Jack <u>must</u> wear better shoes. | (obligation) |

However, it is not only the Finite that serves to express interpersonal meanings such as speaker/writer comment. Certain Adjuncts can be used to similar effect. So-called **mood Adjuncts** express varying degrees of probability and habituality, as well as aspects of intensity and counterexpectancy, and are typically realized by adverbs. However, unlike other types of Adjunct, they directly modify the Mood block, and are therefore not considered part of the Residue. In the examples provided below (Mood block italicized, mood Adjunct underlined), the mood Adjunct is typically positioned just before or just after the Finite operator; although there are exceptions, as in (29) and (30), in which the mood Adjunct is either emphasized by placing it before the Subject or added as an afterthought at the end of the clause.

- | | | |
|------|--|-----------------------------|
| (25) | <i>Jack <u>usually</u> falls</i> when he fetches water | (habituality) |
| (26) | <i>He'll <u>probably</u> fall</i> this time, too | (probability) |
| (27) | <i>He <u>completely</u> ignored</i> Jill's warning | (intensity/degree) |
| (28) | <i>He <u>simply</u> didn't</i> listen | (counterexpectancy) |
| (29) | <u>Usually</u> <i>Jack falls</i> when he fetches water | (habituality, emphasis) |
| (30) | <i>Jack falls</i> when he fetches water <u>usually</u> | (habituality, afterthought) |

Comment Adjuncts encode similar meanings to mood Adjuncts, and are also typically realized by adverbs. They can be asseverative (e.g., *obviously*), qualificative (e.g., *allegedly*), persuasive (e.g., *truly*), and factual (e.g., *actually*), and they can express wisdom and morality (e.g., *cleverly*, *wrongly*), and validity and personal engagement (e.g., *generally*, *personally*). However, unlike mood Adjuncts, comment Adjuncts are less closely involved with the Mood block. They

tend to comment on the whole clause, and are generally classed as part of the Residue.¹⁸ Some examples are given below.

- | | | |
|------|--|--------------------|
| (31) | <u>Clearly</u> it wasn't Jill's fault | (asseverative) |
| (32) | You don't <u>seriously</u> believe that, do you? | (persuasive) |
| (33) | <u>Frankly</u> it was all Jack's own doing | (pers. engagement) |

Texts (1)–(3) can now be described in terms of their interpersonal meaning:

- (1) (Adjunct) In highly comminuted fractures where (Subject) arthrosis (mood Adjunct: habituality) commonly (Finite) occurs, (Subject) a primary surgical arthrodesis (modal Finite) could (Predicator) increase (Complement) the chance of a good patient outcome.
- (2) When (Subject) you or a loved one (Finite) has (Complement) a minor injury, (Subject) you (Finite) want (Predicator) to help (Complement) it (Predicator) heal (Adjunct) as quickly as possible.
- (3) (Subject) Jack (Finite) fell (Adjunct) down and (Finite) broke (Complement) his crown

All three examples are declarative, i.e., they function to provide information. In the first example, modality is expressed by a mood Adjunct and a modal Finite; the adverb *commonly* indicates usuality, and the verb *could* expresses probability. The second example contains no obvious modal elements. However, the Mood blocks contain reference to the reader (the Subject *you*), and therefore signal an attempt by the writer to engage the reader at a more personal level than in examples (1) and (3). The interpersonal analysis of (3) is limited in terms of mood and modality, but contextual factors such as intended audience and participant roles provide further insight into the interpersonal meaning encoded in this and other examples. These issues are explored in section 2.3.

2.2.3 Textual Metafunction

While the ideational and interpersonal metafunctions are both “extra-linguistic phenomena [concerned with] the social and the natural world” (Matthiessen & Halliday 1997: unnumbered), the textual metafunction is concerned with the

¹⁸ Certain comment Adjuncts may also be classed in terms of their experiential meaning, as Circumstances (see discussion in Halliday & Matthiessen 2004: 129–132).

realization of ideational and interpersonal meanings. It is the presentation of human experience and human relationships as text (written or spoken) in a specific context, and relates to information and how the flow of information is presented using resources such as **thematic structure** and **cohesion**.

As discussed in section 2.2.1.2, the positioning of clauses within a clause complex can have an effect on the ideational meaning of the clause complex. Similarly, the positioning of elements *within* a clause plays an important role in encoding the clause's textual message. The structure of this message can be described in terms of **Theme** and **Rheme**. The Theme "serves as the point of departure of the message... [and] locates and orientates the clause within its context" (Halliday & Matthiessen 2004: 64), while the remainder of the clause, the part which develops the Theme, is known as the Rheme. The Theme occupies first position in a clause and comprises "everything up to and including the Participant that functions as the Subject of the clause" (Martin & Rose 2003: 177).¹⁹ For example, the Theme in the first clause of example (3) is *Jack*. In the second clause, the Subject is left out, but the Theme essentially remains the same: *and [Jack]*. In both clauses, the Subject, *Jack*, is described as the topical or **experiential Theme**, since *Jack* as Participant represents the experiential component of the Theme (see section 2.2.1.1 above). In contrast, the conjunction *and* functions to connect the message of the second clause with that of the first, and is known as the **textual Theme**.²⁰ Conjunctive Adjuncts (see discussion below) and continuatives may also function as textual Themes, as demonstrated by the underlined elements in (34) and (35), respectively.

- (3) Jack fell down and [Jack] broke his crown
 (34) However, Jill came tumbling after
 (35) Jack fell down; well, what did you expect?

¹⁹ Theme can also be defined as everything up to and including the "first experiential element" in a clause (e.g., Butt et al. 2000: 135; Halliday & Matthiessen 2004: 66–67, 79; Thompson 2004: 142–143), a definition that does not necessarily include the Subject. Martin & Rose's (2003) definition, however, is used throughout this study, for reasons that will be made more apparent in the discussion of "hyperTheme" below.

²⁰ As discussed in section 2.2.1.2, conjunctions are used to indicate the type of logical relationship between clauses. However, using the term "logical Theme" to describe a conjunction in Theme position may be misleading, since logical meaning refers to the interrelationship of clauses rather than thematic structure within the clause.

A third type of Theme, the **interpersonal Theme**, refers to a thematic element with interpersonal meaning. This is typically realized by a vocative or **modal Adjunct** (collective term for mood and comment Adjuncts; see section 2.2.2 above), e.g., *Jack, you're going to fall* (vocative) or *Unsurprisingly, Jack fell* (comment Adjunct). If present, interpersonal and textual Themes always precede the experiential Theme, and the resulting “thematic complex” is known as a **multiple Theme** (see example below).

- (36) (textual Theme, continuative) Oh (interpersonal Theme, vocative) Jack, (interpersonal Theme, comment Adjunct) of course (experiential Theme: Participant, Actor) you always fall!

In English, the experiential Theme of a declarative clause is typically the Subject, as in the above example (the Actor *you*). However, experiential Themes may occasionally be “marked” by placing an adverbial group, prepositional phrase, or nominal group that is not the Subject of the clause in first position. For example, in *Yesterday Jack fell*, the Theme is marked by a temporal Circumstance, *Yesterday*. By placing the Circumstance in front of the Participant (known as “fronting” or “foregrounding”), the aspect of time becomes the point of departure for the clause (compare with *Jack fell yesterday*). Foregrounding thus allows the speaker/writer to emphasize certain elements of the clause, which can then be developed in subsequent clauses. In the above example, the reason for foregrounding *Yesterday* may become more apparent if the following clause is added: *and now he's in the hospital*. The Theme of this clause is also marked and, like the previous clause, emphasizes an aspect of time (the temporal Circumstance *now*). Together, these marked Themes signal a sequence of events (“*first* this happened, *then* that happened”) that is characteristic of narrative, procedural, and explanatory texts. Indeed, such atypical or **marked Themes** “are often used to signal new phases in a discourse: a new setting in time, or a shift in major participants; that is, they function to scaffold discontinuity” (Martin & Rose 2003: 179) and may thus be considered indicators of **thematic progression**.

Martin & Rose’s (2003: 181–182) notion of **hyperTheme** takes the concept of thematic structure to a level beyond the clause. A hyperTheme defines or summarizes a series of related Themes, and its relation to a text is much like the

relation of a Theme to a clause. Martin (1993a: 245) describes hyperTheme as a “paragraph constituent” commonly realized by a clause (or combination of clauses), but hyperThemes may also be paraphrased by the body of a paragraph (Martin & Rose 2003: 183) or, in some instances, by a single summarizing Theme (Daneš 1974: 201–202). For example, in the nursery rhyme “Jack and Jill” (see below), the experiential Themes are *Jack and Jill*, *Jack*, [*Jack*], and *Jill*. The hyperTheme, the point of departure for this particular text, can thus be summarized as “Jack and Jill.”²¹ Similarly, in the opening paragraph of section 2.2.2 (see (38) below), each Theme, although marked with a Circumstance, includes a Participant that makes reference to “the interpersonal.” The opening Circumstance, a contrastive prepositional phrase, highlights a change in Theme, from the previous section of “discussing the ideational metafunction” to the current section of “discussing the interpersonal metafunction.” In other words, a change in Theme, marked here by a Circumstance, serves as the point of departure for a new phase in the text.²² HyperThemes therefore represent a useful tool for studying thematic progression and how a text develops.

- (37) Jack and Jill went up the hill
 To fetch a pail of water.
Jack fell down and [Jack] broke his crown
And Jill came tumbling after.
- (38) In contrast to the ideational metafunction, the interpersonal metafunction represents “language as action” (Halliday & Matthiessen 2004: 30) and the way human relationships are enacted. In written language, interpersonal meanings refer to the “interaction” between writer and reader, and include the description of mood and modality.

The way a text develops also depends on the way in which clauses and the meanings they construe are connected. These connections are referred to as **cohesive devices**, and include resources such as **conjunction**, **reference**, and **ellipsis**.

²¹ Daneš (1974: 201) refers to this type of thematic progression as “continuous” or “constant.”

²² Note that if the “traditional” definition of Theme were used, only the Circumstances *In contrast to the ideational metafunction* and *In written language* would be recognized as Themes, thus making it difficult to identify the hyperTheme “interpersonal metafunction.”

As illustrated in section 2.2.1.2, conjunctions indicate the status of clause relations. Textually, they also function to guide the development of a text by connecting sentence and clause-complex messages. For example, in (39), the causal conjunctive Adjunct *As a result* refers to both Jack's slipping and his falling as the reason for breaking his crown. Similarly, in a more complex example taken from a cookbook (see (40), from Halliday & Matthiessen 2004: 545), the temporal conjunctive Adjunct *Meanwhile* marks a relationship between *mix the flour with the cayenne pepper in a bowl* and a series of preceding clauses, and indicates a shift in emphasis from *the aubergine slices* to another aspect of the recipe, presumably the preparation of a marinade.

- (39) Jack slipped and fell down. As a result, he broke his crown.
- (40) Place the aubergine slices in a colander, sprinkle with salt and leave to drain for 10 minutes. Rinse and dry thoroughly. Meanwhile, mix the flour with the cayenne pepper in a bowl.

While conjunction connects clause complexes, individual elements within and across clauses can be linked by referencing. For example, in (39), the personal pronoun *he* makes **anaphoric** reference to *Jack* in the previous sentence. Likewise, the possessive determiner *his* refers anaphorically to both *he* in the same clause and *Jack* in the previous sentence, and such links are termed **chains of reference**. Example (40) contains a different type of reference which uses the definite article instead of personals to link elements in the text. By using the definite form in the nominal group *the aubergine slices*, anaphoric reference is being made to a previous referent somewhere in the text. This could be in a prior clause (e.g., *Slice up two aubergines*), in a list of ingredients provided at the start of the recipe, or possibly as a visual referent in the form of a photograph. The same applies in the case of *the flour* and *the cayenne pepper*, while the indefinite articles (e.g., *a colander*) or the zero article (i.e., *salt*), which have no specific referencing function, are used to indicate non-specific, generalized elements. For example, *a colander* implies that “any colander will do.”

Examples (39) and (40) can also be used to demonstrate ellipsis, a resource that allows the omission of certain clauses or clause elements “when they can be presumed from what has gone before” (Halliday & Matthiessen 2004: 535). In (39),

the Subject of the second clause of the first sentence (*Jack*) is left out. The same can also be seen in example (40), in which the Goal of the first clause (*the aubergine slices*) is missing from subsequent clauses. These omissions are economical but still allow the reader to reconstruct the essential missing elements.

A similar way of avoiding the direct repetition of previous clauses and clause elements is a form of pseudo-ellipsis known as **substitution**. Rather than simply omit a clause or clause element, substitution allows previous elements to be replaced by a “linguistic token” (Thompson 2004: 180). In (41), for example, the so-called “pro-verb” *do* is used to represent the Rheme of the previous clause, i.e., *Jill did too = Jill fell down too*. Another commonly encountered linguistic token is the “pro-form” *so*. In example (42), this substitutive element implies one of a number of alternatives: *Jack did fall, Jack fell, he fell, he did*, etc. In other words, it substitutes an entire clause.

- (41) Jack fell down. Jill did too.
 (42) Did Jack fall? Yes, I think so.

In terms of their textual meaning, texts (1)–(3) can be interpreted as follows:

- (1) In highly comminuted fractures where arthrosis commonly occurs, a primary surgical arthrodesis could increase the chance of a good patient outcome.

Theme: marked experiential Theme (Circumstance *In highly comminuted fractures where arthrosis commonly occurs*) + experiential Theme (Participant *a primary surgical arthrodesis*)

Cohesion: no cohesive devices used

- (2) When you or a loved one has a minor injury, you want to help it heal as quickly as possible.

Theme: thematic dependent clause (conditional *When you or a loved one has a minor injury*) + experiential Theme (Participant *you*)²³

²³ When a dependent clause precedes a dominant clause, it acts as the point of departure for the entire clause complex. In other words, the preceding dependent clause can be treated as a constituent of the dominant clause (together known as a “T-unit”; Fries 1994: 229) and can thus be regarded as

Reference: **exophoric** *you* (text-external reference to reader), anaphoric *it* (= *a minor injury*)

(3) Jack fell down / and broke his crown

Theme: experiential Theme (Participant *Jack*), textual Theme (conjunction *and*)

Ellipsis: Subject [*Jack*] in second clause; reference: anaphoric *his*

The Theme of example (1) is marked by a Circumstance which appears before the Subject of the main clause. The reason for this foregrounding depends on the text preceding and following the excerpt, but it may be intended to highlight the importance of a particular medical condition or to signal a new phase in the text. Example (2) is also marked; in this case by a dependent clause, perhaps to emphasize the conditional aspect of the clause complex. In contrast, the Themes of (3) are unmarked, but the clauses themselves contain some interesting cohesive devices. For example, elision of the Subject in the second clause retains the seven-syllable pattern of the first three lines of the nursery rhyme (see (37) above for entire text), and there is anaphoric reference to the Subject of the first clause by the possessive determiner *his* in the second. Example (2) also contains anaphoric reference—*It* refers to the Attribute of the previous clause, the nominal group *a minor injury*—as well as direct exophoric reference to the reader (*you*).

HyperThemes cannot be readily identified in such short text excerpts. Longer samples are required to observe common Themes and thematic changes and, as the following section demonstrates, knowledge of the context in which a text is produced provides further insight into the function and meaning of a text.

2.3 Register

According to Matthiessen & Halliday (1997: unnumbered), “language is embedded in a context of culture or social system, and any instantiation of language as text is embedded in its own context of situation.” Perhaps unsurprisingly, then, SFL divides context into three categories or “dimensions of variation” (Thompson 2004:

thematic (Halliday & Matthiessen 2004: 392–393; Thompson 2004: 154–156). It follows, then, that a dependent clause following its dominant clause is given rhematic status.

40) that correspond to the three metafunctions of language. These are **field** (ideational metafunction), **tenor** (interpersonal metafunction), and **mode** (textual metafunction) (Halliday & Hasan 1985: 12), and together they describe a text's **register**, the functional variety of a language that evolves in a given context (Matthiessen & Halliday 1997: unnumbered).

Field describes the social action, the type of activity taking place, its content and ideas. Tenor refers to the participants; that is, who they are, what roles they have, and what the relationships are between those roles. The third dimension of variation, mode, is concerned with “symbolic organization” (Halliday & Martin 1993: 33): the role the language is playing, participant expectations regarding the language, and the channel of communication.

These three variables—field, tenor, and mode—can be illustrated by re-examining examples (1)–(3) and the contexts in which they were produced.

- (1) In highly comminuted fractures where arthrosis commonly occurs, a primary surgical arthrodesis could increase the chance of a good patient outcome

Field (ideational): content—“treating physical injury”; relatively technical use of language; Processes: material; Participants: Actor (*a primary surgical arthrodesis and arthrosis*), Goal (*the chance of a good patient outcome*); Circumstance: locational prepositional phrase (*In highly comminuted fractures...*)

Tenor (interpersonal): medical researchers addressing group of fellow medical researchers, both groups with an interest in the treatment of physical injury; declarative mood; modal Finite (*could*); habitual mood Adjunct (*commonly*)

Mode (textual): written text; monologue; published in an academic journal; relatively limited readership; marked experiential Theme

Register: “medical writing”

- (2) When you or a loved one has a minor injury, you want to help it heal as quickly as possible

Field: content—“treating physical injury” and “care and responsibility”; relatively non-technical use of language; Processes: relational and mental; Participants: Carrier (*you or a loved one*), Attribute (*a minor injury*), Senser (*you*), Phenomenon (*to help it heal*); Circumstance: adverbial group (*as quickly as possible*)

Tenor: manufacturer to consumers (and wholesalers); relatively informal use of language; declarative mood

Mode: written text; published online, open readership; monologue (but uses second-person pronoun to include reader); marked experiential Theme; anaphoric (*it = a minor injury*) and exophoric reference (*you = reader*)

Register: “advertising”

- (3) Jack fell down and broke his crown

Field: content—“describing an accident”; rhyme; non-technical use of language; Processes: material; Participants: Actor (*Jack*), Goal (*his crown*); Circumstance: adverb (*down*)

Tenor: original author unknown; writer to (adult/child) reader, adult reader to child hearer, adult reciting to child hearer, child reciting, etc.; declarative mood

Mode: written text, which can either be read or performed; published in children’s book of rhyme; open readership, primarily aimed at children; unmarked Themes (second Theme elided)

Register: “nursery rhyme”

Register analysis of this type is perhaps best suited to larger texts, which allow a more precise description of field, tenor, and mode, particularly when the original context is given or more apparent. However, even in these brief analyses, the similarities and differences in register illustrate how a reader might differentiate between three related, yet distinctly different text types. All three texts describe a form of physical injury, perhaps even the same type of injury, and all three are examples of written English. However, the contexts in which they are produced,

their anticipated audiences, and their **lexicogrammar** (i.e., the vocabulary and the grammar; Halliday & Matthiessen 2004: 7) are very different.

It might seem obvious, perhaps even superfluous, to describe (1), originally printed in an academic medical journal, as an example of medical writing—just as it might to define (2) as an advertisement simply because it appears on a corporate website and (3) as a nursery rhyme because it is published in a book of children’s nursery rhymes. However, as Halliday & Hasan (1985: 38) note, “we reconstruct from the text certain aspects of the situation, certain features of the field, of the tenor, of the mode” that might be familiar to us. Therefore, the context, without necessarily being given, may be construed from even short, single-sentence excerpts like those above. It might not, for example, be possible to ascertain that (2) was originally posted on a corporate website, but other aspects of the field, tenor, and mode may indicate to the reader that the text advertises a certain product or service, particularly if the following sentence is included: “Trust Band-Aid.” Similarly, in the case of (1), a reader may not be able to claim with certainty that the text is from a medical journal (it could be from a textbook, with a different intended readership), and there is no indication of whether it was first published in print or online. Indeed, it may not even be an example of medical writing—the mode could be spoken in the setting of a lecture or conference, both with different intended audiences (tenor)—so that without knowing the context, “medical communication” or “medical language” might be a more appropriate label than “medical writing.” Likewise, example (3) may be interpreted in a number of different ways. As demonstrated in the brief register analysis above, there are several plausible interpretations of the tenor, and the mode could just as easily be spoken as written. Indeed, if the nursery rhyme from which the text is taken is unfamiliar to the reader, (3) could feasibly be interpreted as the lyrics of a pop song or the recount of an accident that unintentionally rhymes.

Register analysis tends to focus on longer texts than the examples given here, and preferably on a large corpus of texts (e.g., Biber 1986, Chafe & Danielewicz 1987, Ghadessy 1988). The material is typically analyzed by studying the occurrence (or absence) of certain linguistic features. In large study corpora, patterns may emerge that allow researchers to make generalizations about certain registers. For instance, Chafe & Danielewicz (1987) illustrate how certain features of academic writing differ from other written and oral registers. From the analysis

of their corpus, they conclude, for example, that academic writing is generally “more detached” than, say, the language of a university lecture, which they demonstrate in terms of academic writing’s more extensive use of the passive construction and fewer first-person pronouns.

The generalizations derived from register analysis can in turn be used for pedagogic purposes. For example, in Chafe & Danielewicz’s (1987) above-mentioned study, recurring or absent features are used to describe the linguistic conventions of academic writing in contrast to other registers. These conventions, e.g., avoiding the use of first-person pronouns and a predilection for passive constructions, can be used to make guidelines to aid those who may be unfamiliar with the language and linguistic norms of academic writing.

An interesting example of this can be found at graduate-level academia. Most universities or university departments now offer their students courses on how to write “appropriate” academic papers.²⁴ Some of these courses or guidelines are produced anecdotally by university lecturers, based on their own experience of academic writing and general faculty/departmental requirements. Others are developed from the kinds of studies and results offered by register analysis. More specifically, register analysis (and genre analysis; see section 2.4 below) has been applied at an elementary level of education, most notably in the New South Wales Disadvantaged Schools Program, where systemic-functional linguists at the University of Sydney, Australia, used register analysis to provide local schoolteachers with the pedagogic aids necessary to help elementary-school pupils decipher, edit, and construct their own narrative and descriptive texts (see, for example, Christie 1992, Martin 2000).

In addition to providing pedagogic aids, register analysis can be used at a more abstract level, to study how particular text types are constructed. Such studies, which often have a diachronic perspective, attempt to understand how writers and readers of particular registers construe meaning, by studying the relationship between texts and the social practices they realize (Halliday & Martin 1993: 22–23; see also Couture 1993, Halliday 1988). While they may not offer explicit guidelines

²⁴ See, for example, academic English programs at the University of Oxford (URL: www.lang.ox.ac.uk/eas) and Macquarie University (URL: www.nceltr.mq.edu.au/elp/programs/acadEnglish.html), as well as subject-specific writing programs such as those offered by the University of Oxford’s Department of Medical Sciences (URL: www.medsci.ox.ac.uk/portal/skillstraining/courses/writingskills). All URLs accessed February 5, 2007.

for learners, their research questions and hypotheses are usually informed by the needs of the learner, and, in the case of Halliday & Martin's (1993) studies of scientific language, learner difficulties are highlighted and discussed from a sociolinguistic perspective in terms of issues such as grammatical metaphor, technical taxonomy, and lexical density (Halliday 1993b: 71).

2.4 Genre

Another useful method for the description of different text types is **genre** analysis. Genre is traditionally defined as a category of artistic composition “characterized by similarities in form, style, or subject matter” (Oxford American Dictionary 2006), but its use in literary, rhetoric, and linguistic studies typically goes beyond this somewhat limited definition. Indeed, within the field of linguistics, the analysis of genre has developed so rapidly in recent years, partly due to the wider availability of large electronic corpora and the advent of powerful data-analysis software, that scholars have been prompted to describe the discipline in terms of three distinct, yet overlapping schools of thought (Hyland 2002, Hyon 1996, Yunnik 1997). Each of these groups recognizes genre as a social phenomenon and that specific genres have certain shared characteristics, conventions, and constraints in terms of their language, purpose, and intended audience. However, their differing approaches to genre analysis, and particularly to its practical application, deserve further mention.

2.4.1 English for Specific Purposes

Some of the most extensive and pioneering work in genre analysis research has been carried out by the so-called English for Specific Purposes (ESP) group. As the name suggests, the group's main research interest is of a pedagogic nature, and much of its focus is on secondary and tertiary levels of education, as well as on professional needs such as business English. Consequently, ESP research provides practical guidelines “to help non-native [and native] speakers of English master the

functions and linguistic conventions of texts that they need to read and write in their disciplines and professions” (Hyon 1996: 698).²⁵

Swales, an ESP researcher whose work on genre analysis has been particularly influential, defines genre as “a class of communicative events, the members of which share some set of communicative purposes” (Swales 1990: 58). These genre-specific communicative purposes are recognized and defined, either explicitly or implicitly, by members of the “discourse community,” which Swales describes in terms of six defining characteristics (Swales 1990: 24–27):

1. A discourse community has a broadly agreed set of common public goals. [...]
2. A discourse community has mechanisms of intercommunication among its members. [...]
3. A discourse community uses its participatory mechanisms primarily to provide information and feedback. [...]
4. A discourse community utilizes and hence possesses one or more genres in the communicative furtherance of its aims. [...]
5. In addition to owning genres, a discourse community has acquired some specific lexis. [...]
6. A discourse community has a threshold level of members with a suitable degree of relevant content and discursual expertise. [...]

Here, the term **discourse** refers to language with a specific communicative purpose used in a given social context, and **discourse community** represents the users of that particular discourse.²⁶ By way of example, Swales (1990) discusses a Hong Kong stamp-collectors’ club, of which he is also a member, in terms of these six discourse community characteristics. He emphasizes its “common goals, participatory mechanisms, information exchange, community-specific genres, highly specialized terminology, and high general level of expertise” (Swales 1990: 29), while recognizing that members’ backgrounds and other interests may be vastly different. Furthermore, Swales notes that members of discourse communities have differing levels of expertise, and that the distinction between a member and a non-member is a matter of gradation. Expert members are more easily able to exploit genre conventions than novices, due to their familiarity with the genre and their status within the discourse community. As a result, expert members exert

²⁵ Although most ESP studies are concerned with written texts, some research has also concentrated on the oral conventions of certain genres (e.g., Thompson 1994).

²⁶ Swales’s use of the terms discourse and discourse community is based on the social-constructionist studies of researchers such as Perelman, Kuhn, and Fish, as well as Foucault’s analysis of discursive formations (Swales 1990: 21).

more influence on discourse community-specific genres than peripheral or non-members, and are therefore more likely to be integral in defining a genre's constraints and, ultimately, its development (Askehave & Swales 2001: 198–199; Bhatia 1993: 14–15; 2004: 185).

In ESP genre analysis, the work of Swales (1981, 1990, 2004) and other ESP researchers (e.g., Bhatia 1993, Holmes 1997, Nwogu 1997), like that of SFL register analysis, is based on the study of the occurrence or absence of certain linguistic features. Indeed, ESP research draws extensively on SFL theory (Hyland 2002: 115), but differs in its description of genre-specific structural conventions.

The most widely used ESP genre analysis model, Swales's **structural moves analysis**, e.g., the four-move model (Swales 1981) and the Create a Research Space (CARS) model (Swales 1990), identifies a series of **moves** common to specific genres. These moves, defined as “discoursal or rhetorical units that perform a coherent communicative function in a written or spoken discourse” (Swales 2004: 228), are in turn realized by a series of “submoves” or **steps**, and describe the content and linguistic choices most commonly observed in a genre. The CARS model (Swales 1990: 141, fig. 10), for example, describes the introduction section of scientific research articles in terms of three basic moves—1) establish a territory, 2) establish a niche, 3) occupy the niche—which can be identified and realized by specific grammatical and lexical markers.

Moves, according to Swales (2004: 229), are functional rather than formal units, and Holmes (1997: 325) concedes that there is “a degree of subjectivity that is perhaps unavoidable” in describing a text's rhetorical movement. There are also ESP researchers who reject the linear progression of moves described by Swales in academic research articles in favor of a more “cyclical patterning in the writer's choice of moves” (Hopkins & Dudley-Evans 1988: 116–117). These inconsistencies may be exacerbated by a lack of consensus in the ESP literature concerning the definition of a move. Swales's own definition is somewhat vague (see above), and Holmes (1997: 325) describes a move merely as “a segment of text that is shaped and constrained by a specific communicative function.” A more thorough description, however, is that provided by Nwogu (1991, 1997). Nwogu's research into genre analysis, especially concerning the genre of the medical research article, is particularly relevant to this study (see section 2.6), and, in a later paper, Nwogu (1997: 122) defines a move as “a text segment made up of a bundle

of linguistic features (lexical meaning, propositional meanings, illocutionary force, etc.) which give the segment a uniform orientation and signal the content of discourse in it.” Each move in a particular section is in turn realized by a series of steps, or “constituent elements” (Nwogu 1997: 122), some of which are obligatory, some of which are optional (Swales 1990: 142). Again, it is Nwogu’s definition that seems to be the most explicit: “constituent elements or slots [...] combine in identifiable ways to constitute information in the move” (Nwogu 1997: 122). Nwogu goes on to state that these elements can be determined partly from contextual inference and partly from specific textual markers, including explicit prefacing statements, explicit lexical items, discourse conjuncts, and summary statements.

Nwogu (1997: 123–124) describes the process by which he identified moves and steps in his study corpus:

1. Focusing on the propositions in the texts and identifying important information.
2. Searching for linguistic clues such as function words, explicit lexemes and expressions, verb forms, discourse conjuncts and markers, structural headings and subheadings, summary statements, etc.
3. Classifying and paraphrasing the context of discourse based on linguistic clues.
4. Assigning discourse functions to the overall information in segments of text as well as constituent elements of information in the segments.
5. Establishing whether or not the function identified is a general one by reference to other texts in the corpus.

The first four “steps” of Nwogu’s five-step method are reminiscent of some of the SFL analyses described in section 2.2. For example, by “identifying important information,” searching for “discourse conjuncts” and “summary statements,” and “assigning discourse functions to the overall information in segments of the text,” Nwogu is effectively identifying Themes and hyperThemes as well as other functional elements in the discourse. Using this methodology, Nwogu identifies an 11-move schema comprising a series of 26 steps, which together provide a language-, content-, and structure-specific description of a corpus of medical research articles. The subsequent schema can be applied to help medical writers produce “clear, coherent and logically organized research reports” (Nwogu 1997: 119). From a pedagogic perspective, such a description of genre

may be useful in instructing learners (Swales 1990: 1–8), and the above-mentioned models of Swales (1990, 2004) and Nwogu (1997) are typical of the descriptive or even “prescriptive” (Dudley-Evans 1987: 1) pedagogic application of ESP research.²⁷

2.4.2 *The Sydney School: Systemic-Functional Genrists*

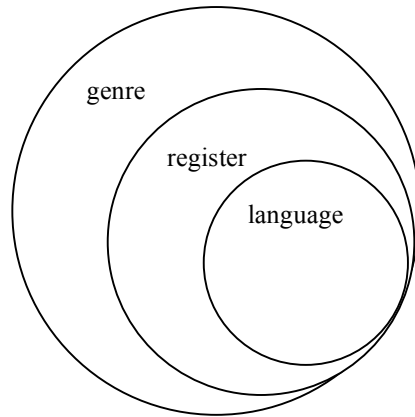
SFL has provided much of the theoretical basis for sociolinguistic research in recent years. In addition to their work with register analysis, systemic-functional linguists also use genre as a means of describing language use in specific contexts. Such researchers are often referred to as Systemic-Functional Genrists (SFG), or the “Sydney School” as they are more commonly known in North America (Freedman & Medway 1994: 3; Hyland 2002: 115).

Martin (2001: 163), a systemic-functional linguist at the University of Sydney, Australia, defines genres in terms of “purposeful goal-oriented activities” or social processes, and, in a later work (Martin & Rose 2003: 254–255), emphasizes how register is positioned within the wider sphere of genre (see Figure 2.1). This semiotic sphere, according to Thompson (2004: 42–43, emphasis added), includes “register plus *purpose*... [i.e.] the more general idea of what the interactants are doing through language, and how they organize the language event [...] in order to achieve that purpose.” However, this simple distinction can lead to an apparent contradiction in the literature (Ferguson 1994: 15; Swales 1990: 40), as some researchers define register and register variation explicitly in terms of *purpose* (e.g., Biber 1994: 32; 1995: 7; Miller 2004: 11; Romaine 1994: 64). Couture (1986) offers an alternative explanation to avoid the conflation of genre and register by claiming that registers are constrained by features at the lexicogrammatical level, while generic constraints operate at the level of discourse structure. Moreover, according to Couture (ibid.: 82), genres “can only be realized in completed texts or texts that can be projected as complete, for genre does more than specify kinds of codes extant in a group of related texts; it specifies conditions for beginning, continuing, and ending a text.” This allows Couture (ibid.: 86–87) to distinguish genres such as research reports, business reports, and newspaper articles from their accompanying registers—scientific language, business language, and the

²⁷ See, for example, Swales & Feak’s (2000) textbook on the writing of academic English.

language of journalism—while also emphasizing the pedagogic importance of understanding the relationship between the two systems.²⁸

Figure 2.1. Semiotic relationship of language, register, and genre (adapted from Martin & Rose 2003: 254, fig. 8.2).



In this study, it is accepted that genre can be distinguished from register in terms of both structure and purpose, and that genre is realized by register (Martin & Rose 2003: 254; Thompson 2004: 43). For example, a text classed as “medical writing,” as in the example used at the start of section 2, may be found in a medical textbook or a medical research article. The structure and the purpose of these two genres may differ, e.g., different formats and different intended readerships, but the register they use remains essentially the same.

Like ESP, SFG research has been used to produce descriptive frameworks that can be used by teachers and students to interpret, edit, and construct texts of specific genres (e.g., Christie 1992). However, unlike ESP, Systemic-Functional Genrists tend to be more explicit in terms of the social ideology of genres (see, for example, Freedman & Medway 1994, Martin 2000), and their work has explicitly attempted to empower underprivileged groups by providing access to “the cultural capital of socially valued genres through an explicit grammar of linguistic choices” (Hyland 2002: 115). As such, SFG research traditionally differs from ESP in its sites of application—i.e., elementary schools and migrant language programs (SFG) versus tertiary education and professional settings (ESP)—as well as in its

²⁸ Hasan (chapter 4 in Halliday & Hasan 1985; Hasan 1996) and Martin (1985) offer alternative explanations for distinguishing between register and genre. These are not discussed further in this study.

teaching strategies. Students focus on factual writing, such as reports, procedures, expositions, and explanations, and a scaffolded style of teaching, known as the teaching–learning cycle (Martin 2000: 118), is employed which provides teachers and students with explicit guidelines on generic staging and lexicogrammatical features (see examples in Butt et al. 2000: 9–14).

2.4.3 *New Rhetoricians*

Another influential perspective on genre studies is that offered by the “New Rhetoricians,” a group of predominantly North American scholars whose research focuses on rhetoric and the teaching of university composition classes. In a seminal paper, Miller (1984: 159) defines genres as “typified rhetorical actions based in recurrent events,” and emphasizes the social and historical significance of genres. Indeed, Miller’s concern, like that of other New Rhetoricians (e.g., Bazerman 1988, Freedman & Medway 1994), is primarily with the situational contexts in which genres occur and their social purposes (or *actions*), rather than with form or grammatical tendencies. Genres are specifically referred to as dynamic rather than fixed phenomena (Freedman & Medway 1994: 9; Miller 1984: 163), and students (primarily native English speakers in tertiary education) are encouraged to be aware of the social functions of genres, “the purpose of the text, the audience of the text, and the circumstances of the writing” (Hyon 1996: 704), rather than being given the descriptive and possibly even “authoritarian” frameworks prescribed by the Systemic-Functional Genrists (Freedman 1994: 192).

While the genre analysis of the New Rhetoricians provides little in the way of explicit teaching material, it does provide an important counterbalance to the research traditions and applications of ESP and SFG. An understanding of the social purposes of texts, the New Rhetoricians argue, is paramount to learners’ increased awareness and use of specific genres, and it is only by increased exposure to authentic genre-specific texts that the conventions and applications of genres can truly be mastered (Freedman 1994).²⁹ Therefore, New Rhetoric research is primarily concerned with investigating contexts. Its methodologies “tend to be

²⁹ Freedman’s view on genre competency is influenced by Krashen’s model of second-language acquisition (Krashen 1982). According to Krashen, the explicit rules of a language are too complex to be taught and can only be acquired through increased contact with, and participation in, the target language (Freedman 1994: 196).

ethnographic, rather than text analytic, with the aim of uncovering something of the attitudes, values, and beliefs of the communities of text users that genres imply and construct” (Hyland 2002: 114). Bazerman’s (1984, 1988) study of experimental reports in physics is an often-cited example of this kind of ethnographic genre-based approach. In *Shaping Written Knowledge*, Bazerman (1988) employs a diachronic perspective and equates changes in the genre to the changing needs of the discourse community. While there is a necessary amount of text analysis in the description of these changes, e.g., use of the passive, temporal and causal subordination, sentence length, etc., the study concentrates primarily on the development of physical science and how scientific researchers have effected and adapted to change over the last century. Such changes in genre over time, Bazerman (1988: 315) concludes, give the static description of generic features and structure little value, and simply emphasize the need to be aware of the processes at work behind the text. Writers and researchers are therefore encouraged to look not at static features such as structure and linguistic conventions, but to consider historical developments and current scientific practice. Bazerman (1988: 323–329) summarizes these considerations as follows:

1. Consider the writer’s fundamental assumptions, goals, and projects.
2. Consider the structure of the literature, the structure of the community, and the writer’s place in both.
3. Consider the immediate rhetorical situation and rhetorical task.
4. Consider the investigative and symbolic tools.
5. Consider the processes of knowledge production.
6. Accept the dialectics of emergent knowledge.

2.4.4 *Summarizing Genre*

A summary of the three above-mentioned genre perspectives is provided in Table 2.2. However, as mentioned at the beginning of section 2.4, it is worth noting that, despite the differences emphasized in the table below, there is considerable overlap between these three approaches in terms of the recognition and analysis of genre (see discussion in Hyland 2002: 114–115).

Table 2.2. Summarizing genre analysis

	Context	Framework	Goals
ESP	EAP, EPC, L2 tertiary education	structural moves analysis	guide for student/professional communication
SFG	primary and secondary education, and adult migrant education	explicit genre descriptions	student empowerment
New Rhetoricians	L1 tertiary education	no explicit framework	student awareness of “social action” of genres

EAP: English for Academic Purposes; EPC: English for Professional Contexts; L1/L2: first/second language

2.4.5 Text Types

Another method for describing language variation, which appears in the literature but is not discussed further in this paper, is in terms of **text type**. Unlike genre and register, text types are defined in “strictly linguistic terms” (Biber 1994: 52; Biber 1995: 10), and “external criteria” such as participant relationships, discourse community, purpose, etc. are ignored (Biber & Finegan 1986, cited in Ghadessy 1999: 127). Text types are first identified on formal grounds and grouped according to their linguistic characteristics before later being interpreted functionally (Biber 1995: 10). In this study, however, the term “text type” is used as a general cover term in the opening sections of the paper to refer to different registers or genres.

2.5 Previous Genre-Based Research Concerning Medical Writing

In terms of the methods of register and genre analysis described above, ESP research has been the most productive in describing the conventions of written medical discourse. Swales’s (1984, 1990, 2004) structural moves analysis, despite the shortcomings discussed in section 2.4.1, provides a well-established theoretical framework that has been used to describe various research report genres (Holmes 1997, Kanoksilapatham 2005, Peng 1987, Swales 1984, 1990, 2004), specific sections of the medical research article (Bruce 1984, Dubois 1997), and the medical research article as a whole (Nwogu 1997, Skelton 1994).

Of particular relevance to this study, Nwogu (1997: 135) carried out a structural moves analysis of a corpus of 15 medical research articles and consequently identified 11 schematic units or moves, which together constitute “the overall semantic macrostructure” of the genre. A similar methodology to that used by Nwogu is adopted for the textual analysis of this study corpus (see sections 2.4.1 and 3.3.2 for details). Nwogu’s study material was published during the years 1985–1987 (this study: 2004–2006) and is selected from similar publication sources to those of this study (see Collection of Study Material, section 3.1). This allows for a useful diachronic comparison between the two studies.

Another interesting study that adopts Swales’s structural moves analysis is that of Skelton (1994), which is based on a corpus of 50 medical research articles published in the *British Journal of General Practice* between 1989 and 1993. As the study is published in the journal itself, Skelton’s linguistic analysis, which is intended to be read by fellow medical researchers (Skelton is also a general practitioner), is less comprehensive than Nwogu’s (1997) and less specific in terms of how moves are defined, recognized, and categorized. However, since the same structural moves methodology has apparently been used, Skelton’s results, which differ from those of Nwogu (15 moves instead of 11), provide additional comparative material for this study’s textual analysis.

Although Skelton’s (1994) and Nwogu’s (1997) studies both draw on the genre analysis model of Swales, they contain little, if any, reference to the “communicative purposes” (Swales 1990: 58) of the medical research article and the context in which it is produced. Three useful studies in this respect are Swales (1990), Bazerman (1988), and Halliday & Martin (1993). Each considers register/genre in terms of its social context. Swales (1990: 21–27) emphasizes the importance of the discourse community; Bazerman (1988: 323–331) focuses on writer assumptions and goals in relation to the perceived audience; and Halliday & Martin (1993: 22–23) take a bottom-up approach to context using SFL. It should be noted, however, that none of these studies is exclusively concerned with medical writing. Nevertheless, their discussions of scientific discourse in general provide a foundation for this study’s analysis of the situational context of written medical discourse (see section 3.3.1 for methodological details).

Other linguistic studies of scientific or medical discourse have tended to focus on specific lexicogrammatical features of texts rather than on the organization

of information or the rhetorical purpose of the research article genre, e.g., modality (Adams Smith 1984), personals (Martínez 2005, Tarone et al. 1998), Processes (MacDonald 2002), tense/voice (Heslot 1982, cited in Swales 1990), Theme (Gosden 1992), and citations (Hyland 1999, Thompson & Tribble 2001). Such studies provide useful linguistic descriptions of the characteristics of the medical research article, and provide further comparative material for this study (see section 3.3.2).

2.6 Current Guidelines for the Writing of Medical Research Articles

Writing programs in English for specific or general academic purposes are designed to provide students with the skills they need to understand and, more importantly, to produce their own texts in their chosen academic field. Such courses are widespread (see examples in section 2.3) and offer guidelines on “effective writing style” and “standard scientific structures.”³⁰ Similar courses are also offered to more experienced researchers, particularly with respect to the publication of research papers.³¹

In addition to writing courses, and occasionally as an accompaniment to such courses (Wager 2005), a number of books have been published that offer advice to undergraduates, postgraduates, and researchers on the writing of medical reports and medical research articles (e.g., Day 1994, Goodman & Edwards 2006, Hall 2003, Young 1996). These books are written by journal editors and medical researchers, all with extensive writing and publishing experience and, in the case of Young (1996), a background in linguistics and literary analysis. In all the examples cited above, the authors describe the structure and rhetorical purpose of the medical research article, and provide instruction on how to replicate article content and style. Hall (2003), for example, a former editor of the *British Medical Journal*, presents contributions by a number of authors, many of whom are medical journal editors themselves, on the content and conventions of traditional medical research article sections (i.e., the Introduction, Methods, Results, and Discussion; see below)

³⁰ From the Department of Medical Sciences, University of Oxford (URL: www.medsci.ox.ac.uk/portal/skillstraining/courses/writingskills/writingskills2006_7, accessed February 5, 2007).

³¹ “Getting Research Published,” Department of Continuing Education, University of Oxford (URL: www.conted.ox.ac.uk/cpd/personaldev/courses/getting_research_published.asp, accessed February 5, 2007).

as well as on the overall publishing process. Similarly, Young (1996) discusses the linguistic and structural conventions adopted in the medical research article, as well as offering advice on the writing of case reports and the preparation of oral presentations.

In all of the above-mentioned publications, reference is invariably made to the requirements and conventions of the so-called Vancouver Group. Originally a small consortium of medical journal editors, the group first met in Vancouver, Canada, in 1978 to establish a set of guidelines for the formatting of manuscripts submitted to its journals (ICMJE 2006: 3). Officially referred to as the International Committee of Medical Journal Editors (ICMJE), its *Uniform Requirements for Manuscripts Submitted to Biomedical Journals* (ICMJE 2006) are currently endorsed by 638 medical journals, and updated editions of the guidelines are published on an annual basis.³² The intention of the ICMJE is “to help authors and editors in their mutual task of creating and distributing accurate, clear, easily accessible reports of biomedical studies” by providing “ethical principles in the conduct and reporting of research, and [...] recommendations relating to specific elements of editing and writing” (ICMJE 2006: 4). In regard to the latter, ICMJE recommends the Introduction–Methods–Results–Discussion (IMRD, also referred to as IMRAD) structure for the writing of original medical research articles, and provides under each of these headings information on how to structure content (see Appendix). According to ICMJE (2006: 24), the IMRD structure represents “a direct reflection of the process of scientific discovery,” and its origins in medical research, which emerged from a need to standardize research procedures as the discipline of medical research developed (Atkinson 1992, Young 1996), can be traced back to the early 1940s (Sollaci & Pereira 2004). Today, IMRD is considered the standard form for the writing and publishing of original medical research articles.

The ICMJE guidelines do not include a description of the linguistic features common to each of the IMRD sections, but a discussion of the generic characteristics of the medical research article would not be complete without considering the influence of the ICMJE and its guidelines on medical publishing.

³² Figure as of March 6, 2007 (see URL: www.icmje.org/jrnlist.html).

Both the textual and contextual analyses in this study therefore refer to ICMJE's *Uniform Requirements* and the influence it has on the genre.

As noted by ICMJE (2006: 4), journal-specific requirements also need to be considered when writing (and indeed analyzing) medical research articles. While many medical journals recommend the ICMJE manuscript guidelines to potential journal contributors (see above), they may have their own unique additional requirements, which may be technical (e.g., the reproduction of figures and tables), orthographic (e.g., insisting on certain spellings or abbreviations), or content specific (e.g., research articles versus case reports, or clinical versus experimental studies).³³ Some medical associations even publish their own "style manuals" (e.g., the American Medical Association; Iverson et al. 1998), which provide further detailed guidance for manuscript preparation.

2.7 Register and Genre in this Study: A Multidimensional Approach

In this study, the most important distinction between register and genre lies not necessarily in the description of purpose or discourse structure but in their differing methods of analysis; that is, the use of the register components field, tenor, and mode versus the application of genre-based structural moves analysis. Register and genre provide valuable methodological frameworks for analyzing text and context, and this paper attempts to combine both models in its textual and contextual description of the medical research article.

The register discourse components of field, tenor, and mode are a useful means of describing the context of situation as well as the discourse community involved in the production of the texts of this study corpus (see section 2.3). In terms of genre, ESP moves analysis provides a valuable framework for describing the structural and stylistic composition of a text (see section 2.4.1), and the work of the Systemic-Functional Genrists a means of describing function and thematic structure (see sections 2.2 and 2.4.2). The ethnographic approach to genre, as espoused by the New Rhetoricians, complements the primarily textual frameworks of ESP/SFG by emphasizing the relationship between text production and social

³³ See, for example, the *New England Journal of Medicine* (URL: <http://authors.nejm.org/Misc/NewMS.asp>, accessed March 9, 2007) and the *Lancet* (www.thelancet.com/authors/lancet/authorinfo, accessed March 9, 2007).

context (see section 2.4.3). Thus, this paper proposes a combined register–genre approach to discourse analysis that includes elements of all four of the above-mentioned perspectives on genre and register (see section 3.3 for detailed methodologies).

Such an approach to genre is motivated by the study of systemic-functional linguistics and genre analysis, as described above, and by Bhatia’s (2004) proposed four-space model. Bhatia’s research (e.g., Bhatia 1993, 2004) is influenced by Swales (1984, 1990), concentrating primarily on business and legal texts and the educational aspects of ESP genre analysis. However, it differs somewhat in its description of genre as a multidisciplinary activity, an approach which includes elements of linguistic as well as sociological and psychological research (Bhatia 1993: 16–22). This model of genre analysis is later refined in *Worlds of Written Discourse*, in which “a multidimensional analytical perspective” is proposed (Bhatia 2004: 163). Bhatia’s four-space genre-based model of analysis thus adopts textual, ethnographic, socio-cognitive, and socio-critical perspectives. Such a model is more comprehensive—and consequently more difficult to reproduce—than the model proposed here (see section 3.3), but it nevertheless expresses the range of possibilities available to genre research and, like the model in this paper, emphasizes the need for both textual and social analyses in the study of genre.

3 Material and Methods

3.1 Collection of Study Material

Articles for analysis were collected from the electronic archives of four international medical journals: the *British Medical Journal* (BMJ), the *Journal of the American Medical Association* (JAMA), the *Lancet* (LAN), and the *New England Journal of Medicine* (NEJM). These journals were selected on the basis of their reputation in the field of medicine, and are widely regarded as the leading periodicals in their field. Together they have the four highest “journal impact factors”—an index measuring how often a journal cites and is cited by other journals within a two-year period—in the Institute for Scientific Information’s “Medical, General” category (Garfield 1994, Institute for Scientific Information 2005). They are also the only international English-language medical research journals published on a weekly basis.

According to Bhatia (2004: 189), generic integrity in written academic discourse is largely maintained by the peer-review process and editorial intervention. Therefore, journals published as frequently as those listed above, and with such high standing within the medical research community, could arguably be regarded as the most significant organs for maintaining and developing the patterns and conventions of the medical research article, and were hence chosen as highly representative of the genre.

All articles included in the study were original research articles. Other submissions such as review articles, case reports, and short communications were excluded. In order to limit selection of the study material, database searches were restricted to articles dealing with obesity and overweight. The decision to concentrate on this particular field of research was based on familiarity with the content and context, and also the ease with which comparisons might be made among articles of similar content (see Methodology, section 3.3). Furthermore, only articles published in the years 2004–2006 were included. Again, this limited the search and, because genres are liable to change over time (Bazerman 1984, 1988; Bhatia 1993, 2004; Freedman & Miller 1994; Miller 1984), it was an attempt to

avoid including older articles that might not be considered “typical” of contemporary written medical discourse.

The search for relevant articles was conducted in March 2006, using the search engines of each of the four periodicals (see section 3.2 for journal URLs). Searches were restricted to original research articles published in an approximately two-year period between January 2004 and February 2006, using the terms *obes** (truncated search term) and *overweight*. These terms were entered into an open field (abstract, text, and title search) in order to yield as many returns as possible. Articles were then reviewed for content relevance, i.e., whether the study dealt specifically with obesity/overweight or whether reference was made to the condition in some other health-related context. Articles not dealing specifically with obesity/overweight were excluded.

In the initial search, certain articles were considered for exclusion if the main (first three) authors were affiliated to a non-English-medium research institute. This was done to avoid possible discrepancies in language use and structure that might not be considered typical of the genre and could therefore skew the results of the overall analysis (see, for example, Block & Chi 1995 on the differing citation traditions in Chinese and English academic writing). However, such an exclusion seemed neither feasible nor of particular value when describing medical research articles. It is not possible to distinguish native from non-native English speakers on the grounds of academic affiliation alone, nor is it possible to know what stages of language editing a manuscript may have gone through before being accepted for publication; and, even if it were, to exclude such articles from the analysis would be to suggest that only native English speakers are responsible for defining the conventions of the genre. This seems absurd given that the percentage of English-language medical research articles published by researchers from non-English-speaking countries may be as high as 50% (Dawson et al. 1998: 68–69). Exclusions were therefore only based on article type and subject relevance.

Seventeen articles were subsequently identified for analysis according to the above-mentioned selection criteria: five from BMJ, five from JAMA, four from NEJM, and three from LAN (see Table 3.1, section 3.2 for details). Articles were archived as electronic portable document format (PDF) files and comprised a corpus of approximately 104,000 words, including figures, tables, and references (see Table 3.2 for details).

3.2 Description of Study Corpus

For ease of reference, each article was coded according to its source journal and date of publication. A full list of the articles in the study corpus is provided below, in Table 3.1, and a summary follows in Table 3.2.

Table 3.1. Study article references and codes according to source journal

Corpus code	Article reference
BMJ-1	Ebrahim, S., Montaner, D., & Lawlor, D. A. (2004). Clustering of risk factors and social class in childhood and adulthood in British women's heart and health study: cross sectional analysis. <i>British Medical Journal</i> 328: 861–865.
BMJ-2	Baird, J., Fisher, D., Lucas, P., Kleijnen, J., Roberts, H., & Law, C. (2005). Being big or growing fast: systematic review of size and growth in infancy and later obesity. <i>British Medical Journal</i> 331: 929–934.
BMJ-3	Viner, R. M. & Cole, T. J. (2005). Adult socioeconomic, educational, social, and psychological outcomes of childhood obesity: a national birth cohort. <i>British Medical Journal</i> 330: 1354–1356.
BMJ-4	Reilly, J. J., Armstrong, J., Dorosty, A. R., Emmett, P. M., Ness, A., Rogers, I., Steer, C. & Sherriff, A., for the Avon Longitudinal Study of Parents and Children Study Team. (2005). Early life risk factors for obesity in childhood: cohort study. <i>British Medical Journal</i> 330: 1357–1362.
BMJ-5	Truby, H., Baic, S., deLooy, A., Fox, K. R., Livingstone, M. B. E., Logan, C. M., Macdonald, I. A., Morgan, L. M., Taylor, M. A., & Millward, D. J. (2006). Randomised controlled trial of four commercial weight loss programmes in the UK: initial findings from the BBC "diet trials." <i>British Medical Journal</i> 332: 1309–1314.
JAMA-1	Williams, J., Wake, M., Hesketh, K., Maher, E., & Waters, E. (2005). Health-related quality of life of overweight and obese children. <i>Journal of the American Medical Association</i> 293: 70–76.
JAMA-2	Lange, B. J., Gerbing, R. B., Feusner, J., Skolnik, J., Sacks, N., Smith, F. O., & Alonzo, T. A. (2005). Mortality in overweight and underweight children with acute myeloid leukemia. <i>Journal of the American Medical Association</i> 293: 203–211.
JAMA-3	Taylor, E. N., Stampfer, M. J., & Curhan, G. C. (2005). Obesity, weight gain, and the risk of kidney stones. <i>Journal of the American Medical Association</i> 293: 455–462.
JAMA-4	Flegal, K. M., Graubard, B. I., Williamson, D. F., & Gail, M. H. (2005). Excess deaths associated with underweight, overweight, and obesity. <i>Journal of the American Medical Association</i> 293: 1861–1867.
JAMA-5	Gregg, E. W., Cheng, Y. J., Cadwell, B. L., Imperatore, G., Williams, D. E., Flegal, K. M., Narayan, K. M. V., & Williamson, D. F. (2005). Secular trends in cardiovascular disease risk factors according to body mass index in US adults. <i>Journal of the American Medical Association</i> 293: 1868–1874.
LAN-1	Hancox, R. J., Milne, B. J., & Poulton, R. (2004). Association between child and adolescent television viewing and adult health: a longitudinal birth cohort study. <i>Lancet</i> 364: 257–262.
LAN-2	Pereira, M. A., Kartashov, A. I., Ebbeling, C. B., Van Horn, L., Slattery, M. L., Jacobs, D. R. Jr., & Ludwig, D. S. (2005). Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis. <i>Lancet</i> 365: 36–42.

LAN-3	Kimm, S. Y. S., Glynn, N. W., Obarzanek, E., Kriska, A. M., Daniels, S. R., Barton, B. A., & Liu, K. (2005). Relation between the changes in physical activity and body-mass index during adolescence: a multicentre longitudinal study. <i>Lancet</i> 366: 301–307.
NEJM-1	Bhargava, S. K., Sachdev, H. S., Fall, C. H. D., Osmond, C., Lakshmy, R., Barker, D. J. P., Biswas, S. K. D., Ramji, S., Prabhakaran, D., & Reddy, K. S. (2004). Relation of serial changes in childhood body-mass index to impaired glucose tolerance in young adulthood. <i>New England Journal of Medicine</i> 350: 865–875.
NEJM-2	Weiss, R., Dziura, J., Burgert, T. S., Tamborlane, W. V., Taksali, S. E., Yockel, C. W., Allen, K., Lopes, M., Savoye, M., Morrison, J., Sherwin, R. S., & Caprio, S. (2004). Obesity and the metabolic syndrome in children and adolescents. <i>New England Journal of Medicine</i> 350: 2362–2374.
NEJM-3	Hu, F. B., Willett, W. C., Li, T., Stampfer, M. J., Colditz, G. A., & Manson, J. E. (2004). Adiposity as compared with physical activity in predicting mortality among women. <i>New England Journal of Medicine</i> 351: 2694–2703.
NEJM-4	Wadden, T. A., Berkowitz, R. I., Womble, L. G., Sarwer, D. B., Phelan, S., Cato, R. K., Hesson, L. A., Osei, S. Y., Kaplan, R., & Stunkard, A. J. (2005). Randomized trial of lifestyle modification and pharmacotherapy for obesity. <i>New England Journal of Medicine</i> 353: 2111–2120.

Table 3.2. Statistical description of the study corpus

	Corpus	Per article, mean (range)
Words	104,201	6129 (4282–7523)
Pages	131	8 (5–13)
Authors	111	7 (2–12)
Tables / figures	59 / 20	3 (2–5) / 1 (0–3)
References	502	30 (17–48)

A summary of the publication details of the source journals is given below. This information was extracted from journal homepages, printed journal colophons, and the 2004 *Journal Citation Reports* of the Institute for Scientific Information (2005).

British Medical Journal

Corpus abbreviation: BMJ

URL: <http://www.bmj.com>

Founded: 1840

Ownership: British Medical Association

Publication details: 52 issues, 623 research articles, 7000–8000 submissions per year

Readership/circulation: print run 122,000, 1.2 million online users/month

Impact factor: 7.038

Content availability: all research articles freely available online without subscription, non-research material freely available 12 months after initial publication

Journal of the American Medical Association

Corpus abbreviation: JAMA

URL: <http://jama.ama-assn.org>

Founded: 1883

Ownership: American Medical Association

Publication details: 48 issues, 351 research articles, 6000 submissions per year

Readership/circulation: print run 350,000

Impact factor: 24.831

Content availability: selected research articles freely available online without subscription

Lancet

Corpus abbreviation: LAN

URL: <http://www.thelancet.com>

Founded: 1823

Ownership: Reed-Elsevier Group

Publication details: 52 issues, 415 research articles per year, submissions per year unknown

Readership/circulation: unknown

Impact factor: 21.713

Content availability: selected research articles freely available online without subscription

New England Journal of Medicine

Corpus abbreviation: NEJM

URL: <http://content.nejm.org>

Founded: 1812

Ownership: Massachusetts Medical Society

Publication details: 52 issues, 316 research articles, and 5000 submissions per year

Readership/circulation: 200,000 paying subscribers, 1.2–1.6 million online users/month

Impact factor: 38.570

Content availability: research articles freely available online without subscription 6 months after initial publication

3.3 Methodology

The multidimensional model used to analyze this study corpus combines four major sociolinguistic concepts: register, discourse community, structural moves, and thematic structure. These concepts are used to describe genre from two important perspectives. The first attempts to contextualize the medical research article in terms of its social and communicative purpose, by considering the context of

situation and the discourse community in which it is produced. The second provides a textual description of the medical research article, examining organizational structure and lexicogrammatical features by a combination of structural moves analysis and the functional grammar of SFL.

Although the textual and contextual analyses were conducted separately, this paper takes a functional view of language in which text and context are inextricably linked. Context plays a part in determining the text, and the text in determining the context (Halliday 1978: 3; Halliday & Hasan 1985: 47), so that if language is to be meaningful, one is dependent on the other. The proposed multidimensional model therefore provides an overall description of the genre of the medical research article, in which elements of the text can be predicted from, and are a product of, the context, and vice versa (Halliday & Hasan 1985: 34–37).

3.3.1 Contextual Analysis

An “ethnographic” approach was used to describe as comprehensively as possible the contextual setting of the medical research article. Texts concerning studies of weight-related health issues were chosen as the study corpus due to my familiarity with this area of research. As a language editor for a series of international medical research journals in which weight-related health issues are frequently discussed, and with 10 years’ editorial experience in English-language medical publishing, I consider myself at least a peripheral member of the discourse community that produces such texts (cf. Swales 1990: 29). This position and experience alone may not amount to the kind of ethnographic investigation endorsed by Bhatia (1993: 22; 2004: 163) and Bazerman (1988: 323–331), but they do provide additional insight into the content and context of the study material, which is particularly valuable in defining the discourse community and the register components of the discourse. Source journals (BMJ, JAMA, LAN, and NEJM) and the *Uniform Requirements of the ICMJE* (2006) provided additional resources for describing the contextual setting.

3.3.1.1 Context of Situation

Register was used to define the context of situation in which the medical research articles of the study corpus were constructed. The context of situation, i.e., the

immediate environment in which the texts were produced, can be described in terms of field, tenor, and mode. Examples of how these components can be applied to texts of different registers are provided in section 2.3. In the case of this study corpus, however, a more comprehensive description of each component was possible based on the availability of full-length texts and knowledge of the context in which the texts were produced.³⁴ Field, tenor, and mode were analyzed based on a thorough reading of the study texts in combination with additional published material. This additional material included “Instructions for Authors” and general manuscript submission guidelines in the source journals, as well as the *Uniform Requirements* of the ICMJE (2006).

Discourse components were defined based on the work of Butt et al. (2000: 182–195), Halliday & Hasan (1985: 12, 24–26), and Miller (2004: 162–174), and can be summarized as follows:

1. Field
 - a. the content of the text, and
 - b. the social activity taking place
2. Tenor
 - a. the participants, and
 - b. their discourse roles
3. Mode
 - a. the channel of communication, and
 - b. the text’s rhetorical function

3.3.1.2 Discourse Community

The discourse community involved in the construction of the texts of this study corpus was defined according to Swales (1990: 24–27; see section 2.4.1). Six discourse community characteristics were identified, and can be summarized as follows:

1. A set of common public goals

³⁴ Note that a lexicogrammatical analysis was not conducted in describing the context of situation. Detailed analysis of ideational, interpersonal, and textual meanings was instead conducted as part of the textual examination of the study corpus (see section 3.3.2).

2. The mechanisms of intercommunication among members
3. The information and feedback provided by members
4. The genre(s) utilized
5. Specific lexis
6. Membership: expert and non-expert members

These characteristics were described based on a thorough reading of the corpus texts as well as additional material such as the mission statements of the source journals (see section 3.2 for journal URLs) and the ICMJE *Uniform Requirements* (ICMJE 2006). Description of the discourse community was also based on personal observations of the editorial and publishing process.

3.3.2 Textual Analysis

As discussed above, research articles concerning weight-related health issues were chosen due to a familiarity with the subject. An additional reason for choosing texts of similar subject matter, rather than more diverse medical disciplines, was to allow easier recognition and comparison of structural and thematic patterns. Similar topics may be more easily generalized in terms of their content, which could make structural and thematic comparisons across texts easier. The disadvantage of such a decision, however, may be that it subsequently becomes more difficult to generalize the results to other medical research disciplines.

A methodology similar to that described by Nwogu (1997: 123–124; see section 2.4.1) was adopted for the structural analysis of this corpus of texts. The procedure, however, was modified to include the SFL concept of Theme/hyperTheme (Halliday & Matthiessen 2004, Martin & Rose 2003; see section 2.2.3) in order to identify and confirm changes in thematic structure. This additional step was taken in order to strengthen the theoretical basis for identifying changes in the discourse, as structural moves analysis alone was not considered sufficient to describe “rhetorical movement” within the texts (see, for example, Bhatia 1993: 31–32, and discussion in section 2.4.1). The corpus texts were also analyzed in terms of their ideational and interpersonal meanings, as well as other features of the textual metafunction, namely cohesion. Verb forms (tense, aspect, and voice) and the use of personals (personal, possessive, and reflexive pronouns)

were noted for comparative purposes (Martínez 2005, Swales 1990, Tarone et al. 1998). The use of non-integral references compared with the use of integral reference citations was also considered for comparison with previous studies (e.g., Hyland 1999, Swales 1990, Thompson & Tribble 2001). Non-integral references are citations “*outside* the sentence, usually placed within brackets, and which play no explicit grammatical role in the sentence” (Thompson & Tribble 2001: 92, authors’ italics), while integral references play a more explicit grammatical role, with the author’s name (or the name of the study) occurring in the citing sentence. The decision to use one form over the other is considered a matter of emphasis. Non-integral references background and generalize the research/researchers being cited, while integral references act to foreground specific research and researchers (Hyland 1999: 344). Examples are given in (43) and (44).

- (43) The frequency of obesity has risen at an alarming rate in all age and ethnic groups in the USA.^{1,2} (non-integral, superscript nos. refer to list at end of article; from LAN-2)
- (44) Dickson et al.²⁸ also found that underweight patients experienced higher treatment-related mortality. (integral; from JAMA-2)

Moves, steps, and their lexicogrammatical features were thus identified according to the following procedure:

1. Read through section to understand general content.
2. Highlight Themes and summarize content as a series of simplified bullet points.
3. Identify hyperThemes and assign appropriate discourse functions.
4. Note ideational, interpersonal, and textual features such as Processes, Participants, and Circumstances, clause types and clause relations, modality, and cohesion, as well as verb forms, the use of personals, and the type of reference citation.
5. Establish commonality of assigned discourse functions and linguistic features by comparison with other study corpus texts.

A rough analysis was initially carried out by reading through the texts, highlighting marked Themes, and noting possible hyperThemes/moves. After this

was done for all articles, a more detailed analysis following the above steps was conducted, in which the Introduction, Methods, Results, and Discussion sections of the corpus articles were analyzed in turn, i.e., all the Introduction sections, all the Methods sections, etc. This was done in order to retain consistency in the analysis of each section and to allow easier categorization of moves and their constituent steps as familiar patterns emerged. Moves and steps were identified by summarizing the content of each section as a series of bullet points and assigning hyperThemes to the text based on changes in thematic structure. Raw data were then compared across texts to identify common thematic features and, in some cases, were refined to reflect similarities.

Other lexicogrammatical features (see step 4 above) were first identified manually on hard copy or in the original article PDF files, either as part of the “hyperTheme/move analysis” or separately after the identification of moves and steps. The frequencies of occurrence and the collocations of certain words and phrases in the corpus were then calculated using either the search function of Adobe PDF Reader version 8.1.0 (Adobe Systems Inc.; URL: www.adobe.com) or the “concord” function of Oxford WordSmith Tools version 4.0 (Oxford University Press; URL: www.lexically.net/wordsmith/index.html). For analyses carried out in WordSmith, PDF files were converted to the text-only (.TXT) file format.

Identified moves and steps were considered “common” if they appeared in more than 50% (nine or more) of the corpus articles. This cut-off level for the identification of moves and steps is somewhat arbitrary, but is in line with the cut-off value used by Nwogu (1997: 124), and was chosen for comparative purposes. In addition to this, “move order stability” (Skelton 1994: 456), i.e., the frequency with which a move was flanked (before and after) by the same moves, was also calculated.³⁵

Unless stated otherwise, all lexicogrammatical characteristics of the study texts are classified and described according to the systemic-functional grammar of Halliday & Matthiessen (2004; see also section 2.2).

³⁵ For moves appearing in either first or last position in an article, only those moves coming either after or before were considered.

4 Contextualizing the Medical Research Article

4.1 Context of Situation

The register discourse components of field, tenor, and mode were used to describe the context of situation in which the corpus of medical research articles was produced. The results of the context-of-situation analysis are presented below. A summary is provided in Table 4.1.

4.1.1 Field

The content of the texts was in effect predefined in the study corpus selection criteria (see section 3.1) and can be summarized as “original research concerning health-related weight issues in human adult and child populations.” This particular aspect of the field of discourse was most clearly demonstrated in the article title, in the abstract (under the heading *Objective* or *Background*), at the end of the Introduction (in which the study aim or hypothesis was presented; see section 5.2.3 in textual analysis), and as a final statement in the Discussion section (the article’s conclusion; see section 5.5.3).

In terms of the activity taking place, all texts in the corpus report the results of scientific studies concerning underweight, overweight, and/or obese patients. These studies are invariably prompted either by apparent gaps identified by previously conducted research, or as a means of verifying previous research results. (The reasons for the studies are stated at several points in the articles, i.e., in the abstract, Introduction, and Discussion sections.) The results of the studies in this corpus are interpreted in relation to the results (and interpretations) of previously published studies, and suggestions are made concerning possible implications and areas requiring further research. The articles in this corpus can therefore be considered part of a wider discussion of research into health-related weight issues that extends beyond the textual confines of each individual article.

4.1.2 Tenor

The participants involved in the discourse can be broadly considered in terms of two groups: authors and readers.

All articles in the corpus are multi-authored texts, with an average of seven (range: 2–12) contributing authors per article (111 authors in total). According to their affiliations, all of these authors are based at universities (67%, 74/111), university-affiliated hospitals (17%, 19/111), or other medical research centers (16%, 18/111), and in only one of 17 articles are all authors based at the same research institute (BMJ-3, two-author article). According to affiliation, the majority of authors in the study corpus are based at institutions in the U.S. (57%, 63/111). Authors from the U.K. are also well represented (20%, 22/111). On a per-journal basis, all authors (19/19) in the BMJ subcorpus (5/17 articles) are based at institutions in the U.K. This is perhaps unsurprising considering the journal's publisher, editors, and the majority of its readers/members are based in the U.K. In the other three source journals, there is a more varied geographic spread (JAMA: U.S. 81%, Australia 19%; LAN: U.S. 76%, New Zealand 18%, Norway 6%; NEJM: U.S. 74%, India 18%, U.K. 8%), although these figures may also be affected by the geographic location of the journals' publishers and editors (see section 3.2).

Author specialties vary, but the most commonly represented medical disciplines according to affiliation are pediatrics (37%, 41/111), dietetics/nutrition (15%, 17/111), and preventive medicine (12%, 13/111). The conventions for displaying authors' academic/professional positions and/or academic qualifications differ among the source journals: BMJ provides author position without qualification, while JAMA, LAN, and NEJM provide author qualification without position (e.g., *Catherine Law, reader in children's health*, BMJ-1; *Joan Williams, PhD*, JAMA-1). In the three journals that provide details of authors' academic qualifications, the majority of authors are qualified as Ph.D. (41%, 34/82) or M.D. (40%, 33/82). Other less-common qualifications include D.M. (Doctor of Medicine; 4%, 3/82) and R.D. (Registered Dietitian; 2%, 2/82). Author roles in carrying out the study and producing the text are stated at the end of each article, and the order in which authors are listed is generally considered a reflection of overall contribution to the study (Rennie et al. 2000: 89; WAME 2007: unnumbered). Such

a measure of contribution, based on “a joint decision by co-authors” (ICJME 2006: 6), may be difficult, and author order is also likely to be a reflection of status, with authors of higher status being listed first (Moss 2003: 47–49).

The potential readers of the articles of this study corpus can be considered in terms of three subgroups corresponding to primary, secondary, and tertiary readerships. The primary addressees of these articles are likely to have similar research interests to the authors, as well as a similar level of background knowledge regarding medical research into health-related weight issues. It would seem reasonable to assume that this group of readers represents the writers’ intended readership, and the relationship between the role of author and primary reader is relatively equal in terms of status.

Another group of potential readers are medical researchers, general practitioners (GPs), and medical students with less specific knowledge (and perhaps less specific interest) in the subject than that of the article authors. This secondary readership comprises a group of readers with broad expertise in medicine, and whose fields of specialization may differ to those of the authors of the study articles. All four source publications are general medical journals, and their subscriber-based readerships are likely to consist of researchers, specialists, GPs, and students with a broad range of medical interests (see section 3.2). While this group may not represent the authors’ primary readership, the role of the authors in relation to a potential secondary reader is likely to be one of relatively equal status; although this may be a matter of gradation, i.e., more equal status among medical researchers, and less equal status between the authors and students of medicine.

A tertiary group of potential readers might also exist—one which the authors may be less likely to consider when preparing their manuscripts for publication. This group consists of readers who may have relatively limited medical knowledge/experience but a particular interest in the subject of weight-related health. This could include patients and their relatives, journalists, politicians, and journal publishing staff, and the status between authors and such a potential tertiary reader is likely to be relatively unequal in terms of shared knowledge and a general familiarity with the medical discourse.

Primary and secondary readerships (and indeed tertiary readers, in theory) are able to respond to the authors’ work by carrying out studies of their own, by submitting letters to the editor, or by “rapid response” commentaries posted online

alongside original research articles (in much the same way as an online discussion forum). For example, the online version of BMJ-2 is accompanied by seven “rapid responses,” three of which are critical of the authors’ study design and conclusions, and four of which lend their support to the authors’ findings.³⁶ The relationship between authors and readers in this case, as in the rest of the corpus, may therefore be more interactive than first appears, and reflects in part what Bazerman (1988: 23) terms the “dialectical interconnectedness” of the texts.

4.1.3 Mode

All the texts of this study corpus are examples of written language, and all were published in print and online formats (in identical versions) in their respective source journals.³⁷ The texts are monologic, but they function as part of a broader research dialogue, which includes comparisons with other studies as well as the possibility of interaction between authors and readers (see analysis of field and tenor above).

In terms of rhetorical function, each article section (IMRD; see section 2.6) serves its own distinct purpose. The Introduction presents the study and its aims in relation to previous research; the Methods section describes the selection of study material, and recounts the procedure and techniques used to analyze the material; the Results section reports data obtained in relation to the methodology; and the Discussion interprets the results in relation to previous studies, debates the implications of the study, and suggests areas in which further research may be warranted. The overall rhetorical function of the medical research article in this corpus can therefore be summarized as a combination of presentation–recount–report–discussion.

³⁶ See <http://www.bmj.com/cgi/content/full/331/7522/929#responses> (accessed May 25, 2006).

³⁷ Online texts, however, are usually made available to readers prior to print publication. See, for example, “early online publication” in the *Lancet*: <http://www.thelancet.com/authors/lancet/authorinfo#eop> (accessed September 18, 2007).

Table 4.1. Summary of the context-of-situation analysis

Field	Tenor	Mode
Content: study data concerning “weight-related health issues” Social activity: presentation and interpretation of study results in relation to other studies	Participants: authors and readers, most of whom have medical background/education, plus potential non-specialists Discourse roles: researchers to primary (equal status), secondary (relatively equal status), and tertiary readerships (unequal status)	Channel of communication: written; paper and online formats; monologue (with potential for “dialogue” with readers) Rhetorical function: presentation – recount – report – discussion

4.2 Medical Discourse Community

Six discourse community characteristics were identified in relation to the study corpus. These characteristics are described below.

4.2.1 Common Public Goals

The goals common to the authors of this study corpus are exemplified in the source journals’ mission statements. BMJ’s aims are “to lead the debate on health, and to engage, inform, and stimulate doctors, researchers and other health professionals in ways that will improve outcomes for patients.”³⁸ JAMA’s key objective is “to promote the science and art of medicine and the betterment of the public health.”³⁹ LAN “seeks to publish high-quality clinical trials that will alter medical practice.”⁴⁰ And NEJM aims to publish “new medical research findings, review articles, and editorial opinion on a wide variety of topics of importance to biomedical science and clinical practice.”⁴¹ The *Uniform Requirements* of the ICMJE, which are endorsed by all the source journals of this study, summarizes these communicative objectives in terms of “creating and distributing accurate, clear, easily accessible reports of biomedical studies” (ICMJE 2006: 4).

³⁸ From URL: <http://resources.bmj.com/bmj/about-bmj> (accessed March 3, 2007).

³⁹ From URL: <http://jama.ama-assn.org/misc/aboutjama.dtl> (accessed March 3, 2007).

⁴⁰ From URL: www.thelancet.com/about (accessed March 3, 2007).

⁴¹ From URL: <http://content.nejm.org/> (accessed March 3, 2007).

There may also be common personal goals among the authors that differ from those described above and that may be less obvious than the public goals of the discourse community, which could include career- and sponsor-related reasons for publishing research material. However, insofar as the authors are part of a specific discourse community, their use of these source journals as a medium for the dissemination of their research results may be considered as tacit agreement with the journals' above-mentioned public goals. The four source journals require all contributing authors to declare potential conflicts of interest, which may be related to authors' personal/private goals, in order to make clear any possible publication bias. A declaration is given at the end of all articles. Four potential conflicts of interest, each related to the receipt of corporate sponsorship, are declared in the 17 research articles of this study corpus. For example, in BMJ-5, this statement reads:

- (45) KRF [author initials] receives consulting fees for serving on the scientific advisory panel of Slimming World, a company that offers a support service for weight loss. This company was not involved in this trial but as it is similar to Rosemary Conley and Weight Watchers, the conclusions may have implications for the company.

4.2.2 Mechanisms of Intercommunication

The most apparent participatory mechanisms concerning the authors of this study corpus are the source journals from which these articles are taken. Additional modes of intercommunication used by these authors may include other scholarly publications, newsletters, conferences and meetings, and online discussion groups.

4.2.3 Information and Feedback

The participatory mechanisms described above provide forums through which an exchange of information can take place between members of the discourse community. As discussed in the context-of-situation analysis (see section 4.1.1), the authors of this study corpus use the source journals to present study data and to prompt comment and further research. This cycle of information and feedback provides the impetus for discourse development and may be considered part of a collective attempt to expand the limits of medical research.

4.2.4 Genre Utilized

The medical discourse community has certain discursive expectations (Swales 1990: 26) in the reporting of original research. The most apparent is the use of the IMRD structure and the expected content within each of those parts (cf. rhetorical function of article sections, section 4.1.3). These discursive expectations or conventions are reflected in the genre utilized by the discourse community, in this case the medical research article. Although other genres may have similar conventions, e.g., the use of IMRD in psychology research articles (American Psychological Association 1983), other discursive expectations related to the language (see 4.2.5 below), content (see 4.1.1 above), and purpose (see 4.2.1 above) of the research article set the genres apart. In this respect, the medical research article genre, as it is (or will be) defined in this paper, is considered unique to the medical discourse community.

Discursive expectations are analyzed and discussed at greater length in the textual analysis of section 5.

4.2.5 Specific Lexis

The study corpus contains a considerable number of specialized lexical items and a high level of **lexical density**.⁴² Much of this lexis is common to medical research in general, and is easily accessible to the primary and secondary readers described in section 4.1.2. However, it might be more difficult for a tertiary reader, a possible “outsider” to the discourse community, to decipher. Specialized terminology is used to refer to medical conditions and syndromes (e.g., *left ventricular hypertrophy*, *hypercholesterolemia*), clinical parameters (e.g., *high-density lipoprotein cholesterol*, *adiponectin*), measuring devices and methodologies (e.g., *Harpenden stadiometer*, *Minnesota codes*), and the description of study material (*ALSPAC – the Avon Longitudinal Study of Parents and Children*). Other forms of specialized lexis such as *P values*, *SD/SEM*, and *confidence intervals* are

⁴² Lexical density: the number of lexical words (content words) per clause. In the opening paragraph of the Introduction of JAMA-4, for example, the average lexical density is 9.7 (range 3–22), compared to typical values of 2 in informal spoken English and 4–6 in general written English (cf. Halliday 1993b: 76). Lexical densities are also provided for examples (46)–(48) below.

mathematical terms and may be common to other academic disciplines in which statistics are used. Some examples from the corpus are provided below.

- (46) Recent interest in clustering of risk factors has focused on the components of insulin resistance syndrome (hyperinsulinaemia, glucose intolerance, obesity, dyslipidaemia, and hypertension).

(from Introduction of BMJ-1; lexical density 16)

- (47) Univariate generalized linear models were used to determine the estimated marginal means of the PedsQL scales and subscales adjusting for the child's age, sex, maternal education, and disadvantage index as covariates.

(from Methods section of JAMA-1; average lexical density 10.5)

- (48) Values for glucose, insulin, insulin resistance, triglycerides, C-reactive protein, interleukin-6, and systolic blood pressure, as well as the prevalence of impaired glucose tolerance, increased significantly with increasing obesity, whereas HDL cholesterol and adiponectin levels decreased with increasing obesity.

(from Results section of NEJM-2; average lexical density 13.5)

In this study corpus, there is relatively little specialized lexis referring specifically to “health-related weight issues.” The few exceptions, such as *adiposity*, *obesity*, and *BMI [body-mass index]*, are terms and acronyms that are either widely recognized, even beyond the specialized medical discourse, or are explicitly defined when introduced. This apparent lack of specific “health-related weight issues” lexis demonstrates one reason why the authors and articles of this study corpus may be considered part of a broader medical research discourse community, in which common community-specific terms and acronyms are used, rather than as part of a separate, relatively limited “health-related weight issues” discourse community (Swales 1990: 26–27; see also section 3.3.2 regarding concern over choice of medical topic). The same can also be said of the common public goals described above (see section 4.2.1). Although there are medical journals that specialize in health-related weight issues, e.g., the *International Journal of Obesity*, whose public goal is to be a “forum for research describing... aspects of obesity and related disorders,” the medical journals of this corpus have

more general medical research objectives that appeal to a general medical research discourse community.⁴³

Further analysis of specific lexis is carried out in the textual analysis of section 5.

4.2.6 Membership

Membership of the discourse community is a matter of gradation. At one end of the scale are those who may be considered expert members, such as the authors of the study articles and their intended readership, and at the other are novices, such as students of medicine and recently qualified practitioners. Expert members share similar background knowledge references and are actively involved in the medical research discourse. Novices, on the other hand, lack the experience and knowledge necessary to participate fully.

In the case of this corpus, the authors are considered expert members. However, some contributors may have a higher level of expertise than others. This could be reflected in the ordering of author names and in their contributions listed at the end of the articles (cf. Rennie et al. 2000: 89–90). However, some authors appear to specialize in different fields (e.g., pediatrics, statistics, nutrition, etc.), and their contributions may equally be a reflection of their specialization—and therefore their subsequent degree of involvement in a particular study—as they are differing levels of expertise within the discourse community.

⁴³ See “aims and scope” of the *International Journal of Obesity*. Available at URL: <http://www.nature.com/ijo/about.html> (accessed September 3, 2007).

5 Textual Analysis of the Medical Research Article

5.1 Structural Moves Analysis

All four source journals endorse the ICMJE guidelines for manuscript preparation (ICMJE 2006; see section 2.6 and Appendix), and all original research articles in the study corpus adhered to the IMRD structure.

According to the textual analysis, the articles of this study corpus followed a distinct structural and thematic pattern. This pattern was governed by the traditional IMRD structure, in which each section has its own predefined rhetorical function (see sections 2.6 and 4.1.3), but there were also discernible structural and thematic patterns integral to each article section. A detailed section-by-section analysis is provided in sections 5.2–5.5, the results of which can be summarized as follows:

Introduction

1. Presentation of study background
 - a) Reporting established knowledge (general to specific)
2. Identification of gap(s) in current research
 - a) Identifying lack of data (or questionable data) in specific area related to established field; b) reason for need to fill gap
3. Statement of research purpose
 - a) Hypothesis/objective; b) brief description of study material and methodological procedure

Methods

4. Description of material/participants
 - a) Size of study sample; b) study period; c) selection criteria; d) type of data collection (e.g., interview, questionnaire, clinical exam); e) frequency of data collection; f) study approval and informed consent
5. Description of experimental procedure
 - a) Measurements taken: techniques/protocol, apparatus (and precision), frequency; b) definition of terms; c) categorizing sample into study subgroups; d) study endpoints/outcomes
6. Description of data-analysis procedure
 - a) Statistical test techniques, reasons for use; b) software used for analysis

Results

7. Report of observations
 - a) Reference to non-verbal material; b) statement of main findings; c) associations and correlations (and lack of), incl. statistical significance; d) adjustments made to analysis

Discussion

8. Discussion of main findings
 - a) Findings in relation to hypothesis/study objective; b) comparison with literature; c) possible mechanisms/causes, and their implications
9. Study limitations
 - a) Strengths and weaknesses of study
10. Conclusion
 - a) Main findings; b) implication of findings; c) recommendations/suggestions for future research

Ten rhetorical moves were identified: three in the Introduction, three in the Methods, one in the Results, and three in the Discussion.⁴⁴ These moves consisted of a series of steps (1–6 steps, 28 in total, indicated by letters in the summary above). The frequencies and order stabilities of these moves and steps are reported in Table 5.1.

Table 5.1. Frequency and stability of moves and steps identified in the corpus

Move / steps	Frequency of move / steps (%) [*]	Order stability, moves only (%)
1 / a	100 / 100	100
2 / a, b	100 / 94, 71	100
3 / a, b	100 / 94, 76	100
4 / a, b, c, d, e, f	100 / 82, 94, 88, 100 [†] , 65, 88 [‡]	100
5 / a, b, c, d	100 / 94, 59, 71, 59	94
6 / a, b	94 / 100, 69	82 (100) [¶]
7 / a, b, c, d	100 / 100, 82, 94, 82	71 (88) [¶]
8 / a, b, c	100 / 100, 100, 71	88
9 / a	88 [§] / 100	94
10 / a, b, c	100 / 65, 82, 88	88

^{*} Move frequency = (no. of move occurrences / no. of articles) × 100; Step frequency = (no. of step occurrences / no. of related moves) × 100. The frequency of a step depends on the occurrence of the move it realizes, not on the number of articles in the corpus, and can therefore have a higher value than its associated move.

[†] This step was a characteristic of move 4 in 11 articles (65%). It also appeared under move 5 (6/17, 35%) in relation to step 5a. The frequency reported here reflects the step's overall occurrence in the study articles.

[‡] This step was a characteristic of move 4 in seven articles (41%). It also appeared under a separate heading at the end of BMJ articles (5/17, 29%) and as a separate move at the end of some Methods sections (3/17, 18%). The frequency reported here reflects the step's overall occurrence in the study articles.

[§] Only one article made no mention of study limitations. In another article, reference to "possible limitations" was made as part of move 8, but was considered too vague to be counted as a separate move.

[¶] Values in parentheses represent adjusted move-order stability if the LAN-only move "role of the funding source" is omitted.

All 10 moves occurred with high frequency (88–100%), and seven (70%) were essential to the rhetorical structure of all articles in the study corpus, i.e., they appeared with a frequency of 100%. An additional move was identified in articles published in LAN only ("role of the funding source", section 5.3.4; 3/17 articles, 18%), but this was not an integral part of the IMRD structure of the rest of the study corpus. In the other articles, information concerning funding was provided after the main body of text, as a reduced-font-size endnote.

⁴⁴ In LAN articles, an additional rhetorical move, "role of the funding source," appeared at the end of the Methods section (see section 5.3.4 for further details).

The frequencies of steps varied more considerably (59–100%) than those of the moves, and only seven steps (7/30, 23%) were an essential part of all articles. In addition to the 28 steps summarized above, a further four possible steps (see sections 5.3–5.5) were identified. Each of these four steps had frequencies of less than 50%.

As shown in Table 5.1, the adjusted order stabilities of the 10 moves ranged from 88% to 100%. This suggests that, in this corpus, the order in which information is presented is relatively fixed. This stability is arguably a result of the IMRD structure, which is consistently used throughout the study corpus. Each section consists of only one to three moves, such that instability within sections, and indeed across sections, is unlikely. Steps, however, were generally less stable than moves, ranging from 6% (e.g., step 4d) to 88% (e.g., step 1a). There could be several plausible reasons for this. Some steps were identified under different moves, e.g., steps 4d and 4f, which were also identified under move 5 and as a separate move in some instances (see Table 5.1). While the frequencies with which these steps appeared in the study corpus were high (step 4d: 100%; step 4f: 88%), their positions in relation to other steps varied considerably, yielding low order-stability values (both 6%). Conversely, some steps had relatively low frequencies (e.g., step 5b; 59%), so that the steps preceding or following them had subsequently low stabilities. The number of possibilities/combinations increases when examining the text from a top-down perspective, i.e., article→section→move→step (1→4→10→28, respectively), and hence the stability decreases. This suggests that the order of steps may be of less importance than the order of sections and moves.

The detailed results of the structural and thematic analysis of this study corpus are presented below according to article section (IMRD), and each section is summarized in terms of selected lexicogrammatical and textual features. Observations are illustrated by selected examples from the study corpus, and these are generally intended to be representative of the moves and steps described. Occasional atypical findings are also highlighted.

5.2 Analysis of Introduction Section

The overall rhetorical function of the Introduction section in this corpus of medical research articles was to present the study in relation to previous research. This was done in terms of three basic structural moves.

5.2.1 Move 1

Move 1, “presenting the study’s background,” was identified in all articles of the study corpus, and it was always the opening move (see Table 5.1 above). As demonstrated in the opening paragraph of LAN-2 below, this phase of discourse (labeled step 1a) begins by referring to generally established knowledge before discussing research more specific to the study—in this case, the effect of fast-food eating habits on bodyweight and insulin resistance in two different ethnic groups of U.S. adolescents.

- (49) The frequency of obesity has risen at an alarming rate in all age and ethnic groups in the USA.^{1,2} The age-adjusted prevalence of obesity, defined as a body-mass index (BMI) of 30 kg/m² or greater, was 30·5% in 1999–2000 compared with 22·9% in 1988–1994, with even higher rates in ethnic minority groups.¹ About two of every three US adults and four of five African-American women were overweight or obese in 1999–2000.¹ In children and adolescents, the prevalence of being overweight rose by 50% in the past decade to about 15%.² (LAN-2)

The Themes in the example above are underlined for the sake of clarity. The first two, both of which are unmarked, highlight the issue of obesity and its prevalence. In the third, also an unmarked Theme, the Subject is modified so as to emphasize numerical value and ethnic relevance. The final clause, which is marked thematically by a Circumstance (the prepositional phrase *In children and adolescents*), foregrounds the age group this particular study is concerned with. Themes in step 1a were generally unmarked, as also demonstrated in the underlined Themes of examples (50) and (51) below, and were usually nouns/nominal groups referring to a medical condition, its prevalence, or other studies.

In (49)–(51), references are continuously made to previous studies—a superscript number referring to a list of numbered references at the end of the article (non-integral references). These references help to situate the study within a

wider research context and thereby emphasize the “intertextuality” of the texts (Bazerman 1993: 20–21; Halliday & Hasan 1985: 47; see also sections 4.1.1 and 4.1.2). This type of referencing was characteristic of moves 1 and 2.

Although the predominant verb form in the example above is the simple past, this was not the case throughout the corpus. Indeed, in the second paragraph of LAN-2, which is still part of move 1, the tense changes to the simple present (see (50) below).

- (50) The medical and economic outcomes of excessive bodyweight are great, including an estimated 300 000 excess deaths and at least US\$100 billion per year in medical expenditures.³⁻⁶ One particularly ominous public-health issue is the occurrence of glucose intolerance⁷ and type 2 diabetes in obese adolescents and young adults.⁸
- (LAN-2)

This change in tense is inverted in some articles and mixed in others. In (51), from JAMA-3, the simple present is used first to describe the current research situation; a situation the authors may feel to be common knowledge among readers. Not until midway through the second paragraph does the verb form change to the simple past, in this case to state the results of two specific previous studies.

- (51) Kidney stones are a major cause of morbidity. The lifetime prevalence of symptomatic nephrolithiasis is approximately 10% in men and 5% in women,¹⁻³ and more than \$2 billion is spent on treatment each year.^{4,5} About 80% of kidney stones contain calcium, and the majority of calcium stones consist primarily of calcium oxalate.^{6,7} [...]
- Obesity is associated with insulin resistance and compensatory hyperinsulinemia, metabolic derangements that may lead to the formation of calcium-containing kidney stones. A recent metabolic trial demonstrated that insulin resistance was associated with defects in renal ammonium production,⁸ and an examination of more than 4500 patients with a history of kidney stones showed that urinary pH was inversely related to body weight.⁹
- (JAMA-3)

The most commonly observed Process type in move 1 was relational, i.e., a Process that functions to relate a Participant to its identity or description. In (51), for example, 10 of 11 Processes (91%) are relational, the only exception being the transitive material Process *spent* in the first paragraph. This is also the only clause in the above examples that is written in the passive voice.

Although personal pronouns such as *we* and *our* were not observed in move 1, “author presence” in the form of modality was expressed by the modal Finite *may* (see second paragraph, first sentence of (51) above). Similar modal Finites were identified in other corpus articles as well as the modal Finites *could* and *might*. Characteristic lexis in this step included Themes highlighting a particular medical condition, e.g., *obesity*, *kidney stones*, *metabolic syndrome*, etc., as well as nominal groups referring to previous studies and recent clinical trials.

5.2.2 Move 2

Move 2, “identifying a gap in the current research,” was present in all the study articles, and always followed directly on from move 1 (see Table 5.1). This particular phase of discourse was realized by two steps.

Step 2a, which appeared in 16 of the 17 articles (94%), identified a lack of data in the existing research or contradictory results in the literature. The underlined multiple Theme in example (52) from NEJM-3, which begins with a textual Theme, the adversative conjunctive Adjunct *However*, and contains an experiential Theme referring to “evidence,” is typical of this step. This same type of multiple Theme was observed as part of step 2a in 50% (8/16) of the study articles.

- (52) It has been suggested that higher levels of physical fitness can eliminate the effect of excess weight and obesity on morbidity and mortality and that, thus, obesity may be a less important determinant of mortality than is fitness. However, evidence in support of this hypothesis has been limited and inconsistent.^{4,5} (NEJM-3)

In (53) below, step 2a is signaled by a conjunction rather than by a conjunctive Adjunct. The effect, however, is much the same as that in (52). The dependent clause of (53) functions in a similar way to the conjunctive Adjunct *however* and the previous clause(s) it contrasts in (52), and the Theme of the independent clause of (53) is similar to that of the experiential Theme of (52). This is best illustrated if (53) is rewritten as two independent clauses linked by a conjunctive Adjunct (see (53a) below).

- (53) Although larger body size may increase the urinary supersaturation of calcium salts, prospective data on the relation between body size and the risk of kidney stone

formation are limited. (JAMA-3)

- (53a) Larger body size may increase the urinary supersaturation of calcium salts. However, prospective data on the relation between body size and the risk of kidney stone formation are limited. (adapted from JAMA-3)

In identifying an apparent gap in the research, a second step (step 2b) was used to provide the reason for the need to fill the gap. This step was identified in 12 (71%) of the study articles. In example (54) below, from BMJ-3, a difficulty is highlighted which is then used to justify the need for the study. Of particular relevance in identifying this step were concessive and causal conjunctive Adjuncts such as *nevertheless*, *yet*, *thus*, and *therefore*, as well as lexis stressing necessity, value, and importance. Conjunctive Adjuncts were usually foregrounded in the Theme position, as in the underlined examples of (54) and (55).

- (54) The study of adult outcomes of childhood obesity is difficult because obesity often continues into adult life and therefore poorer socioeconomic and educational outcomes may actually reflect confounding by adult obesity. Yet identifying outcomes related to obesity confined to childhood is important in determining whether people who are obese in childhood and who later lose weight remain at risk for adult adversity and inequalities. (BMJ-3)
- (55) It is unclear whether the crucial phase of postnatal weight gain is in infancy, childhood, adolescence, or adulthood. Thus, longitudinal studies are required to determine the age at which preventive interventions should be initiated. (NEJM-1)

The verb forms used to identify a gap in the research varied, but were usually either the simple present, e.g., *there is no information on whether childhood television viewing affects adult health* (LAN-1), or the present perfect (see example (52) above). This choice appeared to be influenced by whether the authors, in identifying a gap, specifically referred to the shortcomings of previous studies (present perfect or simple past) or to a more general lack of research in a particular area (simple present).

Process types in move 2 were predominantly relational. For example, (54), from BMJ-3, consists entirely of such Processes. Note that the relational Processes in (52) and (53), used above to illustrate step 2a, are modified by the modal Finite

may. In identifying a gap or a paucity of data in the research, current knowledge is brought into question, and the findings of previous studies are thereby interpreted with caution, perhaps even skepticism. By expressing reservation in step 2a, the authors are also implying a need for their study, regardless of whether this was made explicit in step 2b. In addition to probability, an expression of obligation was also noted in step 2a: *Given the lack of evidence of effective treatments, action to achieve this target must focus mainly on prevention* (BMJ-2). No instances of first-person pronouns were identified.

As in move 1, non-integral references were used throughout move 2 to position the study within its wider research context, and examples of integral references, although significantly fewer than the number of non-integral references (only six instances in move 2), were also observed. For example, in (56) below, the use of integral references (underlined) emphasizes and specifies the perceived scarcity of data, and acts as a reference point for both the aim of the study (step 3a; section 5.2.3 below) and the subsequent comparison of the study findings with those in the literature (step 8b; see section 5.5.1). Note that the dominant clauses in this example are mental (cognitive mental Process *found*) and describe other authors' (Senser) study findings (Phenomenon). (In the first underlined instance, *a recent systematic review* is construed as Senser since it represents a "conscious human collective," i.e., the authors who carried out the review; see Halliday & Matthiessen 2004: 202.)

- (56) A recent systematic review found [...] only two longitudinal studies of the socioeconomic effects of obesity in adolescence.³ Gortmaker et al found that US women who had been obese in late adolescence in 1981 were less likely to be married and had lower incomes seven years later than women who had not been overweight, while men who had been overweight were less likely to be married.⁴ Sargent et al found that UK women, but not men, who had been obese at 16 years in 1974 earned 7.4% less than their non-obese peers at age 23.⁵ (BMJ-3)

5.2.3 Move 3

Move 3, "stating the research purpose," was identified in all articles, and always directly followed move 2 (see Table 5.1). This move was realized by two steps.

Step 3a provided an objective and, in some cases, a hypothesis for the study. This step was identified in 16 (94%) of the corpus articles. In the first example, (57) from LAN-3, the authors state the study's objectives without providing an explicit hypothesis. In the second, (58) from BMJ-1, hypotheses and a study objective are included.

- (57) In this study, we aimed to examine the longitudinal relations between changes in habitual activity and changes in BMI and adiposity during adolescence and to compare changes between girls who remained active and those who were less active. (LAN-3)
- (58) Socioeconomic position in childhood has strong effects on distributions of risk factors in adult life⁹ and is important in determining components of the insulin resistance syndrome¹⁰ and coronary heart disease,⁹⁻¹¹ leading us to hypothesise that socioeconomic position might be associated with differences in clustering of cardiovascular risk factors. [...] We therefore predicted that risk factors measured in adult life would cluster to a greater extent in populations with adverse socioeconomic position. [...] We explored the occurrence and clustering of risk factors for coronary heart disease in a representative sample of older women classified by socioeconomic position in childhood and in adult life. (BMJ-1)

In (57), the opening Theme (underlined) is marked by a locational Circumstance, the prepositional phrase *In this study*. This is used for contrastive purposes—as the previous paragraphs, in which moves 1 and 2 are realized, concentrate on the findings of other clinical studies and trials—and marks the beginning of a new, albeit short, phase of discourse. Example (58), on the other hand, differs in that the study objective is stated without the use of a marked Theme (last sentence, underlined); it also appears to be less explicit than the objective of (57). This is likely due to the hypotheses stated in prior clauses. In (58), it is the hypotheses, and not the objective, that signal the beginning of the step.

Step 3a signifies a shift from the description of previous research (“other people’s research”) to a statement of the intentions of the current study (“our research”), and it is the use of the personal pronouns *we* and *us* that most clearly signals this change in the examples above.⁴⁵

Although the simple-present verb form was occasionally used in step 3a (see (58) above), study hypotheses and objectives were always stated in the simple

⁴⁵ Further analysis concerning the use of personals is provided in section 5.6.

past. Process types were often relational, as in previous moves, but, as demonstrated in (57) and (58) above, cognitive mental Processes such as *aimed*, *hypothesise*, *predicted*, and *explored* were used to present hypotheses and objectives. Such Processes construe perception and cognition, and therefore require a Senser (in this case the personal pronoun *we*), which might be why they were chosen by authors to state supposition and intention. However, this was not always the case. In some instances, the authors opted to use relational Processes. In (59) below, for example, a Carrier (*The objective of this study*) is preferred to a Senser in describing the study objective. Here, what the authors aim to do in the study is rendered as a nominal group, and the authors as participants in the research are omitted. Halliday (1993a: 64–66) refers to this use of relational Processes as a characteristic of the “depersonalized” abstraction of scientific writing, and using such constructions may be a powerful rhetorical tool, considering the objective ideals of scientific research (Gross 1990: 15; Halliday & Matthiessen 2004: 234).

- (59) The objective of this study was to estimate deaths associated with underweight, overweight, and obesity in the United States in 2000 by using all available mortality data from the NHANES [National Health and Nutrition Examination Survey] and to offer an assessment of the uncertainty of those estimates. (JAMA-4)

As noted above, the personal pronoun *we* was frequently used to present hypotheses and research objectives. Modal Finites and modal Adjuncts were rarely observed. One exception, however, was the expression of possibility in the hypothesis of (58), i.e., *leading us to hypothesise that socioeconomic position might be associated with...*

References were used only occasionally in step 3a, and in the case of (58), for example, it could be argued that the two opening clauses of the first sentence, both of which include references to previous studies, are in fact more closely related to move 2.

Step 3b gave a brief description of the study material and methodological procedure. It was less commonly observed than step 3a, appearing in 13 of the 17 articles (76%). In (60), the step describes modifications to a previously used study design; in (61), it briefly describes a data-collection questionnaire; and in (62), the study material. In all articles in which it was identified, step 3b supplemented step

3a and served to clarify the study objective in relation to the study material or methodology.

- (60) We used a different approach from that used previously. Our method was derived from the methods used with the Gail model^{10,11} for predicting breast cancer risk. This method allows us to account for confounding and effect modification, and we provide SEs [standard errors] for the estimates. We also use only data from nationally representative samples with measured heights and weights. We use this approach to make estimates of excess deaths associated with different levels of body weight in the United States in 2000. (JAMA-4)
- (61) The PedsQL is a short survey instrument assessing physical, emotional, social, and school functioning. Strengths of the PedsQL include the availability of parallel reports by a parent-proxy and a child and relatively low ceiling effects, which occur if most participants achieve a near perfect score on a questionnaire. (JAMA-1)
- (62) We identified and quantified risk factors for obesity at age 7 years in children who were participating in the Avon longitudinal study of parents and children (ALSPAC). The study concerns a large contemporary cohort in which confounding variables are being considered and potential risk factors are being tested simultaneously. (BMJ-4)

As demonstrated in the examples above, Themes were generally unmarked in step 3b. Subjects referred to the authors or study methodology, and Processes were a combination of relational, mental, and material. The simple-past tense was used when referring to what the researchers had done (e.g., *We identified and quantified risk factors...*, in (62)), and the simple present was used to indicate generalizability (e.g., *Strengths of the PedsQL include...*, in (61)).

Again, as demonstrated in step 3a, references were rarely observed in step 3b. In (60), they refer the reader to a previously used methodology, avoiding the need to explain them at any great length in this generally short step. No evidence of modality was noted in step 3b.

5.2.4 Summary of Introduction Section

The Themes, Process types, choice of tense/voice, personals and modality, and reference types for moves 1–3 of the Introduction section are summarized below in Table 5.2. These particular features of the discourse were selected due to their

apparent variation throughout the text. They were noted during the initial moves analysis, and were then analyzed separately as part of the lexicogrammatical analysis (stage 4 of the methodological procedure, see section 3.3.2) in an attempt to discern possible patterns of variation across steps, moves, and article sections. Other features such as cohesive ties, clause relations, etc., which are discussed throughout, are not summarized in this or subsequent moves summary tables (Tables 5.3–5.5, sections 5.3.5, 5.4.2, and 5.5.4).

Table 5.2. Moves summary for Introduction sections of corpus articles

Step no.	Theme	Process	Tense, voice	Personals	Modality	References
1a	unmarked	relational	s-pre/s-past active	–	<i>may</i>	NI
2a	multiple	relational	s-pre/pr-per active	–	<i>may</i>	NI/I
2b	multiple	rel./mental	s-pre/pr-per active	–	<i>may</i>	N/I
3a	UM/mult.	rel./mental	s-past active	<i>we</i>	–	–
3b	unmarked	rel./ment./mat.	s-past/s-pre active	<i>we</i>	–	–

Absolute values are not presented. The above summary is observational and based on the examples described in sections 5.2.1–5.2.3. In instances in which combinations are presented, those listed first were the most frequently observed. For example, “rel./mental” indicates that relational Processes were more frequently observed than mental Processes, but that the occurrence of mental Processes was considered significant enough to be classed as a “characteristic” of the step. Abbreviations: UM/mult.: unmarked/multiple Themes; rel./ment./mat.: relational/mental/material clauses; s-pre: simple present; s-past: simple past; pr-per: present perfect; NI/I: non-integral/integral references

5.3 Analysis of Methods Section

The primary rhetorical function of the Methods section was to describe the selection of the study material and to recount the procedure and techniques used to analyze the material. This was done in terms of three structural moves, labeled moves 4–6.

5.3.1 Move 4

Move 4, which can be summarized as “describing the material/study participants,” was identified in all articles, and always directly followed move 3 (see Table 5.1). This phase of discourse consisted of six potential steps.

Step 4a provided details of the size of the study sample and was identified in 14 of 17 articles (82%). In (63), the step begins by naming and briefly describing the study, and by referring to a previous work regarding more precise details of the

cohort (*The Avon longitudinal study... and This study is described in detail elsewhere*). A similar technique, in which authors referred to previous publications for more detailed information concerning study participants and/or methodological procedures (see also step 5a, “study measurements,” section 5.3.2), was used in an additional seven articles (47% of the corpus). All these studies were part of ongoing longitudinal research programs. In the case of (63) below, a summary of the study cohort is provided. This is introduced by a textual Theme (underlined), the summative conjunctive Adjunct *Briefly*. The same Theme was also used in JAMA-1 (not shown). Note that the accompanying experiential Theme (also underlined) contains specific details of the size of the cohort as well as an inclusion criterion (based on expected delivery dates) for study participants (see step 4c, “selection criteria,” below). The two subsequent Themes (underlined) also highlight participant numbers.

- (63) The Avon longitudinal study of parents and children is a longitudinal birth cohort study of the determinants of development, health, and disease during childhood and beyond. This study is described in detail elsewhere.⁶ Briefly, 14 541 pregnant women with an expected date of delivery between April 1991 and December 1992 were enrolled; 13 971 of their children formed the original cohort. About 85–90% of eligible mothers took part in the study. (BMJ-4)

In the example above, the simple present is used to refer to the study, and this was the preferred tense in similar ongoing research projects, e.g., *The NGHS is a multicentre longitudinal study of obesity development in 1213 black and 1166 white girls* (LAN-3). However, the simple-past tense was used when describing participant enrollment and study sample size (see (63) above, and (64) and (65) below).

Processes were typically material in step 4a, and were used to describe inclusion and/or participation, often in the passive voice, such that the Goal of the clause was foregrounded in Theme position and the Actor/Agent was omitted (e.g., *pregnant women... were enrolled*). However, existential and relational processes were also employed to report the size of study samples, particularly in the description of study subgroups. Examples of these are given below ((64), existential; (65), relational). Note that although participant numbers are not included in the Theme in (64), existential *There* prepares the reader “for something

that is about to be introduced” (Halliday & Matthiessen 2004: 257), and as such highlights the information that follows it, in this case overall study numbers. Numerical values relating to the size of the study sample, like those below and those in (63) above, were the most typical lexical markers of step 4a. No modality was observed, and references to other studies were scarce.

- (64) There were 9169 pregnancies, resulting in 8181 live births (8030 singletons and 151 twins), 202 stillbirths, and 867 abortions. (NEJM-1)
- (65) 179 subjects were white (40.8 percent), 135 were black (30.8 percent), 120 were Hispanic (27.3 percent), and 5 subjects were classified as other. (NEJM-2)

Step 4b specified the length of the study period, or, in the case of ongoing longitudinal research, the start of the study and relevant follow-up periods. This step was present in all but one of the corpus articles (94%, 16/17).

In the examples below, the unmarked Themes (underlined) are nouns or nominal groups referring to the study’s data set. Information pertaining to the study period appears in the Rheme as a temporal Circumstance.

- (66) Baseline data were collected between April 1999 and March 2001. (BMJ-1)
- (67) Participants were followed up for 15 years and had six clinical examinations: in 1985–86 (baseline or year 0), 1987–88 (year 2), 1990–91 (year 5), 1992–93 (year 7), 1995–96 (year 10), and 2000–01 (year 15). (LAN-2)
- (68) CCG-2961 [Children’s Cancer Group-2961] opened on August 30, 1996, and closed on December 4, 2002. (JAMA-2)

A marked Theme was used on only one occasion (see (69) below), in which a temporal Circumstance, the prepositional phrase *In 1986*, indicates the start of the study.

- (69) In 1986, 51529 male dentists, optometrists, osteopathic physicians, pharmacists, podiatrists, and veterinarians between the ages of 40 and 75 years completed and returned an initial questionnaire that provided detailed information on diet, medical history, and medications. (JAMA-3)

Note that the underlined Subject (Participant: Actor) in (69) also contains a reference to the size of the study sample (step 4a)—and that the Complement (Goal) also comprises elements of step 4d, “type of data collection.” Indeed, steps 4a and 4b frequently appeared together (77%, 10/13 articles in which both moves were identified) and were difficult to identify as separate steps in some articles. In (70) below, for example, information regarding the size of the study sample (step 4a) appears as an elaborating hypotactic clause (*when 121,700 female registered nurses...*) between two independent clauses describing the study period (step 4b). Note that both step-4b Themes (underlined) are unmarked and refer to the study sample, and that the Circumstances of these clauses relate to the study start and follow-up period, as in examples (66)–(68).

- (70) The cohort of the Nurses’ Health Study was established in 1976, when 121,700 female registered nurses who were 30 to 55 years old completed a mailed questionnaire about their medical history and lifestyle. Women in the study have provided current information regarding lifestyle and health conditions every two years since 1976.
- (NEJM-3)

In cases such as those above, in which two or more steps were identified in the same Theme (e.g., (69), steps 4a and 4b), clause (e.g., (69), steps 4a, 4b, and 4d), or clause complex (e.g., (70), steps 4a and 4b), the presence of multiple steps was registered. Unlike moves, steps rarely represented new phases of discourse, or even new Themes for that matter (e.g., step 4b), but they were nevertheless integral to the move they realized.

As indicated above, the Processes of step 4b were predominantly material and given in the simple past. Omission of the authors as Actors in carrying out study procedures resulted in a relatively high number of passive clauses compared to previous steps, and no other form of author presence, i.e., modality, was observed. References to other studies were not used in step 4b.

Step 4c described the necessary selection criteria for the study material. Such criteria were identified in 88% (15/17) of the corpus articles, and included reasons for both inclusion and exclusion.

In example (71) below, the first Themes of step 4c (including the experiential Theme in the dependent clause of the second sentence) refer to

potential study participants, and exclusion criteria are given in the Rheme. This positioning is reversed in the final sentence of the step, in which the Theme, a series of nominal groups, highlights additional exclusion criteria.

- (71) Eligible subjects were free of uncontrolled hypertension (defined by a blood pressure greater than 140/90 mm Hg); cerebrovascular, cardiovascular, renal, or hepatic disease; and type 1 or 2 diabetes. Women were ineligible if they were pregnant or breast-feeding. The use of medications known to affect body weight, a weight loss of 5 kg or greater in the preceding six months, and the use of selective serotonin-reuptake inhibitors were exclusion criteria. (NEJM-4)

There were also instances in which the Theme contained specific lexis relating to the selection of study participants, e.g., *Exclusion criteria were coronary heart disease, type 1 or type 2 diabetes...* (BMJ-5), or to an Actor involved in the process of inclusion or exclusion, e.g., *We excluded participants from our analysis for the following reasons...* (LAN-2). The latter was also used in the passive form, using the same process, but with the Actor/Agent omitted and the Goal in Theme position: *Patients with Down syndrome, Fanconi anemia [...] were excluded from the study* (JAMA-2). Note that the Theme in this instance is similar to the final underlined Theme in (71) above; the main difference being that the JAMA-2 example is a material clause, while that of (71) is relational. Indeed, the most common Process types in step 4c were material and relational. Material clauses were given in both the passive and active voice. In the passive form, patients and medical conditions as Goal were foregrounded in the Theme position, as in the LAN-2 and JAMA-2 examples in the above paragraph, and no Agent was provided. In the active, the authors as Actor served as the point of departure for the clause (e.g., the first-person pronoun *We* in LAN-2 above).⁴⁶

As demonstrated in the above examples, the past tense was used exclusively in this step. Such a choice seems natural given the description of an act of inclusion/exclusion that invariably took place in the past, at the start of the study. No references to previous studies were observed.

An important step in describing the material/study participants was the description of the type of data collection (step 4d). This step was present in all corpus articles (100%, 17/17)—although in some instances it appeared as part of

⁴⁶ Further analysis concerning the use of personals is provided in section 5.6.

the description of the experimental procedure (move 5, see Table 5.1 and section 5.3.2). Types of data collection and step-specific lexis included questionnaires, interviews, medical examinations, and data retrieval from medical files, as well as a description of how these data were collected.

Themes were rarely marked, and were commonly nouns or nominal groups referring to study participants, databases, or the type of data collection. In (72) below, the Theme of the first sentence, *Participants*, is followed by explicit and implicit anaphoric references to both the participants and the locale at which a questionnaire was administered. The referent and anaphors are underlined, indicating the chain of reference and the series of common Themes. Elided elements have been added for clarity.

- (72) Participants completed a questionnaire and attended a local health centre, where they were interviewed by a research nurse, [where they] were physically examined, and [where they] gave a blood sample. General practitioners' medical records were also reviewed for each participant, and details of diagnoses of cardiovascular disease, diabetes and cancer were extracted. (BMJ-1)

In the second sentence, the Themes refer to a database (*General practitioners' medical records*) and specific data within that database (*and details of diagnoses...*). Note that the Processes in both these clauses are material (*reviewed* and *extracted*), and refer to the act of selection, but Actors/Agents are not specified. In (73) below, however, an Actor (*We*) appears in Theme position, with the type of data collection (*standard questionnaires...*) presented in the Rheme as Goal.

- (73) We used standard questionnaires to maintain consistency in the assessment of demographics... and behavioural information across CARDIA examination visits. (LAN-2)

In some cases, as in (74) below, the type of data collection was foregrounded in Theme position. Again, the clause is material, but no Actor or Agent is specified.

- (74) A longitudinally validated habitual activity questionnaire (HAQ) was used to assess extracurricular... sports and/or physical activities during the past 12 months. (LAN-3)

In only one of the study articles was the opening Theme of step 4d marked, in this case by a temporal Circumstance (*In 1986*) related to step 4b (the study period): *In 1986, women were asked to complete an eight-item questionnaire...* (NEJM-3).

As demonstrated by the examples above, Processes in step 4d were predominantly material, and indicated the collection and use of data from the study sample. The simple past, often in the passive and without the inclusion of an Agent (“agentless passive”), was consistently used throughout. No modality and no references to previous research were recorded.

Step 4e indicated the frequency of data collection. This step appeared in 65% (11/17) of the corpus articles. The relatively low frequency was dependent on the type of study described, since step 4e was unnecessary in studies of a cross-sectional nature, and only appeared in reports of longitudinal data.

An example from LAN-3 (see (75) below) shows that information concerning the frequency of data collection appears as a temporal Circumstance (the prepositional phrase *at years...*) in the Rheme rather than as a marked Theme. This is also the case in (76). Indeed, this was the case in all instances of step 4e in the corpus.

(75) The HAQ was administered at years 1, 3, 5, and annually from years 7 to 10 (a total of seven times). (LAN-3)

(76) Women in the study have provided current information regarding lifestyle and health conditions every two years since 1976. (NEJM-3)

The only marked Themes were those concerning different sample populations or subgroups, such as the underlined examples in (77) below, in which prepositional phrases are used to emphasize data collection in specific samples of the study. There were no Themes marked with a Circumstance (specifically a temporal prepositional phrase or numerative nominal group) that referred directly to the frequency of data collection.

(77) Self-reported weight was updated every 2 years. [...] A version of this food frequency questionnaire [the semiquantitative food frequency questionnaire] has been mailed to study participants every 4 years. [...] In the HPFS and NHS II, thiazide diuretic use

was updated every 2 years. In the NHS I, thiazide use was determined in 1980, 1982, and then every 6 years until 1994, when biennial updates started [...] The participants reported on the interval diagnosis of kidney stones every 2 years. (JAMA-3)

In the above example, step 4e occurred at several points throughout move 4, interspersed between other steps, hence the use of ellipses. In this particular article, the overall study sample was made up of several subsamples, the data from which were collected at different time points.

All the Processes in the above examples are material, i.e., *administered, provided, updated, mailed, determined, started, and reported*.⁴⁷ These Processes refer to the act of collecting and collating data, much like those of step 4d. The frequency of data collection was commonly identified in the Rheme of the clause as a temporal Circumstance. Common tense forms were the simple past when referring to closed studies (e.g., in (77)) and the present perfect when referring to ongoing studies (e.g., in (76)). Actors and Agents were rarely given, resulting in extensive use of the passive, except on occasion when referring to patients (e.g., *The participants reported on...*; JAMA-3). No evidence of modality or reference to previous studies was observed.

A final step in move 4 (step 4f), frequently used throughout the corpus (15/17, 88%), is a declaration concerning study approval and informed consent. Interestingly, although this step most often appeared as part of move 4 (7/17, 41%), it was also included as a separate headed section at the end of the five BMJ articles (5/17, 29%) and as a separate move (after move 6) at the end of the Methods section in three other articles (18%; JAMA-3, NEJM-1, and NEJM-4; see Table 5.1). The step essentially provides a statement declaring “ethical” permission, usually granted by an ethics committee at one or several author institutions, to conduct the study, i.e., permission to perform certain tests/therapies/surgeries on the study material and permission to use what could be potentially sensitive patient data. This permission is usually granted in accordance with internationally agreed protocols, for example the Declaration of Helsinki.⁴⁸ Informed consent concerns written permission by the patients themselves (or their guardians, in the case of

⁴⁷ The verb *reported* is often considered a verbal Process (see section 2.2.1.1). However, in this instance, it refers to the action of completing a hand-written questionnaire and was thus considered material.

⁴⁸ See URL: www.wma.net/e/policy/b3.htm (accessed June 15, 2007).

children) to be included in the study. This is granted with the proviso that patients are informed about the objectives of the study and any impact inclusion in the study might have on their health or wellbeing (ICJME 2006: 10).

As can be seen from the examples below, step 4f was highly standardized.

- (78) The All India Institute of Medical Sciences approved the study. Informed consent was obtained from each subject. (NEJM-1)
- (79) The Yale University School of Medicine human investigation committee approved the study. Written informed consent from parents and written assent from children (where appropriate) and adolescents were obtained. (NEJM-2)
- (80) Written informed consent was obtained from the parents until the child became 18 years old, when she also gave written consent, and study protocols were approved by the institutional review boards at all centres. (LAN-3)
- (81) The study was approved by the ethics in human research committee of the Royal Children's Hospital, the Victorian Department of Education, the Catholic Education Office, and the Independent Schools Office. A parent-proxy provided written informed consent. (JAMA-1)

Commonly occurring Themes included *The study*, *Written informed consent*, and *Local ethics committee approval* (BMJ-5), as well as the names of institutional bodies approving the study, and they were always unmarked. In the example below from LAN-1, the authors foreground themselves in the process of obtaining written informed consent (unmarked Theme *We*). However, this was the only step-4f example in the corpus in which the authors as Actors/Agents were not omitted.

- (82) We obtained written informed consent for each assessment. Our study was approved by the Otago Ethics Committee. (LAN-1)

The most common Process type used in step 4f was material, and the verbs *approve* and *obtain* were the most frequently observed markers of the step. Other important lexis included nouns and nominal groups indicating the names of institutional committees and referring to study participants. The past tense was used in all instances, and the majority of clauses were passive. However, unlike the other

steps of move 4, passive clauses usually included an Agent (e.g., *the Otago Ethics Committee* or *patients, children*, etc.). No modality or references to previous studies were observed.

5.3.2 Move 5

Move 5, “describing the experimental procedure,” was identified in all corpus articles, and always directly followed move 4 (see Table 5.1).⁴⁹ Move 5 was realized by four potential steps, as well as by a less frequently observed fifth step.

Step 5a was the most commonly observed step in move 5 (16/17, 94%). This step described study measurements, and included measurement protocols, the apparatus used (and its precision), and, in some cases, the frequency with which measurements were taken. Examples are given below.

- (83) The NGHS study protocol has been previously described.³² BMI (weight in kg divided by height in m²) was derived from annual measurements of height and weight. Skinfold thickness was measured with Holtain calipers at the triceps, subscapular, and suprailiac sites by centrally trained examiners, and these measures were summed. (LAN-3)

- (84) The subjects consumed a diet containing at least 250 g of carbohydrates per day for three days before the study and [they] refrained from vigorous physical activity. They were evaluated at 8 a.m., after a 12-hour, overnight fast. Their weight and height were measured, and their BMI was calculated. Blood pressure was measured three times while the subjects were seated, and the last two measurements were averaged for analysis. (NEJM-2)

None of the above Themes are marked, and all act to foreground measurement techniques, the measurements themselves, or the study participants. Measurement frequencies were presented in the Rheme as a Circumstance, e.g., the numerative nominal group *three times* in (84). Note also that, as in the excerpt from BMJ-4 in step 4a (“size of study sample,” see example (63), section 5.3.1), (83) refers the reader to a previously published work for more detailed information concerning the study protocol.

⁴⁹ The order stability of move 5 was 94%. This reduced stability was due to the moves that followed it rather than the preceding move 4.

In (84), the Participant in the first Theme in step 5a refers to an Actor, *The subjects*, and the study participants are foregrounded as either Actor or Goal (e.g., *The subjects consumed...* (Actor) or *They were evaluated...* (Goal)) in all but the last sentence. The researchers themselves, however, are omitted or, as in the case of example (83), presented as a backgrounded Agent in the Rheme (*centrally trained examiners*). Indeed, the first person was rarely used in step 5a, and almost all material clauses were written in the passive. This is interesting since, in most instances, the step presumably describes how the researchers themselves carried out study measurements. One exception to this, however, was BMJ-4 (see underlined Themes in (85) below).⁵⁰ Note that this example also contains details of the precision of the apparatus, given each time in the Rheme as a Circumstance (a qualifier to the nouns *height* and *weight*).

- (85) We measured height to 0.1 cm using the Harpenden stadiometer (Holtain; Crymych, Wales). Weight was measured to 0.1 kg. From these values we calculated the body mass index (weight (kg)/height (m)². (BMJ-4)

The past tense was used throughout step 5a, and Processes were commonly material, e.g., *measured*, *calculated*, and *evaluated*.⁵¹ Lexical markers such as the above-mentioned Processes and the names of measuring devices, e.g., *stadiometer* and *calipers*, typified the step. No modality was observed, and the use of references was limited to previously described study protocols (e.g., (83) above).

Step 5b, the definition of specific terms, was less common than step 5a. This step was observed in 10 (59%) of the 17 corpus articles, and served both to clarify standard, central terms used in the study (often definitions of measurements) and to explicitly define less commonly recognized terms. The length of this step varied considerably, from a single-sentence statement (e.g., JAMA-2) to a separately headed two-paragraph section (e.g., NEJM-2, headed “Definitions”). Examples are provided below.

- (86) AML [acute myeloid leukemia] was classified according to French-American-British criteria.⁹ (JAMA-2)

⁵⁰ Further analysis concerning the use of personals is provided in section 5.6.

⁵¹ The verb *evaluated*, as used in (84), refers to the taking of specific measurements rather than a general consideration of the patients’ condition, and was hence considered a material Process rather than a cognitive mental Process.

- (87) We defined obesity on the basis of a threshold BMI z score of 2.0 or more, adjusted for age and sex. The subjects were then classified as moderately obese (a z score of 2.0 to 2.5) or severely obese (a z score above 2.5). Elevated systolic or diastolic blood pressure was defined as a value that exceeded the 95th percentile for age and sex.¹⁹ [...] Impaired glucose tolerance was defined as a glucose level greater than 140 mg per deciliter (7.8 mmol per liter) but less than 200 mg per deciliter (11.1 mmol per liter) at two hours.²¹ (NEJM-2)

The terms to be defined were usually foregrounded as Phenomena in the Theme position (see underlined Themes above), necessitating use of the passive and allowing the Senser (presumably the authors) to be omitted. Other less commonly observed Themes were related to study participants (e.g., *The subjects*, Phenomenon, in (87)) or to the researchers themselves (e.g., *We*, Senser, in (87)). Indeed, mental Processes were the most common Process type observed in step 5b, particularly the verbs *define* and *classify*, which were also the most apparent lexical markers of this step, along with their nominalized forms (*definition* and *classification*) and the specific terms being defined. Like step 5a, this step appeared predominantly in the past tense. However, there were exceptions. In JAMA-5, for example, the present tense was used to make direct reference to defined groups commonly appearing throughout the article: *We refer to BMI groups of <25, 25 to 29.9, and ≥30 as “lean,” “overweight,” and “obese,” respectively* (JAMA-5).⁵² No modality was observed in step 5b. References to other studies, although relatively infrequent, referred to standardized definitions, usually published by specialist medical organizations. The references in (87), for example, cite a “Task Force” report on blood pressure values among children and adolescents, and an “Expert Committee” on the classification of diabetes mellitus.

The use of step 5b, and its relatively low frequency (59%) in the study corpus, may depend on how authors interpret the level of background knowledge of their intended readership. If the authors consider a term to be common knowledge, they may feel that defining it is unnecessary, and perhaps even condescending to their primary and secondary readerships (see section 4.1.2). However, authors (or editors) may insist on defining the term, regardless of whether it is in common

⁵² Note that, in this particular instance, unlike the other examples of step 5b, the Process (*refer*) is verbal rather than mental.

usage, so as to ensure that the methodology used in the study is as transparent as possible in order to allow other researchers to reproduce their results (ICMJE 2006: 27; see Appendix). An interesting example of this could be in the use and definition of the term *body-mass index* (BMI), a controversial measurement central to the classification of over- and underweight patients. The term appeared in every corpus article, but it was only specifically defined in 10 (59%) of them.

Step 5c involved categorizing the study cohort into study subgroups, and was observed in 12 (71%) of the corpus articles. Systematic and/or random categorization was performed in order to compare different characteristics or factors across the overall study sample, for example a comparison of smokers versus nonsmokers, or comparisons among different treatment groups.

(88) Children were classified into 3 mutually exclusive categories of not overweight, overweight, or obese according to the sex- and age-specific cut-off points developed by the International Obesity Task Force.¹⁶ (JAMA-1)

(89) At each test centre, we stratified participants by sex and allocated them to a group [four different commercial weight-loss programs] using random number generation. (BMJ-5)

The above Themes refer to either study participants (Phenomenon) or those responsible for the categorization (Sensor). The Theme in example (89), which is marked with a locational Circumstance (the prepositional phrase *At each test centre*), highlights the fact that the same procedure was carried out by the researchers at all centers involved in the study (five in this case). In the first two Themes in (90) below, the researchers are foregrounded and the nature of the selected subsample is described in the Rheme.⁵³ In the third Theme, the categorization itself, which is marked by a temporal Circumstance, is highlighted.

(90) We randomly selected a subsample of children from the last six months of recruitment (children in focus group), aged from 4 months to 5 years, and [we] invited their parents to bring them in for regular physical examinations. From age 7 years onwards these examinations were extended to the whole cohort. (BMJ-4)

⁵³ The experiential Theme of the second clause is elided but implies repetition of the personal pronoun in the previous clause.

The Processes in step 5c were predominantly mental, e.g., *stratified*, *classified*, and *categorized*, and these Processes were also the defining lexis in the step. Other notable lexical markers were nominalizations such as *categorization* and *classification*, as well as circumstantial Adjuncts such as *randomly* and *selectively*. The past tense was used throughout, and clauses were frequently passive. No modality was observed, and the few instances of references to other studies were concerned with standardized methods of categorization (see (88)).

Step 5d was a relatively low-frequency step in the corpus (10/17 articles, 59%). This step was used to define the study's anticipated endpoints or outcomes. These typically referred to patient status at the end of the study period, such as the presence or absence of a particular condition, and survival and mortality rates. Three articles (JAMA-5, NEJM-3, and NEJM-4) contained separate headed sections referring to study endpoints/outcomes, while the others provided relatively short, usually single-sentence statements.

- (91) The main outcome measures were remission status after courses 1 and 2 of chemotherapy, overall survival, and treatment-related mortality. (JAMA-2)
- (92) The primary outcome was an incident of kidney stone accompanied by pain or hematuria. (JAMA-3)
- (93) The primary endpoint in this analysis was death by any cause. We also conducted analyses according to the causes of death, which were divided into cancer (ICD-8 codes 140.0 through 207.9), cardiovascular disease (ICD-8 codes 390.0 through 458.9 and 795.0 to 795.9), and all other causes. (NEJM-3)

In the above examples, all but one of the Themes refers directly to an outcome or endpoint. The exception, the second Theme in (93), foregrounds the authors (the Actor *We*), while a secondary study endpoint is identified in the Rheme (the Goal *analyses according to causes of death*). Nominal groups referring to study outcomes such as those in the Themes above (e.g., *The main outcome measures*) were typical markers of step 5d.

Unlike the other steps of move 5, step 5d consisted mainly of relational clauses, that is, clauses that “serve to characterize and to identify” (Halliday & Matthiessen 2004: 210)—in this case, clauses that identified study endpoints. In

(92), for example, the Token *an incident of kidney stone accompanied by pain or hematuria* is assigned a Value, *The primary outcome*, by means of an identifying relational Process (*was*), and this Value is foregrounded in the Theme position. An exception to this is the material clause in (93) (*We also conducted analyses...*; see above). Regardless of Process type, the past tense was used throughout step 5d. No modality was observed, and no references to other studies were made.

One step, which is not summarized in the moves analysis above due to its relatively low frequency (6/17, 35%), described certain methodological modifications made by the researchers. These modifications applied to both the description of the experimental procedure (move 5) and the description of the data-analysis procedure (move 6, see below), and justification for the changes was provided.

Themes were generally unmarked and referred either to the methodology or to the researchers. The opening Theme of (95), however, is marked by a thematic non-finite hypotactic enhancement clause, which foregrounds the reason for adjustment. A combination of the active and passive voice was used, and the past tense was used in all cases. Typical lexical markers for this step included the Processes *adjust* and *modify*. No modality or references to other studies were observed.

(94) All analyses were adjusted for sex. Analyses involving VO₂max were also adjusted for bodyweight. (LAN-1)

(95) To account for the fact that a given weight gain in a heavier individual represents a smaller fractional increase than the same weight gain in a lighter individual, we adjusted our weight change analyses for baseline weight. (JAMA-3)

There may be several reasons for the relatively low frequency of this step: 1) no modifications (according to a standard protocol, see step 5a) were made; 2) the modifications made were not deemed significant/relevant to the description of the methodology; or 3) the methodology was described in full, rather than referring to a standard protocol and the subsequent changes made, so that both “standard” and “modified” procedures were included without the need to highlight any “changes” in procedure.

5.3.3 Move 6

Move 6, “describing the data-analysis procedure,” was present in all but one of the corpus articles (16/17, 94%; not used in BMJ-2). In the sixteen articles in which it was used, the move always followed on from move 5 (adjusted order stability 100%; see Table 5.1). This move was realized by two potential steps, as well as a third less frequently observed step.

Step 6a was observed in all instances of move 6 (16/16, 100%). Its function was to describe the statistical test techniques used and to provide reasons for their use.

- (96) We carried out a multivariable analysis for the prevalence of obesity [...] using multivariable binary logistic regression models. Firstly, owing to the strong association between maternal education (as a proxy for socioeconomic status) and childhood obesity, we assessed whether the effect of potential risk factors was confounded by the mother’s education. Secondly, in an effort to reduce colinearity and to minimise the likelihood of producing misleading estimates for the variables (data not shown), we analysed putative risk factors for childhood obesity [...] simultaneously within each of the four risk factor groups. [...] Finally, risk factors that were independently significant ($P < 0.10$) at the within group stage were then entered into a final model in which we analysed all variables simultaneously. (BMJ-4)

In example (96) above, most of the Themes feature the authors themselves (the Senser *we*), and in two instances fronted non-finite dependent clauses (underlined) serve to highlight the reasons for carrying out specific tests. A similar type of foregrounding is demonstrated in (97) (see underlined examples below), in which a locational Circumstance, a finite dependent clause, and a temporal Circumstance highlight certain conditions for the authors’ use of specific statistical tests. In the first two sentences, reasons for analysis are provided in the first of two independent enhancing clauses, while the second describes what was done. A third clause in the second sentence, the elaborative hypotactic clause *which uses...*, provides further characterization. In (98), the authors (Senser *We*) and a statistical test (Goal *The Mantel extension test*) are foregrounded. Information concerning the use of a particular test (*using Cox proportional hazards regression*) and the purpose of a specific test (*to evaluate linear trends...*) follows in the Rheme.

- (97) Attrition rates did not differ between centres, so we analysed data from all participants

together. Baseline weight correlated with total weight loss ($r = 0.33$, $P < 0.001$) and was therefore used as a covariate in the analysis of weight loss over time, which uses repeated measures analysis of covariance. In the intention to treat analysis, we used ANOVA to examine differences between groups; where ANOVA indicated a significant group effect, we performed post hoc pairwise testing with Tukey's HSD (honestly significantly different) test. Before parametric testing, we assessed homogeneity of variance with Levene's statistic and tested for normal distribution; we used Welch's F statistic if variance of the dependent variable was not equal across groups. (BMJ-5)

- (98) We determined the relative risk (RR) of kidney stone formation for each category of body size compared with the referent category using Cox proportional hazards regression. The Mantel extension test was used to evaluate linear trends across categories of body size. (JAMA-3)

Mental and material Processes and the past tense predominated step 6a, and a combination of the passive and active voice was used. Prominent lexical markers included the names of statistical tests (e.g., *the Mantel extension test*), the use of the mathematical functions P and r (probability and correlation), Processes such as *analyzed*, *evaluated*, and *assessed*, and nouns referring to *analysis* and *statistics*. No modality was observed in this step, and no references to previous publications were featured.

A less commonly occurring step, step 6b (11/16, 69%), provided details of the computer software used for the analysis. This step was invariably short, typically one sentence, and appeared as a final step to move 6 in 82% (9/11) of the articles in which it was observed. The examples below demonstrate the relatively standardized way in which step 6b was presented. They also demonstrate typical unmarked Themes, i.e., the data to be analyzed (Phenomenon), the authors (Actor), and the analysis (Goal). On no occasion was the computer software itself foregrounded in the Theme, e.g., "SAS statistical software was used to analyze the data."

- (99) Data were analyzed using SPSS statistical software (version 11.5, SPSS Inc, Chicago, Ill). (JAMA-1)

- (100) We did all analyses with SAS statistical software version 9 (SAS, Cary, NC, USA). (LAN-2)

- (101) All analyses were performed with the use of SAS software (version 8, SAS Institute).
(NEJM-2)

Most Processes were material, and the past tense and passive were used throughout. Lexis indicative of step 6b included the names of computer software packages (as well as manufacturer names and geographic locations) and nouns/nominal groups such as *statistical software* and *analysis*. No modality or references were noted.

A third step, not listed in the moves-analysis summary, was occasionally observed (5/16 articles, 31%). This step recounted the exclusion of data from the final analysis, which was usually as a result of incomplete information from patient questionnaires or missing data from patient records. Examples are provided below.

- (102) To compensate for the tendency for bad news (ie, deaths and relapses) to be reported sooner than ongoing follow-up for patients in continuing remission, events such as deaths and relapses were censored on July 9, 2003, 6 months before data cutoff. (JAMA-2)
- (103) If weight data were missing for more than 2 consecutive time periods, no value was imputed... Missing values for height, waist circumference, and hip circumference were assigned to missing categories... If complete information on diet was missing at the start of a time period, the participant was excluded for that time period. (JAMA-3)
- (104) For all analyses, we excluded women who had reported cancer or cardiovascular disease at baseline. (NEJM-3)

In (102), a causal non-finite hypotactic clause highlights the necessity to exclude certain data; data which is the Theme of the dominant clause, *events such as deaths and relapses*... In (103), two conditional hypotactic clauses foreground the provisions for data exclusion, and the marked Theme in (104) indicates that the exclusion applies to all data analyses. The most common Processes were material, and these were always given in the past tense. Characteristic lexical markers were similar to those for step 4c (study selection criteria), with a focus on *exclusion* and *missing data*. Clauses were frequently agentless passives, and causal-conditional

conjunctions such as *if*, *when*, and *because* were commonly used to indicate reasons and conditions for the exclusion of certain types of data.

The low frequency of this step would appear to depend on 1) whether authors perceive a need to exclude data from their analyses, 2) whether they consider it necessary to inform readers of specific data exclusion, or 3) whether or not such exclusions, regardless of when they take place during the study period, are described in step 4c. Indeed, the content of this step was clearly similar to that of step 4c, but differed notably in terms of its use of Theme and commonly identified Processes (see summary in Table 5.3). As with the other steps in move 6, no modality or references to previous publications were observed.

5.3.4 *Additional Move*

An additional move, “stating the role of the funding source,” was present in the Methods section of the three LAN articles. This move appeared under a separate subheading at the end of the Methods section, and is used consistently in all original research articles published in LAN.⁵⁴ The move acts as a disclaimer for the journal, and attempts to make clear any possible conflict of interest the authors might have in publishing their research article. The other three source journals publish similar statements, but these are provided at the end of articles, at a reduced font size, separate from the main body of text (see example in section 4.2.1).

In the case of LAN, the role-of-the-funding-source move was standardized to such a degree that all three articles included the same text, namely *The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. An additional sentence—The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication—was published in LAN-2 and LAN-3. Both these articles were published in the 2005 volume of the journal; LAN-1 was published in 2004.*

⁵⁴ See “Information for Authors” at URL: www.thelancet.com/authors/lancet/authorinfo (accessed August 20, 2007).

5.3.5 Summary of Methods Section

A summary of selected lexicogrammatical features in moves 4–6 is provided below in Table 5.3.

Table 5.3. Moves summary for Methods sections of corpus articles

Step no.	Theme	Process	Tense, voice	Personals	Modality	References
4a	UM/mult.	material	s-pre/s-past passive	–	–	–
4b	unmarked	material	s-past passive	–	–	–
4c	unmarked	mat./rel.	s-past pass./act.	<i>we</i>	–	–
4d	unmarked	material	s-past passive	–	–	–
4e	unmarked	material	s-past/pr-per passive	–	–	–
4f	unmarked	material	s-past pass. w Agent	–	–	–
5a	unmarked	material	s-past passive	–	–	few NI
5b	unmarked	mental	s-past passive	–	–	few NI
5c	UM/marked	mental	s-past passive	–	–	few NI
5d	unmarked	relational	s-past active	–	–	–
add. step	unmarked	material	s-past act./pass.	<i>we</i>	–	–
6a	UM/TC/MT	ment./mat.	s-past act./pass.	<i>we</i>	–	–
6b	unmarked	material	s-past passive	–	–	–
add. step	TC/MT	material	s-past passive	–	–	–
add. move	unmarked	relational	s-past active	–	–	–

Absolute values are not presented. The above summary is observational and based on the examples described in sections 5.3.1–5.3.4. In instances in which combinations are presented, those listed first were the most frequently observed. Abbreviations: UM/mult.: unmarked/multiple Themes; TC: thematic clause; MT: marked Theme; mat./rel./ment.: material/relational/mental clauses; s-pre: simple present; s-past: simple past; pr-per: present perfect; NI: non-integral references; act./pass.: active/passive voice; add. step/move: additional low-frequency step/move

5.4 Analysis of Results Section

The primary rhetorical function of the Results section was to report the data obtained in relation to the study's methodology. This was done in terms of one essential structural move, as described below.

5.4.1 Move 7

The overall hyperTheme of move 7 can be summarized as “reporting observations.” This move was identified in all corpus articles, and had an adjusted order stability of 88% (see Table 5.1).⁵⁵ This move was realized by a series of four potential steps, as well as by a less frequently observed fifth step.

Step 7a, reference to non-verbal material, was observed in all of the study corpus articles (17/17, 100%). This particular step was used to refer to and comment on “non-verbal material” (Swales 1990: 115) such as tables and figures.⁵⁶ The step frequently appeared at the start of the Results section in reference to baseline study-cohort characteristics (11/17 articles, 65%), although it was also observed throughout the Results section when several figures or tables were presented. Themes in step 7a were predominantly unmarked nominal groups referring to non-verbal material or the data reported in them (see examples below).

- (105) Table 1 shows the characteristics of the 886 men and 640 women in the current sample. Most were married, college graduates (with a bachelor’s degree or above), and not in manual employment. Few women drank alcohol or smoked tobacco. Almost half the subjects were overweight according to the conventional definition,¹¹ and nearly two thirds were overweight when the Asian cutoff value was used.¹² (NEJM-1)
- (106) Demographic and weight characteristics of the sample are shown in TABLE 1. The sample reflects population characteristics in terms of sociodemographics, but as noted in previously published data, children retained in the cohort had a lower mean BMI at baseline than children lost to follow-up (16.9 vs 17.5; $P < .001$).²⁰ (JAMA-1)

Reference to non-verbal material was foregrounded in the Theme position in 29 of 110 instances (as in (105), *Table 1 shows...*). Otherwise, it was presented in the Rheme (8/110; e.g., (106), *... are shown in Table 1*), or, in most cases, as a parenthetical addition at the end of the clause (73/110; see (107) below).

⁵⁵ The order stability of move 7 may be deceptive. Although it was flanked by a number of different moves (5, 6, 8, and 9), it was in fact the only move in the Results section, and could thus be considered entirely “stable,” at least within the section.

⁵⁶ See Table 3.2 (section 3.2) for details of the frequencies of occurrence of tables and figures in the study corpus.

- (107) Age-adjusted fast-food frequency was relatively stable over time in black people but fell in those who were white ($p < 0.0001$ for ethnic origin-time interaction; table 1).

(LAN-2)

The most commonly used Processes were relational (*was, were, are, had*), and non-verbal material was presented either as a Participant:Carrier (see (105)) or as a spatial Circumstance (see (106)). A combination of the past and present tenses was used: the past to indicate the results observed, the present to refer to tables and figures within the article (see (106) and (107) above). Lexis referring to tables and figures, e.g., *Table 1, Figure 2*, etc., was the most apparent characteristic of step 7a. In addition, determiners such as *most, few, almost half, two thirds*, etc. (see (105)) were frequently used to describe and quantify the data presented. References to previous studies were observed, usually with reference to published standards, as a means of comparison (e.g., (105)). No modal Finites, modal Adjuncts, or personals were identified.

Step 7b stated the main findings of the study and was observed in 14 (82%) of the study articles. The frequency of this step, however, may be misleading. In the three articles in which step 7b was not demonstrated, there appeared to be little or no differentiation between major study findings and the associations/correlations of the statistical analyses (step 7c, see below). This is not to say that the articles did not provide details of the study findings—on the contrary, every article reported such findings—rather that a clearly discernible step or thematic change could not be identified.

In (108), the main findings are presented in a paragraph immediately preceding the results of the statistical analysis (step 7c). These findings are directly related to the primary outcome of step 5d and the study aim of step 3a, in this case the incidence of kidney stones, which is foregrounded in the second Theme in the example below. The marked opening Theme (a temporal Circumstance) highlights the extent of the overall data set from which the primary-outcome findings were derived.

- (108) Over a combined 2808334 person-years of follow-up, we documented 4827 new symptomatic kidney stones: 1609 in the HPFS, 1687 in the NHS I, and 1531 in the NHS II. The unadjusted incidence of stones was 301 per 100000 person-years in the HPFS, 117 per 100000 person-years in the NHS I, and 183 per 100000 person-years in

the NHS II.

(JAMA-3)

Similarly, in (109) below, findings in relation to the primary aim of the study (the prevalence of metabolic syndrome among obese children and adolescents) are foregrounded.

- (109) Values for glucose, insulin, insulin resistance [...] as well as the prevalence of impaired glucose tolerance, increased significantly with increasing obesity, whereas HDL cholesterol and adiponectin levels decreased with increasing obesity. [...] The overall prevalence of the metabolic syndrome was 38.7 percent in moderately obese subjects and 49.7 percent in severely obese subjects. (NEJM-2)

Material and relational Processes are used in both the above examples. Material Processes were used to document what was observed, e.g., that certain values *increased*, while relational Processes characterized the main findings, e.g., in terms of an incidence rate or a percentage. The simple-past tense was used throughout step 7b, and characteristic lexis typically referred to terms identified in the study aim or primary outcomes of steps 3a and 5d. No references to previous studies were made in stating the main study findings, and there was no evidence of modality.⁵⁷

In 16 (94%) of the corpus articles, step 7b was followed by a series of statistical associations and correlations (step 7c) related to the primary study observations. These associations were used to identify risk factors and trends across the study material, and represented the results of the statistical test techniques of move 6 (section 5.3.3).

- (110) Increasing birth weight was independently and linearly associated with increasing prevalence of obesity at age 7. Obesity at age 7 was also significantly associated with maternal smoking between 28 and 32 weeks' gestation, with some indication of a dose response (chi-square test for linear trend 27.17). (BMJ-4)
- (111) Child and adolescent viewing [of television] (age 5–15 years) correlated with lower childhood socioeconomic status (n=1013, $r=0.31$, $p<0.0001$), increased parental smoking (n=998, $r=0.11$, $p=0.0005$), higher maternal and paternal body-mass indices

⁵⁷ Note that, in (109), *significantly* and *moderately* are not considered comment Adjuncts, since they refer to statistical ($P<0.05$) and normalized (BMI z score 2.0–2.5) data, respectively.

($n=839$, $r=0.09$, $p=0.0086$; $n=798$, $r=0.11$, $p=0.0013$, respectively), and higher body-mass index at age 5 years ($n=996$, $r=0.11$, $p=0.0004$). Physical activity at 15 years of age did not correlate with overall child and adolescent viewing, but [it] was significantly correlated with fewer hours of adolescent viewing ($n=825$, $r=-0.09$, $p=0.0101$). (LAN-1)

In the examples above, major study parameters are foregrounded in the Theme position (underlined), i.e., early-life risk factors for obesity (in (110)) and the effect of child television viewing on adult health (in (111)). These Themes (primarily Tokens) are then related to other study variables (Values) by Processes such as *correlated* and *associated*, which are in turn typically modified with adverbs such as *significantly*, *linearly*, and *inversely*. Such relational Processes and manner/circumstantial Adjuncts were characteristic of step 7c and described the nature of the relationship between certain parameters. Mathematical functions referring to probability and correlation (P and r) were commonly observed (15/16 articles, 94%), and the simple-past tense was used consistently throughout the step. No modality was observed, and references to others studies were not made.

In 14 (82%) of the corpus articles, step 7d described adjustments made to the statistical analysis. The purpose of these adjustments was to check the extent to which certain variables affected the overall study findings.

- (112) After adjusting for all listed variables included in the model (table 2), changes in HAQ [habitual activity questionnaire] score (independent predictor variable) significantly affected changes in BMI in an inverse relation. Each later year of age at menarche was associated with less gain in BMI, more for white than for black girls. Cigarette smoking was also significantly inversely associated with gain in BMI but only for white girls. (LAN-3)
- (113) After adjustment, overweight and underweight groups were still less likely to survive than middleweight patients, and HR [hazard ratio] for treatment-related mortality was even higher than in univariate analysis. (JAMA-2)

In the first underlined example in (112), a temporal hypotactic enhancement clause highlights the adjustment of study variables, and the Theme of the dominant clause which it enhances (*changes in HAQ score...*) refers to one of the adjusted variables that had a “significant” effect on the study outcome. Unmarked Themes

follow, in which other relevant adjusted variables are foregrounded. Similarly, in (113), the first Theme is marked by a temporal Circumstance, *After adjustment*, and Subjects (Carriers) refer to study subgroups and the study's main outcome measurement, mortality in over- and underweight children.

As demonstrated by the examples above, and in (114) below, Circumstances and hypotactic clauses were commonly thematic, but they were also backgrounded in some instances. In (115), for example, emphasis is on the main study outcome and the effect the adjustments had, as indicated by the modified material Process *increased significantly*. The adjustments themselves follow in the Rheme.

- (114) When we restricted the analyses to women who had never smoked, in order to minimize confounding by this major cause of death,²¹ we observed a direct monotonic relationship between the body-mass index and mortality. (NEJM-3)
- (115) The prevalence of the metabolic syndrome increased significantly with increasing insulin resistance (P for trend, <0.001) after adjustment for race or ethnic background and obesity group. (NEJM-2)

In examples (112)–(115), the majority of Processes (7/10, 70%) are material, although relational processes were also used to characterize the results after adjustments had been made (e.g., in (113)). Processes were always in the past-tense form (usually the simple past), and verbs such as *adjusted*, *restricted*, and *affected* were indicative of the step. Other step-specific markers included nouns such as *adjustment*, *relationship*, and *change*, lexis related to study outcome (see step 5d), and temporal and conditional conjunctions such as *after* and *when*. Again, as in previous move-7 steps, no modality was observed. References were also scarce. In example (114), the reference, which was one of only three examples in this step, calls the reader's attention to a study that discusses "body weight and longevity," an article co-authored by three of the six authors of NEJM-3.

An additional low-frequency step was observed as part of move 7 (in 4/17 articles, 24%). This step is not listed in the summary of Table 5.1, but it provided basic interpretation of the study results in relation to previous studies, a step generally reserved for the Discussion section (steps 8b and 8c, section 5.5.1 below; see also ICMJE 2006: 28 and Appendix).

- (116) The years of birth ranged from 1945 to 1982, suggesting that these relations [risk of adolescent obesity vs. childhood weight] have been consistent over time. [...] Year of birth in the studies of adults ranged between 1929 and 1970, suggesting that associations [adult obesity vs. childhood weight] have been consistent over time. (BMJ-2)

In the above example, the Themes in the dominant clauses refer to study parameters, while those of the dependent *that*-clauses refer to the relations between parameters. The use of the -ing participle *suggesting*, however, is the key marker in this step, and indicates a tentative interpretation of the study results. From a functional perspective, *suggesting* has both coordinating and subordinating properties, and could for example be replaced with “and [this/it] suggests” or “which suggests.” This functional “ambiguity” backgrounds information in a way that the use of the coordinator *and* and the relative pronoun *which* cannot. Coordination with *and* implies a certain parataxis between clauses, so that no obvious backgrounding effect is observed. The relative pronoun *which*, on the other hand, introduces a hypotactic clause, which backgrounds the second clause in relation to the first. However, unlike the -ing participle, the relative pronoun refers directly to the entire preceding clause and may indicate greater emphasis on the relation between the study results and the authors’ interpretation of them (i.e., hypotactic elaboration). A similar example can be seen in (117), in which the -ing participle *indicating* is used to convey (and, at the same time, background) the authors’ comment on the results.⁵⁸

- (117) The survey x BMI group interaction comparing diagnosed diabetes prevalence in the first vs fifth surveys was statistically significant for obese persons ($P=.03$) but not overweight persons ($P=.28$), indicating that secular increases in prevalence of diagnosed diabetes were greater among obese than lean persons. (JAMA-5)

Although this step has much in common with steps 8b and 8c (see section 5.5.1 below), the interpretations were relatively brief and always backgrounded in dependent clauses. No overall conclusions were drawn concerning the main study outcomes.

⁵⁸ Note that the verbs *suggest* and *indicate* represent identifying relational Processes.

5.4.2 Summary of Results Section

The lexicogrammatical features of steps 7a–7d are summarized in Table 5.4.

Table 5.4. Moves summary for Results sections of corpus articles

Step no.	Theme	Process	Tense, voice	Personals	Modality	References
7a	unmarked	relational	s-past/pr-per active	–	–	NI
7b	unmarked	rel./mat.	s-past active	–	–	–
7c	unmarked	relational	s-past active	–	–	–
7d	UM/TC/MT	mat./rel.	s-past active	<i>we</i>	–	few NI
add. step	unmarked	relational	s-past/pr-per active	–	–	–

Absolute values are not presented. The above summary is observational and based on the examples described in section 5.4.1. In instances in which combinations are presented, those listed first were the most frequently observed. Abbreviations: UM/MT: unmarked/marked Themes; TC: thematic clause; rel./mat.: relational/material clauses; s-past: simple past; pr-per: present perfect; NI: non-integral references; add. step: additional low-frequency step

5.5 Analysis of Discussion Section

The main rhetorical function of the Discussion section was to interpret the results of the study in relation to previous studies, to discuss the implications of the study, and to suggest areas in which further research might be warranted. This was done in terms of three structural moves (moves 8–10).

5.5.1 Move 8

Move 8, “discussing the main findings,” was identified in all the study corpus articles (see Table 5.1). The move had an order stability of 88% and was realized by three steps, as well as by a less frequently observed fourth step.

Step 8a stated the main study findings in relation to the hypotheses and/or study objectives of step 3a (see section 5.2.3). This step was identified in all the corpus articles. These hypothesis-related findings were sometimes stated explicitly (5/17 articles, 29%), as in (118) below, and, in this instance, the cognitive mental Process *hypothesised* most clearly signals the start of the step. The final Theme (underlined), which includes a causal conjunction and a Participant referring to the hypothesis, concludes the step, at least regarding this particular hypothesis-related finding.

- (118) We had hypothesised that clustering would have been more marked in women who had experienced greater social disadvantage throughout their lives... Although risk factors were more common in women from manual social classes in either childhood or adult life, they showed broadly similar patterns of clustering in all four social class groups. Thus, our main hypothesis was not supported. (BMJ-1)

However, more commonly (in 12/17 articles, 71%), hypothesis-related findings were implicit. As illustrated in step 3a, the study objective of JAMA-4 (see example (59) below, repeated from section 5.2.3) was to *estimate deaths associated with underweight, overweight, and obesity*. In (119), the authors respond directly to that objective.

- (59) The objective of this study was to estimate deaths associated with underweight, overweight, and obesity in the United States in 2000 by using all available mortality data from the NHANES [National Health and Nutrition Examination Survey] and to offer an assessment of the uncertainty of those estimates. (JAMA-4)
- (119) Our results show increased mortality associated with underweight and with obesity, particularly with higher levels of obesity, relative to the normal weight category. (JAMA-4)

Similarly, the authors of LAN-1 (see (120) below) begin their Discussion section by responding to their study objective of examining *the association between child and adolescent television viewing and a range of adult health indicators*.

- (120) Our results show that television viewing during childhood and adolescence is associated with overweight, poor cardiorespiratory fitness, raised serum cholesterol, and cigarette smoking in early adulthood. We found no significant association between television viewing and blood pressure. (LAN-1)

In these two examples, the opening Participants and Processes of step 8a are identical. Indeed, this particular Participant–Process combination (*Our results show*) was frequently observed at the start of the Discussion section (6/17 articles, 35%; and 6/12 [50%] articles without explicit mention of hypotheses). Participants (Tokens) were modified with either possessive determiners or definite articles, i.e., *Our results* or *The results*, and relational Processes such as *show* and *demonstrate* were used to identify the meaning of the results (Values). Alternative Participant–

Process combinations at the start of the Discussion section included *We found* (BMJ-4), *We demonstrated* (JAMA-1), and *We documented* (JAMA-5).⁵⁹

The majority of Processes in step 8a were relational, but mental Processes were also identified (e.g., *We found* in (120) and BMJ-4). The simple present was used when referring to elements within the article, e.g., *Our results show...* ((120) above), but otherwise the past-tense form was preferred. No modality or references to previous research appeared in step 8a.

An important step in summarizing the main findings, which was identified in all the corpus articles, was to compare the results of the study with those in the medical literature. This step (step 8b) was characterized by a number of Themes marked with Circumstances referring to the study and/or previous studies. In general, these Circumstances as well as certain dependent clauses served to compare and contrast the study results with those of previous publications. For example, in (121) below, the first Theme is marked with a locational Circumstance (the prepositional phrase *In our cohort*), which highlights the current study in relation to data from *another pediatric sample*. The concessive hypotactic enhancement clause of the second sentence serves to qualify the meaning of the subsequent dominant clause by foregrounding a concession concerning the study's findings (signaled by the conjunction *Although*). Again, as in the previous sentence, the findings are compared favorably with those of a previous study.

- (121) In our cohort, C-reactive protein levels tended to rise with increases in the z score for the body-mass index — a finding similar to that in another pediatric sample.³⁷
Although these levels were at the high end of the normal range, such levels have been associated with adverse outcomes.³⁸ (NEJM-2)

A similar compare-and-contrast technique is used in (122) (see below). In this instance, the opening Theme contains reference to the literature (marked with a comparative Circumstance, *As in earlier reports*) and to the current study (the Participant: *we*), and the dependent *that*-clause indicates how a specific study parameter, *small size at birth*, compares across the studies. In the following sentence, a difference between the findings in the literature and those of the study is then emphasized by the adversative conjunctive Adjunct *however*.

⁵⁹ Further analysis concerning the use of personals is provided in section 5.6.

- (122) As in earlier reports,^{4,8,20} we also found that a small size at birth, defined by a low birth weight or ponderal index, was associated with increased plasma glucose and insulin concentrations and insulin resistance during adulthood. It was not, however, associated with the occurrence of impaired glucose tolerance or diabetes in our study. (NEJM-1)

However, not all Themes signaling the start of step 8b were marked. In (123), for example, the Theme is unmarked, and the concessive and comparative Circumstances *despite* and *in contrast* appear backgrounded in the Rheme. The opening Theme is also unmarked in (124). Here, comparison of the study with the literature is expressed by the identifying relational Process *contrast*.

- (123) We did not observe an independent protective effect of exclusive breast feeding on obesity in our final model, despite strong univariable associations, and in contrast to our previous studies.¹² (BMJ-4)
- (124) These results contrast with those in most adult cancers in which underweight patients have no excess mortality and overweight patients have excess cancer-related death rather than death from excessive toxicity.²⁰⁻²⁵ (JAMA-2)

Step 8b can be likened to the literature review of moves 1 and 2, and the compare-and-contrast technique described above is similar to that of step 2a (identifying a lack of data in the literature). Step 8b acts to resituate the study in relation to the literature after having presented the study's observations in move 7 and, as in moves 1 and 2, non-integral references were a common feature; integral references, however, were only observed in one article (two instances in LAN-2). Processes were predominantly relational or mental, as in the examples above, and a combination of simple-past and simple-present verb forms was used. The present tense was generally used to make generalizations based on the study findings, while the past tense served to highlight specific study findings (see above examples).

Step 8c discussed possible mechanisms/causes for the study findings as well as their potential implications, and was noted in 12 (71%) of the corpus articles. This particular step was marked by the use of modal finites and mood Adjuncts (commonly known as "hedging"; Swales 1990: 112), as the authors speculated on the possible explanations for their results. These interpretations were often accompanied by citations of previous studies (non-integral references only; see

(125) below), which were used to lend support to the authors' arguments.

The following examples from BMJ-4 and JAMA-4 illustrate the use of modal finites (*may* in both examples) and mood Adjuncts (*perhaps* in JAMA-4) in step 8c. Other commonly observed modal finites were *could* and *might*, and additional mood Adjuncts, which were generally less common than modal finites, included *possibly* and *probably*.

(125) Television viewing may confer risk through a reduction in energy expenditure because watching television is associated with dietary intake, or because large amounts of time spent sedentary may contribute to impairment of the regulation of energy balance by uncoupling food intake from energy expenditure.^{18 19} (BMJ-4)

(126) The differences between NHANES I and the later surveys suggest that the association of obesity with total mortality may have decreased over time, perhaps because of improvements in public health or medical care for obesity-related conditions. (JAMA-4)

Note the use of the verb *suggest* in (126). This identifying relational Process assigns a Value to the observed differences (a decrease in obesity-related mortality over time) that expresses greater uncertainty than if other relational processes such as *show* or *prove* had been used. The use of this particular Process was characteristic of step 8c, as were similarly "cautious" identifying relational Processes such as *imply* and *indicate*.⁶⁰ The simple-present tense was used throughout.

Themes in step 8c were generally unmarked, and usually referred to study findings. These findings were typically assigned Values or "possible meanings" by relational Processes and their accompanying modal finites (as demonstrated above).

A fourth step, not listed in the moves-analysis summary due to its low frequency (3/17, 18%), provided brief background information for the study and was essentially a summary of moves 1 and 2. Themes in this step were unmarked, and Subjects referred to study parameters or the study material. Processes were relational, and the present-tense verb form was used. An example from BMJ-1,

⁶⁰ The verb *suggest* may also be classed a projecting verbal Process, e.g., *Jill suggested that Jack should be careful*. In the example above, however, *suggest* was considered a Process of demonstration rather than one of recommendation (see also Halliday & Matthiessen 2004: 235). The same applies to the verb *imply*.

which also includes a modal Finite (*might*), is provided below.

- (127) People who are obese, smoke, and have hypertension and hypercholesterolaemia might be considered common high risk stereotypic patients who require multiple risk factor intervention. (BMJ-1)

Such a statement seems reasonable, given that it appears at the start of the Discussion before step 8a and provides background information for the study's main findings subsequently given in step 8a. However, the need to avoid repetition in the Discussion section, as stressed by the ICMJE (2006: 28; see also Appendix), could be a possible explanation for this step's relatively low frequency in the study corpus.

5.5.2 Move 9

Move 9 was summarized as “indicating the study's limitations,” and was presented as a headed subsection in four articles (BMJ-1, -2, -3, and -4). It was observed in 15 (88%) of the study corpus articles and had an order stability of 94% (see Table 5.1). The move was realized by one elemental step, which is described below.

Step 9a, indicating the strengths and weaknesses of the study, was integral to move 9 and appeared in all instances of the move (15/15 articles, 100%). Its purpose was to indicate to the reader, particularly in the case of study weaknesses, areas in which the study might be improved. It also served to address possible questions the reader might raise in connection with the study's aims and methodology, and to indicate how certain restrictions in the study were avoided or accounted for.

- (128) The strengths of our study include: long-term prospective study design with high rates of follow-up over 15 years... [five additional study strengths]; and the demographics of the cohort—young adult black and white men and women from four US metropolitan areas who have been examined during a period of life when substantial weight gain occurs and chronic diseases arise. The limitations of this study relate mainly to its observational nature, with the possibility of residual confounding precluding definitive conclusions about causality, and to reliance on self-reported diet and other lifestyle factors. (LAN-2)

- (129) Weaknesses of this study include the use of self reported height and weight in adulthood. (BMJ-3)

As demonstrated above, the Themes of step 9a were generally unmarked, and Subjects referred to *limitations*, *strengths*, and *weaknesses*. Other Themes in the step included “questionable” study parameters such as the first underlined Theme in (130). Note that textual Themes were also used to emphasize that a possible limitation (in this case, self-reported body-size measurements) might not necessarily be limiting. This was achieved by using the adversative conjunctive Adjunct *however* and the additive conjunctive Adjunct *in addition*. A concessive dependent clause (signaled by the conjunction *Although*) was used to similar effect in BMJ-2 (see (131) below) to indicate that the consequence remained unchanged.

- (130) The limitations of our study deserve mention. The measures of body size used in our study were self-reported. However, validation studies demonstrated the accuracy of these reports. In addition, any misclassification is likely to be random with respect to case status and therefore would bias the study results toward the null. (JAMA-3)

- (131) Systematic reviews are subject to publication bias. Although we attempted to limit the impact of this through contact with first authors and experts, we did not identify any unpublished analyses. (BMJ-2)

In all the examples above, the predominant Process type is relational. This was typical of step 9a, as was a combination of simple-past and simple-present verb forms. Hedging was also observed, namely the noun *possibility* in (128) and the adjective *likely* in (130).

5.5.3 Move 10

A conclusion was provided in the final move (move 10) of the study corpus. This move was observed in all the corpus articles (17/17, 100%; see Table 5.1), and was indicated by a separate subheading in four of the BMJ articles (BMJ-1, -2, -3, and -4). The move had an order stability of 88%, and consisted of three potential steps.

Step 10a reiterated the main study findings of step 8a and was identified in 11/17 articles (65%). As can be seen in the examples below, the opening Theme of this particular step was frequently marked with a Circumstance, the prepositional

phrase *In conclusion*; indeed, this was observed in 6/11 articles (55%). Subjects varied and included individual study parameters, overall results, and the authors of the article. This was also the case when Themes were unmarked, as in (134) below.

- (132) In conclusion, our results show that obesity and weight gain are associated with an increased risk of symptomatic nephrolithiasis. (JAMA-3)
- (133) In conclusion, we found that both the body-mass index and the level of physical activity were important and independent predictors of mortality and that a higher level of physical activity does not appear to negate the risk associated with adiposity. (NEJM-3)
- (134) Obesity limited to childhood has little impact on adult socioeconomic, educational, social, and psychological outcomes. Persistent child and adult obesity in women is associated with somewhat poorer employment and relationship outcomes, although adversity is substantially less than previously reported. (BMJ-3)

Processes were predominantly relational. However, there were exceptions, such as in (133) above, in which *found* represents a cognitive mental Process. A combination of the simple-past and simple-present tense was used throughout the step, depending on whether findings were generalized (simple present, e.g., in (132) and (134)) or study specific (simple past, e.g., in (133)). There was no evidence of modality or references to previous studies as a part of this step.

Step 10b was used to state the implications of the study findings, and was observed in 14 (82%) of the corpus articles. Although it was related to step 8c (“mechanisms/causes,” section 5.5.1), step 10b appeared with a higher frequency (82% versus 71%) and was therefore not always a summary of step 8c. In (135), for example, a discussion of the study implications was not identified in move 8, but possible implications are clearly stated as part of move 10 (see below).

- (135) Clustering of risk factors—in distinction to the co-occurrence of risk factors—implies that they are not independent of each other and may therefore reflect an underlying causal or pathogenetic mechanism. The clustering we observed... does not seem to be particularly associated with causal mechanisms operating in childhood. Clustering of risk factors may be of relevance in explaining observed variations in risk for coronary heart disease. If clustering is more pronounced in geographically or socially defined groups and clusters of risk factors operate synergistically... then much of any

“unexplained” variation may be explained by risk factor clustering. (BMJ-1)

As in step 8c, step 10b made similar use of modal finites and mood Adjuncts as an expression of uncertainty. Excerpt (135) above, which is the longest example of step 10c in the study corpus, is a prime example of this. Note the rather extensive use of the modal finite *may*, the “modalized” attributive relational Process *seem*, as well as the mood Adjunct *particularly* (all underlined). In the first sentence of this excerpt, the authors also use an identifying relational Process, *implies*, as a way of cautiously assigning a Value (that of dependency) to the *clustering of risk factors*. This Value is in turn interpreted with caution (and given a Value itself) using a modal finite (*may*), a causal conjunct (*therefore*), and an additional identifying relational Process (*reflect*).

In another example from BMJ (see (136) below), the first clause relates to the study findings of step 10a, while the second suggests possible implications (step 10b). There is a noticeable difference in modality between the two clauses, and the level of probability expressed by the modal auxiliary in the first clause (*can*) is greater than that in the second (*could*). In fact, it could be argued that the meaning expressed by *can* contains relatively little epistemic modality, and that *can expect* is a quality assigned to the patients by the author rather than a comment by the author on the possibility of the proposition.

(136) Our study provides data on how much weight patients can expect to lose by dieting, and these data could help practitioners in managing patients’ sometimes unrealistic expectations of weight loss targets. (BMJ-5)

As demonstrated above, Themes were generally unmarked and Subjects referred to the study, study data, or individual study parameters. Processes were commonly relational and modalized in such a way as to express a degree of uncertainty. This effect was created by using a combination of modal finites, modal Adjuncts, and “modalized” relational Processes such as *suggest*, *seem*, *appear*, and *imply*. The simple-present verb form was most commonly used, often as a way of generalizing the study’s findings (as in (136)), but instances of the past tense were also observed on occasion (e.g., *The clustering we observed...*; example (135) above).

Step 10c provided recommendations for future research. It was identified in 15 (88%) of the corpus articles and frequently appeared as a final step at the end of move 10 (12/15, 80%). In a similar way to move 2 of the Introduction, step 10c indicated possible gaps or shortcomings in the study that warranted further research. It also provided suggestions for how the study's findings might be used in clinical practice or public health programs.

As illustrated in the examples below, Themes were generally unmarked, and cognitive mental Processes such as *determine*, *investigate*, and *explore* were used in a similar way to those in step 3a, to suggest possible objectives for future research. Participants, however, were slightly different to those of step 3a, in that they generally referred to a “product of human consciousness” (Halliday & Matthiessen 2004: 203), e.g., *Future studies*, rather than a more typical “human” Senser. The simple-present tense was most commonly used, but other verb forms such as the present perfect were also observed (see (139) below).

- (137) Further research is needed to determine if these findings can be replicated in different age groups and countries, and to investigate the temporal relationships between overweight and obesity and QOL [quality of life] in children. (JAMA-1)
- (138) Future studies should explore the effect of obesity and sex on urine composition, and weight loss should be explored as a potential treatment to prevent kidney stone formation. For now, clinicians have an additional reason to encourage weight control in their patients. (JAMA-3)
- (139) Prevention strategies for childhood obesity to date have usually been unsuccessful and typically focus on change in lifestyle during childhood or adolescence. Future interventions might focus on environmental changes targeted at relatively short periods in early life, attempting to modify factors in utero, in infancy, or in early childhood, which are independently related to later risk of obesity. (BMJ-4)

There were a few instances of marked Theme in step 10c. In (138), the Theme of the second sentence is marked by a temporal Circumstance (underlined), and in (140) below, a study-parameter Participant is fronted by a concessive Circumstance (underlined). The former highlights how clinicians, in wait of future research, might use the findings of the current study, while the latter appears to acknowledge a less-than-favorable study finding.

- (140) Despite our encouraging findings, a considerable proportion of lean as well as obese persons still have elevated levels of modifiable risk factors, particularly when one considers that the current definitions of risk factor control are more aggressive than the definitions used in this trend analyses. Clinical and public health efforts should continue to emphasize maintenance of healthy lifestyle behaviors for lean as well as overweight and obese persons. (JAMA-5)

A degree of modality was also demonstrated in this step, but it was different to that of steps 8c and 10b. While authors emphasized caution, uncertainty, and probability in steps 8c and 10b by using the modal finites *may* or *could*, necessity and obligation were stressed in step 10c by using the modal finite *should*, e.g., *Clinical and public health efforts should continue to emphasize...* (see (140)). An additional expression of modality was the use of interpersonal grammatical metaphor to express author opinion, but this occurred in only one of the corpus articles, LAN-1: *We believe that reducing television viewing should become a population health priority.*

5.5.4 Summary of Discussion Section

The lexicogrammatical features of moves 8–10 are summarized in Table 5.5.

Table 5.5. Moves summary for Discussion sections of corpus articles

Step no.	Theme	Process	Tense, voice	Personals	Modality	References
8a	Unmarked	rel./ment.	s-past/s-pres active	<i>we, our</i>	–	–
8b	MT/UM	rel./ment.	s-past/s-pres active	<i>we, our</i>	–	NI/I
8c	Unmarked	relational	s-pres active	–	<i>may, could, might, possibly, probably</i>	NI
add. step	Unmarked	relational	s-pres active	–	<i>might</i>	–
9a	UM/mult.	relational	s-past/s-pres active	<i>we, our</i>	–	–
10a	MT/UM	relational	s-pres/s-past active	<i>we, our</i>	–	–
10b	unmarked	relational	s-pres/s-past active	<i>we, our</i>	<i>may, could, particularly</i>	–
10c	UM/MT	ment./rel.	s-pres/s-past active	–	<i>should</i>	–

Absolute values are not presented. The above summary is observational and based on the examples described in section 5.5.1–5.5.3. In instances in which combinations are presented, those listed first were the most frequently observed.

Abbreviations: MT/UM/mult.: marked/unmarked/multiple Themes; rel./ment.: relational/mental clauses; s-past: simple past; s-pres: simple present; pr-per: present perfect; NI/I: non-integral/integral references; add. step: additional low-frequency step

5.6 On the Use of Personals

As noted in the analysis above, the personal pronoun *we* and the possessive pronoun *our(s)* appear in certain steps throughout the study articles, and occurred 434 and 176 times, respectively, in the 104,201-word corpus (overall frequencies: 4.2 and 1.7 per 1000 words). The personal pronoun *us* was observed on nine occasions (frequency: 0.09 per 1000 words). Table 5.6 indicates the distributions of *we* and *our* across the four sections (IMRD) of the study articles, recorded using the “concord” function of WordSmith 4.0 (.TXT files) and the search tool of Adobe PDF Reader version 8.1.0 (PDF files).

Table 5.6. Distribution of personals (*we/our*) in the study corpus

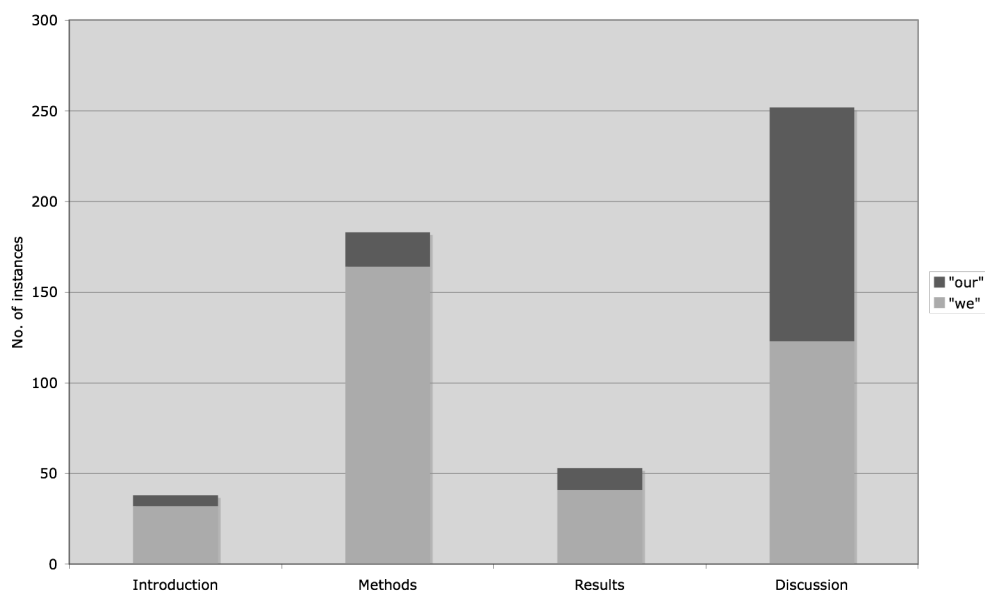
Source	Introduction	Methods	Results	Discussion	Others*	Total
BMJ-1	2 / 0	10 / 0	3 / 0	6 / 9	6 / 1	27 / 10
BMJ-2	3 / 0	10 / 1	6 / 1	4 / 6	3 / 2	27 / 10
BMJ-3	2 / 0	9 / 0	1 / 1	25 / 8	0 / 0	37 / 9
BMJ-4	2 / 0	17 / 1	1 / 0	7 / 8	2 / 0	29 / 9
BMJ-5	1 / 1	21 / 0	2 / 0	2 / 2	0 / 1	26 / 4
JAMA-1	3 / 0	1 / 0	1 / 2	3 / 9	6 / 0	14 / 11
JAMA-2	0 / 0	0 / 0	1 / 0	2 / 1	9 / 0	12 / 1
JAMA-3	3 / 1	9 / 2	3 / 0	7 / 9	7 / 4	29 / 16
JAMA-4	4 / 1	11 / 0	8 / 3	5 / 17	6 / 0	34 / 21
JAMA-5	3 / 0	14 / 2	1 / 1	17 / 12	10 / 0	45 / 15
LAN-1	1 / 0	14 / 3	0 / 0	21 / 3	5 / 1	41 / 7
LAN-2	1 / 1	25 / 2	3 / 1	5 / 7	4 / 0	38 / 11
LAN-3	1 / 0	1 / 2	0 / 1	1 / 10	5 / 0	8 / 13
NEJM-1	1 / 0	5 / 0	2 / 0	5 / 7	2 / 0	15 / 7
NEJM-2	2 / 0	4 / 1	3 / 1	6 / 9	4 / 1	19 / 12
NEJM-3	1 / 1	13 / 5	6 / 1	2 / 10	2 / 0	24 / 17
NEJM-4	2 / 1	0 / 0	0 / 0	5 / 2	2 / 0	9 / 3
Total	32 / 6	164 / 19	41 / 12	123 / 129	72 / 10	434 / 176

* “Others” refers to other sections of the text, including the abstract, acknowledgments, and references, as well as occasional reading errors such as “*we ight*” and “*bet we en*”.

The personal pronoun *we* was used most frequently in the Methods (166/362, 46%) and Discussion (123/362, 34%).⁶¹ In the Methods section, *we* commonly collocated with the material Process *used* (30/164, 18%; e.g., *we used ANOVA to examine differences...*, example (97), step 6a), while in the Discussion section it most frequently collocated with the mental Process *found* (29/123, 24%; e.g., *We found no significant association...*, example (120), step 8a). Other collocations, which occurred in both the Methods and Discussion sections, included *we examined* (19/287, 7%), *we calculated* (9/287, 3%), and *we adjusted* (9/287, 3%). The possessive pronoun *our* was used primarily in the Discussion section (129/166, 78%), and most commonly collocated with the head nouns *study* (38/166, 23%), *results* (24/166, 14%), and *findings* (16/166, 10%).⁶²

Figure 5.1 (generated from the data in Table 5.6) demonstrates that personals (*we* and *our*) tended to cluster in the Methods (185/528, 35%) and Discussion (252/528, 48%), with relatively few examples occurring in the Introduction (38/528, 7%) and Results (53/528, 10%).⁶³ The corpus articles contained a mean of 36 (range: 12–60) personals per article.

Figure 5.1. Distribution of personals in the study corpus



⁶¹ The denominator was corrected for the occurrence of the personal pronoun *we* in IMRD sections only, i.e., occurrence in IMRD = total occurrence (434) – occurrence in “other” sections (72) = 362.

⁶² The denominator was corrected for the occurrence of the possessive pronoun *our* in IMRD sections only, i.e., occurrence in IMRD = total occurrence (176) – occurrence in “other” sections (10) = 166.

⁶³ The total number of personal and possessive pronouns in the IMRD sections was: 362 + 166 = 528.

6 Discussion of Study Results

6.1 Comparison with the Literature

6.1.1 Structural Moves Analysis

According to the textual analysis above, the 17 medical research articles of this study corpus followed a structural and thematic pattern that comprised a series of 10 commonly used structural moves. These moves were observed with frequencies of 88 to 100% and tended to appear in the same order throughout the corpus (adjusted order stabilities: 88–100%). Each move consisted of one to six constituent elements or steps (28 steps in total), which appeared with frequencies of between 59 and 100%.

In a previous study by Nwogu (1997), 11 structural moves (frequencies: 47–100%; order stabilities: unknown) were identified in a corpus of 15 medical research articles published between the years 1985 and 1987 in *BMJ*, *JAMA*, *LAN*, *NEJM*, and the *Journal of Clinical Investigation* (*JCI*). These 11 moves each consisted of one to five steps (total: 26 steps).

Skelton (1994) conducted a moves analysis of 50 medical research articles published in the *British Journal of General Practice* (*BJGP*) in 1989–1993, and identified 15 moves common to the study material. These moves appeared with frequencies of 66 to 100% and had order stabilities in the range of zero to 100%. Steps were not identified.

In addition to these studies, ICMJE (2006: 4) offers guidelines on “the conduct and reporting of [medical] research.” These guidelines include instruction on how to prepare medical research articles for publication. Recommendations for the writing of original research articles using the IMRD structure are included in the Appendix and are numbered for ease of reference. Each of these numbers corresponds to a move, e.g., move 1: *Provide a context or background for the study*; move 2: *State the specific purpose or research objective*, etc. (ICMJE 2006: 26; see Appendix).⁶⁴ Twelve moves were identified.

Table 6.1 provides a comparison of the moves analyses of the four above-

⁶⁴ Moves were identified using the same methodology as that described in section 3.3.2.

mentioned studies/publications. A total of 19 different moves were identified across these four sources and are summarized in the first column of the table; their relevant move numbers are given under the name of each study/publication. Some moves were common to all four sources, either as moves assigned unique move numbers (e.g., “stating the research purpose,” which was given separate, unique move numbers by all four sources), or as moves that generalized two or more move descriptions (e.g., move 7 in this study, which described both “consistent observations” and “non-consistent observations” under one move, i.e., “reporting observations,” but was classed as two separate moves by Nwogu (1997)). There were also instances in which a move in one study was classed as a step in another (e.g., “limitations,” classed as move 9 in this study and step 10(5) in Nwogu (1997)), and others in which no comparable move or step was identified (e.g., “describing mechanisms/causes,” which was not observed in Skelton’s (1994) corpus).

Table 6.1. Multi-source comparison of structural moves analyses of medical research articles

Move description	Move no.			
	This study	Nwogu (1997)*	Skelton (1994)*	ICMJE (2006)*
<i>Introduction</i>				
Study background	1	1	2	1
Importance of study	2b	–	1	–
Gap in literature	2	2	3	1
Research purpose	3	3	4	2
<i>Methods</i>				
Material	4	4	5	3
Experimental procedure	5	5	6	4
Data-analysis procedure	6	6	7	5
<i>Results</i>				
Consistent observations	7	7	10	6
Non-consistent observations	7	8	10	6
Non-verbal material	7a	7	9	6
Adjustments	7d	–	8	–
Assessment	–†	7(2)	11	–
<i>Discussion</i>				
Main findings	8	9	13	7

Specific outcomes	8a	10	13	7
Comparison with literature	8b	10(4)	14	9
Mechanisms/causes	8c	10(2)	–	8
Limitations	9	10(5)	12	10
Recommendations	10c	11(2)	15	11
Conclusion	10	11	–	12

* Steps are not available for Skelton (1994) or ICMJE (2006). For Nwogu (1997), figures in parentheses refer to step number.

† Registered as a low-frequency step (24%; see section 5.3.1).

Of the 19 moves summarized in Table 6.1, four were considered common to all studies/publications and were assigned unique move numbers in each study: “stating the research purpose,” “describing the selection of study material,” “describing the experimental procedure,” and “describing the data-analysis procedure.” These moves apply to the Introduction and Methods sections only, and suggest that, at least in the studies described above, these sections of the medical research article may have a more standardized thematic structure than the Results and Discussion. Indeed, in all the above-mentioned studies, the Methods section was described in terms of the same three structural moves, presented in the same chronological order (order stabilities: 94–100% in this study, 74–80% in Skelton). This, in addition to the relatively high frequencies with which the three moves were observed (94–100% in this study, 86–100% in Skelton, 60–100% in Nwogu), suggests that the methodologies used in medical research, or at least in the writing of medical research, are highly standardized. That the Results section seems less standardized, i.e., no commonly observed moves across the above-mentioned studies (see Table 6.1), is perhaps surprising, considering its function is to report data obtained in relation to the study’s methodology (see sections 4.1.3 and 5.4). One might expect the Results section to reflect the structural homogeneity of the Methods section, but this appears not to be the case. On the contrary, according to this cross-study moves criterion of standardization, the Introduction section demonstrates greater homogeneity than the Results section (one versus no common cross-study moves).

Skelton (1994: 457) observed that the moves of the Methods and Results sections were more *heterogeneous*, due to their being more “strongly content-driven,” than the Introduction and Discussion sections. This is in contrast with the comparison presented in Table 6.1, in which the moves of the Results and Discussion sections appear more heterogeneous than those of the Introduction and

Methods sections. The reason for this discrepancy may be partly contextual. Although all three studies (Nwogu 1997, Skelton 1994, and the current study) attempt to characterize the generic structure of the medical research article, they differ in terms of their purpose and intended audience. Nwogu's (1997: 120) study, much like the current study, "is written for specialists in the field of linguistics" and aims to provide a linguistically focused description of the genre, while Skelton's (1994: 455) study, "an aid to writing original papers for publication," primarily addresses the writing needs of general practitioners. These differing approaches to genre analysis are perhaps understandable given the academic backgrounds and interests of the authors (linguistics and general practice medicine), but they may lead to fundamental differences in methodology and hence differing results (see Nwogu 1997: 120; Skelton 1994: 455–456).

The findings of the structural moves analysis of Nwogu (1997) and this study appear to be in closer agreement (i.e., eight common moves, and 11 versus 10 moves in total; see Table 6.1) than those of Skelton (1994) and this study (seven common moves, 15 versus 10 moves) or Skelton (1994) and Nwogu (1997) (six common moves, 15 versus 11 moves). In addition to the possible contextual reason highlighted above, these move discrepancies may also be related to the choice of study material. This study used similar sources to Nwogu (1997) for the selection of corpus articles (BMJ, LAN, JAMA, and NEJM, with the addition of JCI in Nwogu's study corpus). Skelton's (1994) material, on the other hand, was selected from one source only (BJGP), but represented a much larger corpus (50 articles) than the corpora of this study (17 articles) and Nwogu (1997; 15 articles). The differences in the number and commonality of moves could therefore be a reflection of the size of the study sample or of different research and/or writing traditions in general practice (specifically in BJGP) compared to those in other areas of medicine (i.e., in BMJ, LAN, JAMA, NEJM, and JCI). It should also be noted that BJGP does not officially endorse the guidelines of the ICJME (2006) and may therefore have different journal-specific requirements to the source publications used in Nwogu (1997) and this study.⁶⁵

Another potential reason for deviation between the results of studies using apparently similar methodologies for moves analysis is the issue of subjectivity

⁶⁵ See ICJME "Journal List" at URL: www.icmje.org/jrnlist.html (accessed March 6, 2007).

(Holmes 1997: 325; Hyland 2002: 116; see also discussion in section 2.4.1). Assigning a move to a particular segment of text depends on a number of variables. These include summarizing the content of the text, identifying lexicogrammatical markers, and assigning discourse functions, all of which are more or less prone to broad interpretation. Some moves identified by structural moves analysis subsequently operate at higher levels of generalization than others (see Lewin et al. 2001: 18), and a segment of text classed as a move in one study may be interpreted as a step or part of a different move in another. For example, in Table 6.1, the move description “adjustments” in the Results section is interpreted as a move in Skelton (1994) and as a step in this study, while it is not mentioned as either in Nwogu (1997). Furthermore, steps are not discussed by Skelton (1994), which allows for the possibility that certain moves in that study might otherwise have been interpreted as steps had a similar methodology to Nwogu (1997) or this study been used.

In order to avoid some of the potential difficulties in assigning structural moves and steps, this study also examined the occurrence of hyperThemes (Martin 1993a: 244–247; Martin & Rose 2003: 181–184; see also section 2.2.3). A hyperTheme signals the point of departure for a particular phase of discourse in a text and summarizes a series of subsequent related Themes. HyperThemes can be realized by clausal elements or paragraph constituents (i.e., a sentence or clause), and although their identification arguably entails a certain amount of subjectivity, they may be easier to identify than moves and steps, and thus serve a valuable supplementary role in describing the development of a text.

6.1.2 Lexicogrammar

Skelton’s (1994: 457) comments above regarding the heterogeneity of moves in the Methods and Results compared to the Introduction and Discussion of the medical research article, although not supported by this study’s findings or the cross-study moves comparison in Table 6.1, serve to illustrate the close relationship between the Methods and Results, on the one hand, and the Introduction and Discussion, on the other. Methods and Results sections rely on a recount-and-report structure that primarily refers to events taking place within the study (study-internal references), while Introduction and Discussion sections use a present-and-discuss structure to

contextualize the study within a wider research setting (study-external references). For example, in the articles of this study corpus, the number of references to other studies was much higher in the Introduction and Discussion than in the Methods and Results. This is in accordance with Thompson & Tribble's (2001) observations among a series of 16 agricultural botany Ph.D. theses in which they observed that the number of citations in the Introduction and Discussion was between four and seven times that in the Methods and Results. The citation conventions in these Ph.D. theses were different to those in the research articles of this study. The former uses the author-date system, while in medicine (as in many other physical-science disciplines such as physics, engineering, etc.) the convention is to use the numerical-endnote form, i.e., a superscript number referring to a list of references at the end of the article. In disciplines that use numerical-endnote referencing, the prominence with which authors of other studies are cited is reduced (Hyland 1999: 346). For example, research articles in physics demonstrate significantly higher percentages of non-integral compared to integral references (83.1% non-integral versus 16.9% integral) than research articles in, say, philosophy (35.4% non-integral versus 64.6% integral) (ibid.: 347). In this study, the figures were even more skewed toward almost exclusive non-integral referencing than those observed in Hyland's (ibid.) sample of 10 physics articles. There were only eight examples of integral references recorded in the entire corpus of this study (see moves 2 and 8, sections 5.2.2 and 5.5.1). Considering there are 502 references listed in the corpus articles (see Table 3.2, section 3.2), a conservative estimate reveals that, in terms of these two forms of referencing, approximately 98% (494/502) were non-integral compared to 2% (8/502) integral. Since the 502 references in this corpus may have been cited more than once, it is likely that the actual percentage of non-integral references is even greater. According to Hyland (ibid.: 355), this overwhelming preference for non-integral referencing reflects "the conventions of impersonality in science articles" and "the assumption that the person who publishes a claim is largely immaterial to its accuracy."

The relationship between the Introduction and Discussion and the Methods and Results is not only manifested in references to previous studies, but also in aspects of the lexicogrammar of the texts. For example, in the Methods and Results sections of this study, Processes were generally material and given in the simple-past tense (describing what was done and what happened), while in the Introduction

and Discussion sections they were more commonly relational and given in the simple-present or simple-past tense (identifying entities and ascribing attributes to them in relation to previous research). A similar pattern was observed by Heslot (1982, cited in Swales 1990: 135) in a series of 16 research articles from the plant pathology journal *Phytopathology*; namely that the Methods and Results sections were almost entirely dominated by use of the simple-past tense (94% of verb forms in those sections of the corpus), while the Introduction and Discussion sections contained a combination of simple-present and simple-past verb forms, the majority of which were in the present tense (48.5% simple present versus 35% and 39% simple past, respectively). Although a section-by-section analysis of Process types in medical research articles was not found in the literature, MacDonald (2002: 461–462) showed that, in a sample of four medical research articles published in *BMJ* in 1992, relational and material Processes (as in this study) were dominant throughout, accounting for 43.3% and 27.2% of all Processes, respectively.

There were also distinct section-related differences in terms of “author presence,” i.e., the way in which authors expressed their involvement in and comment on various aspects of the research. In the Introduction and Discussion sections, a combination of modality (modal Finites and modal Adjuncts) and personals (pronouns such as *we* and *our*) was used to indicate authors’ opinions on and interpretations of the research (both study-internal and study-external). In contrast, in the Methods and Results sections, there were few instances of modality, and author presence was almost exclusively expressed by personals, indicating the authors’ participation in conducting the research (study-internal).

These observations concerning author presence in the IMRD sections of medical research articles are in agreement with those of Adams Smith (1984: 26) who noted that, in a sample of six medical research articles from *BMJ*, “author’s comment” (modality) was far more common in the Introduction and Discussion sections than the Methods and Results sections. Adams Smith’s (ibid.: 35) comment-per-line ratios were 0.46 for the Discussion, 0.34 for the Introduction, and 0.05 for the Methods and Results. Similar results were observed by MacDonald (2002: 457–458), who found that “by far the highest proportion of speaker’s comment [modality] falls in the Introduction and Discussion sections,” with comment-per-line ratios in these sections (both 0.5) being more than eight times that of the Methods section (0.06) and over twice that of the Results section (0.22).

With reference to personals, which were not discussed by Adams Smith (1984) or MacDonald (2002), the results of this study show that the pronouns *we* and *our* occurred throughout the four main article sections (IMRD), but tended to cluster in the Methods and Discussion (see section 5.6). Heslot (1982, cited in Swales 1990: 135), on the other hand, commented that first-person pronouns did not occur in the Methods and Results sections but did in the Introduction and Discussion. No figures were provided to substantiate this claim (*ibid.*: 135), but the difference may be methodological and/or due to different writing traditions in the journal *Phytopathology* compared to the medical research journals of this study. In other words, the two study corpora may represent different genres or subgenres of the academic research article. Similarly, Martínez (2005), in the study of a one-million-word corpus of English-language biology research articles published in a series of 14 biology journals in 1995–2001, observed that the pronoun *we* was primarily used in the Results section, but also appeared with relatively high frequency in the Introduction and Discussion sections. The Methods section, however, contained relatively few instances (approximately one-tenth of that of the Results section). Again, the difference between these findings and those of the current study could be methodological or may indicate fundamental differences in the way biology research articles are written compared to medical research articles. These possible genre-specific differences appear to be supported by Tarone et al. (1998: 131–132) who examined two astrophysics research articles and noted that the personal pronoun *we* occurred with similar frequencies throughout *all* sections of the articles. Neither of these astrophysics articles adopted the IMRD structure, and the size of the study material is not suitable for generalization. However, based on their findings, Tarone et al. (*ibid.*: 115) state that the astrophysics research article as well as research articles of other academic disciplines such as mathematics and theoretical linguistics do not necessarily follow the structural moves accorded to the academic research article by Swales (1990: 141 ff.), and therefore represent distinct genres rather than subgenres of the more general “research article genre” (see discussion below in section 6.2 regarding the possible implications of this).

In terms of voice, passive constructions were noted throughout all sections of the corpus. However, they were only dominant over the active voice in the Methods section (see summary in Table 5.3, section 5.3.5). This observation is

supported by Heslot (1982, in Swales 1990: 136–137) who also observed more extensive use of the passive in the Methods section than in the Introduction, Results, and Discussion. Martin (1993b: 193–194) offers an interesting explanation for this, based on a secondary-school science text, in which the passive allows authors to foreground apparatus and study material as a series of Themes that define the point of departure as the organization of the methodology rather than the researchers conducting the experiment. This can be seen in example (83) of step 5a, repeated from section 5.3.2 (see below), in which the methodology and study parameters are foregrounded and the researchers as Agents are omitted or backgrounded in the Rheme. A very different effect is achieved if the clauses of (83) are made active by “thematizing” the researchers (see (83a) below). Note, however, that in this study, a combination of the active and passive was generally more common than either entirely passive or entirely active steps, as in example (87) of step 5b (repeated below from section 5.3.2), which foregrounds the authors, study material, and study parameters in turn.

- (83) The NGHS study protocol has been previously described.³² BMI (weight in kg divided by height in m²) was derived from annual measurements of height and weight. Skinfold thickness was measured with Holtain calipers at the triceps, subscapular, and suprailiac sites by centrally trained examiners, and these measures were summed. (LAN-3)
- (83a) Other authors have previously described the NGHS study protocol.³² We derived BMI (weight in kg divided by height in m²) from annual measurements of height and weight. Centrally trained examiners measured skinfold thickness with Holtain calipers at the triceps, subscapular, and suprailiac sites, and they [or we] summed these measures.
- (87) We defined obesity on the basis of a threshold BMI z score of 2.0 or more, adjusted for age and sex. The subjects were then classified as moderately obese (a z score of 2.0 to 2.5) or severely obese (a z score above 2.5). Elevated systolic or diastolic blood pressure was defined as a value that exceeded the 95th percentile for age and sex.¹⁹ [...] Impaired glucose tolerance was defined as a glucose level greater than 140 mg per deciliter (7.8 mmol per liter) but less than 200 mg per deciliter (11.1 mmol per liter) at two hours.²¹ (NEJM-2)

Written scientific discourse has been described as “depersonalized” (e.g., Halliday 1993a: 66; Kreutzberg 2005: 395), and this is frequently attributed to a relative lack of personal pronouns and to extensive use of the passive voice (see, for example, Chafe & Danielewicz 1987, Kreutzberg 2005). While no attempt was made to verify this by comparing the corpus texts with those of “non-scientific” discourse, it may be worth noting that BMJ encourages authors to “write in the active and use the first person where necessary.”⁶⁶ No evidence of increased use of the first-person pronoun, for example, was observed in the BMJ subcorpus compared to the other journals (see Table 5.6, section 5.6). However, this prescriptive appeal to authors, although widely open to interpretation (*where necessary*), may be suggestive of a change in the way in which authors as participants in the research (i.e., as Actors or Sensors) are expected to present themselves in the medical discourse.

The identification of Themes, particularly marked Themes, accounted for a significant part of the analysis in this study. As discussed in section 2.2.3, marked Themes may signal new phases in the discourse, and this makes them useful in identifying hyperThemes and moves. Although the vast majority of Themes in this corpus were unmarked, new steps in the discourse were often (but by no means always) signaled by marked or multiple Themes (e.g., steps 3a and 7d), and in steps 8b and 10a of the Discussion, marked Themes were in fact more common than unmarked Themes (see Table 5.5, section 5.5.4). However, beyond these observations, no distinct Theme-related patterns were observed in the corpus, and a more thorough analysis was not carried out. Indeed, relatively little research has focused on the function of Theme (or hyperTheme) in scientific research articles. One exception is Gosden (1992), who discussed the discourse function of marked Themes in a series of 36 physics, chemistry, and biology research articles. Gosden’s (ibid.: 208–209) definition of marked Theme, adapted to include textual and interpersonal (i.e., multiple) Themes, differs from that of this study, as do the discourse function labels applied to “sentence initial elements” (i.e., textual, interpersonal, and experiential elements other than the Subject that appear in the Theme position). However, even though a meaningful comparison between the two studies is difficult, Gosden’s (ibid.: 218) finding that the majority of “marked

⁶⁶ See “Guidance on House Style” at URL: <http://resources.bmj.com/bmj/authors/bmj-house-style> (accessed October 16, 2007).

Themes” (40.3%) were accounted for in the Discussion section is noteworthy in light of the observation in this study that marked Themes were dominant relative to unmarked Themes in two steps (8b and 10a) of the Discussion section.

6.1.3 Context

Although both genre and register analysis, as described in sections 2.3 and 2.4, stress the importance of understanding the contextual and social setting in which texts are produced, very little research has specifically aimed to describe the context of medical discourse or the medical research article. Neither of the two main studies used for comparison with this study (Nwogu 1997, Skelton 1994) discuss the social setting or discourse community in which medical research articles are produced, and the few studies that do exist in this field focus on doctor–patient exchanges in medical consultations (see review in MacDonald 2002: 447–448).

While there are no suitable comparisons to be made with the literature regarding the contextual analysis of section 4, it is worth noting that much of the textual analysis of section 5 is informed by the preceding contextual investigation. For example, the discourse components of field, tenor, and mode not only describe the social activity taking place, the participants involved, and the channel of communication; they also define the content of the text, the roles of participants within the discourse, and the text’s rhetorical function (see section 4.1). Similarly, a description of the discourse community entails textual as well as contextual analyses (e.g., a description of the specific lexis adopted by community members; section 4.2.5), and supplementary material such as the guidelines of the ICMJE (2006, and Appendix) contain valuable information for describing the social activity of medical research and provide important comparative data for the textual analysis.

6.2 Implications

Research concerning academic discourse, and particularly genre analysis, is primarily motivated by the needs of the learner, i.e., someone for whom a particular discourse type is unfamiliar (sections 2.3 and 2.4). This pedagogic approach to genre provides an important foundation for understanding discourse and the

expectations involved in producing genre-specific texts. Defining the medical research article as a distinct genre rather than in terms of a more general academic research article may have important implications for the understanding of medical discourse and the teaching of English for academic/specific purposes (EAP). Although a number of academic disciplines adopt the IMRD structure for research articles (e.g., biology, plant pathology, physics, psychology, medicine, etc.; see above), there may be distinct differences in the way these article sections are used, which means that the description of structural and lexicogrammatical features across a range of disciplines may be of little value to those interested in acquiring writing skills within a specific discipline. For example, Swales's much-cited studies of genre (1981, 1990) describe a general model for research article introductions (see section 2.4.1) based on research articles in a variety of academic fields (including chemistry, engineering, and agriculture, among others). However, the level of generalization at which these moves are described may be insufficient to describe the discursive expectations of specific academic fields, particularly that of medicine, which was not a part of Swales's (1981) original corpus.

If indeed it is valuable to teach students the structural and lexicogrammatical conventions of a genre, as ESP researchers and Systemic-Functional Genrists have argued (e.g., Martin 1993a: 255; Swales 1990: 4), then the generic differences among research articles of various academic disciplines highlight the need to focus on discipline-specific rather than simply scientific discourse in an EAP setting. The moves analysis of this study could in theory be adapted for such pedagogic purposes, but the findings from this relatively limited-sized corpus should be interpreted with caution and cannot be generalized to the extent that they provide a complete description of the generic features of the medical research article—a necessary prerequisite if the results of such studies are to be applied in a practical educational setting (see discussion of study limitations, section 6.3).

The value of teaching generic conventions as a means of empowering students, or at least as a means of providing learners with the tools to help them participate in certain discourses, is a matter of contention (see section 2.4.3). Genres are generally considered dynamic phenomena (Freedman & Medway 1994: 9; Miller 1984: 163; Bhatia 2004: 23–24), and some researchers therefore argue that the static structural and lexicogrammatical description of a genre is of limited

value to the learner (Bazerman 1988: 315; Freedman 1994: 192). The results of this study and the subsequent comparison with the literature (see section 6.1) raise an interesting point in this respect. Although the 10 structural moves identified in this study were observed with high frequency and high order stability throughout the corpus (both 88–100%), the number and descriptions of moves were in relatively poor agreement with those of previous studies (Nwogu 1997, Skelton 1994; see section 6.1).⁶⁷ Regardless of the reason for this discrepancy—whether a matter of context, subjectivity, differing study material, and/or as a result of changes in the genre over time—the notable differences between this study and those of Nwogu (1997) and Skelton (1994) suggest that any corpus-based description of the medical research article is unlikely to be able to describe fully the complexity and diversity of the genre.

There appears, then, to be a valid argument for avoiding prescriptive guidelines that are either too narrow based on a limited corpus or too general based on a corpus of wide-ranging academic disciplines. However, it should be noted that the ideological argument against the explicit teaching of generic features—that it does not encourage critical thinking in relation to discourse and genre (Freedman 1994)—raises one particularly important issue. New Rhetoricians argue that students learn primarily through contact with and experience of authentic texts related to their chosen academic discipline rather than through the use of a set of prescriptive guidelines (*ibid.*: 192, 196).⁶⁸ However, there may lie an inherent contradiction in such a claim. Providing students with “authentic” genre-specific texts implies that, in the case of medical research articles, the texts used to familiarize the student with the genre are selected from sources such as the medical publications used in this study. This then raises a number of questions concerning the selection of appropriate texts: What are the selection criteria used? Do these criteria imply “appropriateness” and “typicality/atypicality,” and if so, what is the basis for such criteria? Are the selected texts used for genre familiarization and writing skills, or do they form part of the student’s general medical education, or both? And how and to what extent does a student acquire sufficient genre

⁶⁷ Lexicogrammatical features may also have differed, but they were not discussed to any great extent in Nwogu (1997) or Skelton (1994), and therefore cannot be compared.

⁶⁸ ESP and SFG also advocate the use of authentic material in the teaching of genre (Hyland 2002: 126–127), but this is usually provided in addition to descriptive frameworks that learners can use to decipher and construct texts of their own.

experience by being exposed to a selected number of these “genuine” texts? The findings of this study cannot possibly provide answers to such complex questions, but they do suggest, as discussed above, that there is good reason to provide learners of medicine, native speakers or non-native speakers of English (or any other language for that matter), with a practical framework to help decipher and construct genre-specific texts such as the medical research article. However, at the same time, it is important that learners are made aware of the potential pitfalls of following a step-by-step guide to genre, and that they are encouraged to develop their writing skills by thorough reading of genre-authentic texts, preferably selected on the basis of genre analysis.

As a final comment on the pedagogic implications of genre, it should be noted that the various theoretical and practical approaches to genre presented here and in section 2.3 are influenced by the contexts in which they are applied. Students’ needs and previously acquired skills may vary according to these contexts, and this demands different educational approaches, preferably based on detailed needs analyses (Brindley 1989).

6.3 Limitations

Before making any concluding remarks regarding the study, there are a number of important study limitations to consider, some of which have been briefly discussed in sections 6.1 and 6.2.

The first concerns the study sample. The selection criteria used were relatively restrictive (see section 3.1) and consequently limited the study material to 17 articles; the reasons for these restrictive selection criteria, e.g., years of publication, article content, etc., are given in sections 3.1, 4.1, and 4.2. A relatively small sample allowed for a more detailed (manual) analysis of individual articles than if a much larger corpus had been used. This was considered particularly important for the description of lexicogrammatical features and the identification of Themes and hyperThemes. Moreover, the size of the sample was of the same order of magnitude as that of a highly relevant previous study that was used for comparative purposes (Nwogu 1997; 15 medical research articles). However, as a result of the limited size of the study sample, the findings from this corpus cannot be claimed to be representative of the entire medical research article genre. A larger

sample might have allowed the results to be generalized with greater confidence but, within the scope of this study, that would have limited the extent of the analysis of individual articles, providing fewer data for comparison with the literature.

Another limitation with regard to the study material is related to the content of the selected articles. A specific area of medicine (health issues related to obesity and overweight) was chosen as part of the selection criteria, rather than a wider range of medical topics, in order to allow easier content summarization and comparison across articles. This was particularly important in the assignment of moves and steps (see procedure in section 4.2). Limiting the content in this way, however, may have restricted the genre analysis to a description of a very specific part of the medical discourse, namely that dealing with weight-related health issues, rather than a more general description of the medical research article. Nevertheless, part of the examination of the discourse community in section 4.2 indicated that the authors of the study articles might be considered part of a general medical discourse community rather than a more narrowly defined discourse community concerned with “health-related weight issues” (see section 4.2.5).

An additional concern regarding the content of the corpus articles is that, as a student of linguistics and not medicine, there is greater potential for misunderstanding when analyzing medical research articles, particularly when summarizing medical content and when describing certain aspects of the register and the discourse community. As discussed in sections 2.4.1 and 6.1.1, the assignment of moves and steps involves “a degree of subjectivity that is perhaps unavoidable” (Holmes 1997: 325), and this subjectivity may become more apparent if the content of the text is misunderstood. While it is possible that certain moves or steps were incorrectly assigned due to misunderstanding of the content, efforts were made to reduce this effect by analyzing the text in terms of hyperTheme. This method, although not normally used in connection with structural moves analysis, may have increased the level of certainty with which moves and steps were assigned to the texts. This potential source of error could have been avoided had the analysis been conducted in collaboration with a medical researcher or medical student, but such a measure was not feasible within the confines of a Master’s thesis. However, the possibility of misinterpretation may have been reduced by my experience and familiarity with the discourse (see section 3.3.1).

Of particular note regarding the methodology was the decision to divide the analysis into two separate parts (sections 4 and 5). This was done to emphasize the importance of context, particularly in light of previous studies that have tended to neglect the social aspects of genre in their analyses (Nwogu 1997, Skelton 1994). However, this division may be misleading, since much of the analysis presented in sections 4 and 5 was conducted simultaneously and not as two separate procedures. It is worth re-emphasizing here the statement made at the beginning of section 3, namely that text and context are interdependent and cannot easily be separated in a systemic-functional analysis such as this. For example, the register analysis in this study was restricted to contextual elements of the field, tenor, and mode, such as the discussion of participant roles and the channel of communication. However, register analysis also involves a thorough examination of the ideational, interpersonal, and textual meanings expressed in the text (see section 2.3); aspects that were instead included in section 5 as part of the moves analysis.

This study adopts a broad analytical perspective as a means of describing the generic features of the medical research article. Within the confines of this study, however, there may have been certain features that were either overlooked or examined in insufficient detail. This was a difficult balance to achieve, and it is possible that this study would have benefited by focusing on different (or perhaps fewer) features of the medical research article. Although the features selected for analysis were chosen based on a thorough review of the available literature, analyzing other lexicogrammatical characteristics such as nominalization (Banks 2003, Halliday & Matthiessen 1999) or examining non-IMRD sections such as abstracts, reference lists, and acknowledgments, or non-verbal material such as figures and tables, may have yielded interesting results.

6.4 Concluding Remarks

The multidimensional genre-based discourse analysis of this study demonstrates that the structural and thematic progression of 17 IMRD-structured medical research articles can be described in terms of 10 rhetorical moves, each of which is realized by a series of 1–6 “constituent elements” or steps (28 in total). Together, these moves and steps, seven of which were identified as mandatory (seven moves

and seven steps; see section 5.1), define the rhetorical function of the four main sections (IMRD) of the medical research article.

Moves and steps were identified in the study articles by a series of “characteristic markers.” These included content-related lexis, as well as specific Process types, Themes (and hyperThemes), clause types, Adjuncts, tense, and voice (see Tables 5.2–5.5 for section/moves summaries), and the repetition of certain markers served as a form of “rhetorical cohesion” between moves and steps, which demonstrated a complex interrelationship both within and across article sections.

A comparison of these findings with the literature revealed that there was much in common with previous register- and genre-related studies, but that there were also distinct differences in terms of the use of certain lexicogrammatical features within and across genres (see section 6.1.2), and in the moves analyses of apparently similar genres (see section 6.1.1). With regard to the latter, there was no evidence to suggest that these differences were due to changes in the genre over time (see Aims and Scope, section 1.1), since patterns of change could not be ascertained by comparing only three studies (this study, Nwogu 1997, and Skelton 1994). One change, however, which can be accounted for, and which was not identified in the corpora of Nwogu (1997) and Skelton (1994), was the LAN-only move, “role of the funding source” (see section 5.3.4). Although this move did not appear as part of the IMRD structure of the other source journals, and was therefore not included in the overall 10-move summary, a statement to this effect is provided as an endnote alongside the acknowledgements and references in *BMJ*, *JAMA*, and *NEJM*. This requirement by the ICMJE to include a statement regarding the receipt of funding and the potential conflict of interest arising from such funding was first introduced in 2001 (ICMJE 2001), which is after the publication of the material analyzed by Nwogu (1997; material published 1985–1987) and Skelton (1994; material published 1989–1993). Other differences between this study and the two previous related studies are likely due to a combination of variables, i.e., genre change, differing study material, subjectivity, and contextual/reader expectations (see sections 6.1 and 6.2 above).

It seems, then, that the findings of this study raise a number of questions and possibilities that warrant further research. In addition to the limitations and potential improvements discussed above, areas for further research might include the following: a detailed analysis of how Theme and hyperTheme can be used as

markers for identifying rhetorical moves; a diachronic investigation of the medical research article to identify changes in the genre over time; a study of how generic frameworks and genre-authentic texts are used in an educational setting; and an analysis of the extent to which editorial intervention affects the rhetorical structure of the medical research article; to name but a few.

References

- Adams Smith, D. E. (1984). Medical discourse: aspects of author's comment. *ESP Journal* 3: 25–36.
- American Psychological Association (1983). *Publication manual*. 3rd ed. Washington, D.C.: American Psychological Association.
- Askehave, I. & Swales, J. M. (2001). Genre identification and communicative purpose: a problem and a possible solution. *Applied Linguistics* 22: 195–212.
- Atkinson, D. (1992). The evolution of medical research writing from 1735 to 1985: the case of the *Edinburgh Medical Journal*. *Applied Linguistics* 13: 337–374.
- Banks, D. (2003). The evolution of grammatical metaphor in scientific writing. In: Simon-Vandenberg, A.-M., Taverniers, M., & Ravelli, L. (Eds.), *Grammatical metaphor: views from systemic functional linguistics* (pp. 127–147). Amsterdam: John Benjamins Publishing Co.
- Bazerman, C. (1984). Modern evolution of the experimental report in physics: spectroscopic articles in *Physical Review*, 1893–1980. *Social Studies of Science* 14: 163–196.
- Bazerman, C. (1988). *Shaping written knowledge: the genre and the activity of the experimental article in science*. Madison: University of Wisconsin Press.
- Bazerman, C. (1993). Intertextual self-fashioning: Gould and Lewontin's representations of the literature. In Selzer, J. (Ed.), *Understanding scientific prose* (pp. 20–41). Madison: University of Wisconsin Press.
- Biber, D. (1986). Spoken and written textual dimensions in English: resolving the contradictory findings. *Language* 62: 384–414.
- Biber, D. (1994). An analytical framework for register studies. In Biber, D. & Finegan, E. (Eds.), *Sociolinguistic perspectives on register* (pp. 31–56). Oxford: Oxford University Press.
- Biber, D. (1995). *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Bhatia, V. K. (1993). *Analysing genre: language use in professional settings*. London: Longman.

- Bhatia, V. K. (2004). *Worlds of written discourse: a genre-based view*. London: Continuum.
- Block, J. & Chi, L. (1995). A comparison of the use of citations in Chinese and English academic discourse. In Belcher, D. & Braine, G. (Eds.), *Academic writing in a second language* (pp. 231–274). Norwood, N.J.: Ablex Publishing Corporation.
- Brindley, G. (1989). The role of needs analysis in adult ESL programme design. In Johnson, R. K. (Ed.), *The second language curriculum* (pp. 63–78). Cambridge: Cambridge University Press.
- Bruce, N. J. (1984). Rhetorical constraints and information structure in medical research report writing. *English for Medical Purposes Newsletter* 1: 5–18.
- Butt, D., Fahey, R., Feez, S., Spinks, S., & Yallop, C. (2000). *Using functional grammar: an explorer's guide*. Second edition. Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Chafe, W. & Danielewicz, J. (1987). Properties of spoken and written language. In Horowitz, R. & Samuels, S. J. (Eds.), *Comprehending oral and written language* (pp. 83–113). San Diego: Academic Press.
- Christie, F. (1992). Literacy in Australia. In Grabe, W. (Ed.), *Annual Review of Applied Linguistics, 12. Literacy* (pp. 142–155). New York: Cambridge University Press.
- Couture, B. (1986). Effective ideation in written text: a functional approach to clarity and exigence. In Couture, B. (Ed.), *Functional approaches to writing: research perspectives* (pp. 69–92). London: Frances Pinter.
- Couture, B. (1993). Provocative architecture: a structural analysis of Gould and Lewontin's "The Spandrels of San Marco." In Selzer, J. (Ed.), *Understanding scientific prose* (pp. 276–309). Madison: University of Wisconsin Press.
- Daneš, F. (1974). Functional sentence perspective and the organization of the text. In Daneš, F. (Ed.), *Papers on functional sentence perspective* (pp. 106–128). Prague: Academica.
- Day, R. A. (1994). *How to write and publish a scientific paper*. 4th edition. Phoenix: Oryx Press.

- Dawson, G., Lucocq, B., Cottrell, R., & Lewison, G. (1998). *Mapping the landscape: national biomedical research outputs 1988–1995*. London: Wellcome Trust.
- Dubois, B. L. (1997). *The biomedical discussion section in context*. Advances in Discourse Processes Series. Greenwich, C.T.: Ablex.
- Dudley-Evans, T. (1987). Guest editor's introduction. *ELR Journal* 1: 1–9.
- Ferguson, C. A. (1994). Dialect, register, and genre: working assumptions about conventionalization. In Biber, D. & Finegan, E. (Eds.), *Sociolinguistic perspectives on register* (pp. 15–30). Oxford: Oxford University Press.
- Fries, P. H. (1994). On Theme, Rheme and discourse goals. In Coulthard, M. (Ed.), *Advances in written text analysis* (pp. 229–249). London: Routledge.
- Freedman, A. (1994). “Do as I say”: the relationship between teaching and learning new genres. In Freedman, A. & Medway, P. (Eds.), *Genre and the new rhetoric* (pp. 191–210). London: Taylor & Francis.
- Freedman, A. & Medway, P. (1994). Locating genre studies: antecedents and prospects. In Freedman, A. & Medway, P. (Eds.), *Genre and the new rhetoric* (pp. 1–20). London: Taylor & Francis.
- Garfield, E. (1994). The ISI impact factor. Available at URL: <http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor> (accessed: September 28, 2006).
- Ghadessy, M. (1988). The language of written sports commentary: soccer—a description. In Ghadessy, M. (Ed.), *Registers of written English: situational factors and linguistic features* (pp. 17–51). London: Pinter Publishers.
- Ghadessy, M. (1999). Textual features and contextual factors for register identification. In Ghadessy, M. (Ed.), *Text and context in functional linguistics* (pp. 125–139). Amsterdam: John Benjamins.
- Gosden, H. (1992). Discourse functions of marked themes in scientific research articles. *English for Specific Purposes* 11: 207–224.
- Gross, A. G. (1990). *The rhetoric of science*. Cambridge, MA: Harvard University Press.
- Hall, G. M. (2003). *How to write a paper*. Third edition. London: BMJ Publishing.
- Halliday, M. A. K. (1978). *Language as social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.

- Halliday, M. A. K. (1988). On the language of physical science. In Ghadessy, M. (Ed.), *Registers of written English: situational factors and linguistic features* (pp. 162–178). London: Pinter Publishers.
- Halliday, M. A. K. (1993). Some grammatical problems in scientific English. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 69–85). London: Falmer Press.
- Halliday, M. A. K. (1993a). On the language of physical science. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 54–68). London: Falmer Press.
- Halliday, M. A. K. (1993b). Some grammatical problems in scientific English. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 69–85). London: Falmer Press.
- Halliday, M. A. K. & Hasan, R. (1985). *Language, context, and text: aspects of language in a social-semiotic perspective*. Geelong: Deakin University Press.
- Halliday, M. A. K. & Martin, J. R. (1993). General orientation. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 2–21) London: Falmer Press.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (1999). *Construing experience through meaning: a language-based approach to cognition*. London: Continuum.
- Halliday, M. A. K. & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. Third edition. London: Arnold.
- Hasan, R. (1996). The nursery tale as a genre. In Cloran, C., Butt, D., & Williams, G. (Eds.), *Ways of saying: ways of meaning* (pp. 51–72). London: Cassell.
- 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* 16: 321–337.
- Hopkins, A. & Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes* 7: 113–121.
- Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics* 20: 341–367.

- Hyland, K. (2002). Genre: language, context and literacy. *Annual Review of Applied Linguistics* 22: 113-35.
- Hyon, S. (1996). Genre in three traditions: implications for ESL. *TESOL Quarterly* 30: 693–722.
- Institute for Scientific Information (2005). *Journal citation reports: medical – general*. Report available on CD-ROM. Philadelphia: Thomson Scientific.
- ICMJE (International Committee of Medical Journal Editors) (2001). Sponsorship, authorship, and accountability. *Annals of Internal Medicine* 135: 463–466.
- ICMJE (International Committee of Medical Journal Editors) (2006). *Uniform requirements for manuscripts submitted to biomedical journals: writing and editing for biomedical publication*. Philadelphia: ICMJE. Available at URL: www.icmje.org (accessed: September 15, 2006).
- Iverson, C., Flanagan, A., & Fontanarosa, P. B. (1998). *American Medical Association manual of style: a guide for authors and editors*. Baltimore, M.D.: Williams & Wilkins.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24: 269–292.
- Krashen, S. (1982). *Principles and practice in second language learning and acquisition*. Oxford: Pergamon Press.
- Kreutzberg, G. W. (2005). Scientists and the marketplace of opinions. *European Molecular Biology Organization [EMBO] Reports* 6: 393–396.
- Lewin, B. A., Fine, J. & Young, L. (2001). *Expository discourse: a genre-based approach to social science research texts*. London: Continuum.
- Maher, J. (1986). The development of English as the international language of medicine. *Applied Linguistics* 7: 206–218.
- Martin, J. R. (1985). *Factual writing: exploring and challenging social reality*. Geelong: Deakin University Press.
- Martin, J. R. (1993a). Life as a noun: arresting the universe in science and humanities. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 221–267). London: Falmer Press.
- Martin, J. R. (1993b). Literacy in science: learning to handle text as technology. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: literacy and discursive power* (pp. 166–202). London: Falmer Press.

- Martin, J. R. (2000). Design and practice: enacting functional linguistics. *Annual Review of Applied Linguistics* 20: 116–126.
- Martin, J. R. (2001). Language, register and genre. In Burns, A. & Coffin, C. (Eds.), *Analysing English in a global context: a reader* (pp. 149–166). London: Routledge.
- Martin, J. R. & Rose, D. (2003). *Working with discourse: meaning beyond the clause*. London: Continuum.
- Matthiessen, C. M. I. M. & Halliday, M. A. K. (1997). Systemic functional grammar: a first step into the theory. Available at URL: <http://minerva.ling.mq.edu.au> (accessed: November 17, 2006).
- Miller, D. R. (2004). *Language as purposeful: functional varieties of texts*. Bologna: University of Bologna. Available at URL: http://amsacta.cib.unibo.it/archive/00000866/01/language_as_purposeful-donna_miller.pdf (accessed: December 10, 2006).
- Moss, F. (2003). Titles, abstracts, and authors. In Hall, G. M. (Ed.), *How to write a paper* (pp. 42–50). Third edition. London: BMJ Publishing.
- Nwogu, K. N. (1991). Structure of science popularization: a genre-analysis approach to the schema of popularized medical texts. *English for Specific Purposes* 10: 111–123.
- Nwogu, K. N. (1997). The medical research paper: structure and functions. *English for Specific Purposes* 16: 119–138.
- Peng, J. (1987). Organisational features in chemical engineering research articles. *English Language Research Journal* 1: 79–116.
- Rennie, D., Flanagan, A., & Yank, V. (2000). The contributions of authors. *Journal of the American Medical Association* 284: 89–91.
- Romaine, S. (1994). On the creation and expansion of registers: sports reporting in Tok Pisin. In Biber, D. & Finegan, E. (Eds.), *Sociolinguistic perspectives on register* (pp. 59–81). Oxford: Oxford University Press.
- Sollaci, L. B. & Pereira, M. G. (2004). The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *Journal of the Medical Library Association* 92: 364–371.
- Skelton, J. (1994). Analysis of the structure of original research papers: an aid to writing original papers for publication. *British Journal of General Practice* 44: 455–459.

- Swales, J. M. (1981). *Aspects of article introductions*. Birmingham: University of Aston Language Studies Unit.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales, J. M. (2004). *Research genres: explorations and applications*. Cambridge: Cambridge University Press.
- Swales, J. M. & Feak, C. B. (2000). *English in today's research world: a writing guide*. Ann Arbor, MA: University of Michigan Press.
- Tarone, E., Dwyer, S., Gillette, S., & Icke, V. (1998). On the use of the passive and active voice in astrophysics journal papers: with extensions to other languages and other fields. *English for Specific Purposes* 17: 113–132.
- Thompson, G. (2004). *Introducing functional grammar*. Second edition. London: Arnold.
- Thompson, P. & Tribble, C. (2001). Looking at citations: using corpora in English for academic purposes. *Language Learning & Technology* 5: 91–105.
- Thompson, S. (1994). Frameworks and contexts: a genre-based approach to analysing lecture introductions. *English for Specific Purposes* 13: 171–186.
- Wager, E. (2005). *Getting research published: an A to Z of publication strategy*. Oxford: Radcliffe Publishing.
- WAME (World Association of Medical Editors) (2007). *WAME policy statements*. Available at URL: <http://www.wame.org/resources/policies/> (accessed: February 23, 2007).
- Young, P. (1996). *The art and science of writing: a handbook for health science students*. London: Chapman & Hall.
- Yunnik, S. (1997). Genre, registers and sociolinguistics. *World Englishes* 16: 321–336.

Appendix

Excerpt from: International Committee of Medical Journal Editors (2006) *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication* (pp. 26–28). Toronto: ICMJE. Numbering has been added for comparative purposes (see section 6.1 and Table 6.1).⁶⁹

IV.A.5. Introduction

1) Provide a context or background for the study (i.e., the nature of the problem and its significance). 2) State the specific purpose or research objective of, or hypothesis tested by, the study or observation; the research objective is often more sharply focused when stated as a question. Both the main and secondary objectives should be made clear, and any pre-specified subgroup analyses should be described. Give only strictly pertinent references and do not include data or conclusions from the work being reported.

IV.A.6. Methods

The Methods section should include only information that was available at the time the plan or protocol for the study was written; all information obtained during the conduct of the study belongs in the Results section.

IV.A.6.a. Selection and Description of Participants

3) Describe your selection of the observational or experimental participants (patients or laboratory animals, including controls) clearly, including eligibility and exclusion criteria and a description of the source population. Because the relevance of such variables as age and sex to the object of research is not always clear, authors should explain their use when they are included in a study report; for example, authors should explain why only subjects of certain ages were included or why women were excluded. The guiding principle should be clarity about how and why a study was done in a particular way. When authors use variables such as race or ethnicity, they should define how they measured the variables and justify their relevance.

⁶⁹ Cited with permission for “educational, not-for-profit purposes” from the International Committee of Medical Journal Editors (ICMJE 2006: 3).

IV.A.6.b. Technical information

4) Identify the methods, apparatus (give the manufacturer's name and address in parentheses), and procedures in sufficient detail to allow other workers to reproduce the results. Give references to established methods, including statistical methods (see below); provide references and brief descriptions for methods that have been published but are not well known; describe new or substantially modified methods, give reasons for using them, and evaluate their limitations. Identify precisely all drugs and chemicals used, including generic name(s), dose(s), and route(s) of administration. Authors submitting review manuscripts should include a section describing the methods used for locating, selecting, extracting, and synthesizing data. These methods should also be summarized in the abstract.

IV.A.6.c. Statistics

5) Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results. When possible, quantify findings and present them with appropriate indicators of measurement error or uncertainty (such as confidence intervals). Avoid relying solely on statistical hypothesis testing, such as the use of P values, which fails to convey important information about effect size. References for the design of the study and statistical methods should be to standard works when possible (with pages stated). Define statistical terms, abbreviations, and most symbols. Specify the computer software used.

IV.A.7. Results

6) Present your results in logical sequence in the text, tables, and illustrations, giving the main or most important findings first. Do not repeat in the text all the data in the tables or illustrations; emphasize or summarize only important observations. Extra or supplementary materials and technical detail can be placed in an appendix where it will be accessible but will not interrupt the flow of the text; alternatively, it can be published only in the electronic version of the journal.

When data are summarized in the Results section, give numeric results not only as derivatives (for example, percentages) but also as the absolute numbers from which the derivatives were calculated, and specify the statistical methods used to analyze them. Restrict tables and figures to those needed to explain the argument of the

paper and to assess its support. Use graphs as an alternative to tables with many entries; do not duplicate data in graphs and tables. Avoid nontechnical uses of technical terms in statistics, such as “random” (which implies a randomizing device), “normal,” “significant,” “correlations,” and “sample.” Where scientifically appropriate, analyses of the data by variables such as age and sex should be included.

IV.A.8. Discussion

Emphasize the new and important aspects of the study and the conclusions that follow from them. Do not repeat in detail data or other material given in the Introduction or the Results section. For experimental studies it is useful to begin the discussion by 7) summarizing briefly the main findings, then 8) explore possible mechanisms or explanations for these findings, 9) compare and contrast the results with other relevant studies, 10) state the limitations of the study, and 11) explore the implications of the findings for future research and for clinical practice. 12) Link the conclusions with the goals of the study but avoid unqualified statements and conclusions not adequately supported by the data. In particular, authors should avoid making statements on economic benefits and costs unless their manuscript includes the appropriate economic data and analyses. Avoid claiming priority and alluding to work that has not been completed. State new hypotheses when warranted, but clearly label them as such.

Index

Page numbers in bold indicate the introduction and definition of a term.

- academic writing, 25
- Adjunct, **13**, 14, 15, 16, 17, 18, 20, 127
- adverbial group, 18
- aspect, 48
- Circumstance, 7, 9, 18, 19, 22
- clause, 7, 9, 17, 19, 20, 21
 - embedded clause, 12
- clause complex, 10, 12, 20
- cohesion, 17, 21, 48
- cohesive devices, 19
- collocation, 50
- communicative purpose, 28, 36, 45
- Complement, **13**, 14
- conjunction, 11, 17, 19, 20, 22
- context, 46, 47, 121
- context of situation, 6, 22, 46, 47, 51
- continuative, 17
- coordination, 11
- definite article, 20
- diachronic, 34, 128
- dimensions of variation. *See* register
- discoursal expectation, 57
- discourse, **28**, 115
- discourse community, 28, 36, 46, 47, 55
 - expert member, 59
 - novice, 59
- editorial intervention, 41, 128
- elision. *See* ellipsis
- ellipsis, 19, 20, 21
- English for Specific Purposes (ESP), **27–31**, 35, 39
- ethnographic, 34, 39, 46
- expansion, 10
- Finite operator, 13, 14, 15
- foregrounding, 18, 22, 49
- fronting. *See* foregrounding
- function, 6
- function label, 7
- generic integrity, 41
- genre, **27–35**, 39
 - ideology, 32
 - pedagogy, 27, 121–24
 - sites of application, 32
- genre-register conflation, 31
- grammatical metaphor, 27
- hyperTheme, 18, 19, 22, 48, 115, 120, 127
- ICMJE, 38, 47
- IMRD structure, 38, 57, 61, 127
- indefinite article, 20
- journal impact factor, 41
- lexical density, 27, 57
- lexicogrammar, 25, 36, 46, 47, 49, 50, 115, 116
- lexis, 127
- linguistic token. *See* substitution
- medical English, 5
- metafunction, **6–22**
 - ideational, **6–12**, 23
 - experiential, 7, 6–10
 - logical, **10–12**
 - logico-semantic relationship, 10
 - interpersonal, **12–16**, 23
 - textual, **16–22**, 23
- modality, 12, 14, 37, 117
 - modal Finite, **14**, 16
- mood, 12, 13
 - imperative, **12**, 14
 - indicative, **14**
 - declarative, 12, 14, 16, 18
 - interrogative, 12
 - WH- interrogative, 13
 - yes/no interrogative, 13, 14
 - queclarative, 14
- Mood block, 13, 14, 15, 16
- move, **29**, 49, 126
 - common, 50
 - move 1, 64
 - move 10, 104
 - move 2, 66
 - move 3, 68
 - move 4, 72
 - move 5, 81
 - move 6, 87
 - move 7, 92
 - move 8, 98
 - move 9, 103
 - subjectivity, 29, 114
- move order stability, 50
- multidimensional model, 6, 39, 45, 46, 126
- New Rhetoricians, **33–34**, 39
- New South Wales Disadvantaged Schools Program, 26
- nominal group, 18, 20, 22, 78, 81
- Participant, 7, 8, 9, 17, 19, 21, 22
- passive, 26
- pedagogy, 26
- personals, 20, 26, 37, 48, 109, 117
- possessive determiner, 20, 22
- Predicator, **13**, 14
- prepositional phrase, 18, 19
- Process, 7, 8, 9, 37, 116, 127
 - behavioral, 8
 - existential, 8
 - material, 7, 116
 - mental, 7, 11
 - relational, 7, 117
 - circumstantial, 8
 - intensive, 8
 - possessive, 8
 - verbal, 8, 11
- projection, 10
- reference, 19, 20
 - anaphoric, 20, 22
 - chains of reference, 20
 - exophoric, 22

- integral, 49
- non-integral, 49
- referent, 20
- study-external, 116
- study-internal, 115
- register, **23**, 22–27, 39, 45, 51
 - field, 23, 47, 51, 55
 - mode, 23, 47, 54, 55
 - tenor, 23, 47, 52, 55
- repetition, 21
- Residue, 13, 15, 16
- Rheme, 17, 21
- rhetorical action, 33
- rhetorical movement, 29
- rhetorical purpose, 37
- step, **29**, 49, 126
 - step 10a, 104
 - step 10b, 105
 - step 10c, 107
 - step 1a, 64
 - step 2a, 66
 - step 2b, 67
 - step 3a, 69
 - step 3b, 70
 - step 4a, 72
 - step 4b, 74
 - step 4c, 75
 - step 4d, 76
 - step 4e, 78
 - step 4f, 79
 - step 5a, 81
 - step 5b, 82
 - step 5c, 84
 - step 5d, 85
 - step 6a, 87
 - step 6b, 88
 - step 7a, 92
 - step 7b, 93
 - step 7c, 94
 - step 7d, 95
 - step 8a, 98
 - step 8b, 100
 - step 8c, 101
 - step 9a, 103
- structural moves analysis, **29–31**, 35, 45, 61, 111, 114
 - Create a Research Space (CARS) model, 29
 - four-move model, 29
- study limitations, 124–26
- study material, 41
- study objective, 2
- Subject, 13, 16, 17, 18, 21, 22
- submove. *See* step
- subordination, 11
- substitution, 21
 - pro-form, 21
 - pro-verb, 21
- Sydney School. *See* Systemic-Functional Genrists (SFG)
- Systemic-Functional Genrists (SFG), 31–33, 39
- systemic-functional linguistics (SFL), **6–22**, 46, 50
- taxis, 10
 - hypotaxis, 10, 11
 - parataxis, 10, 11
- teaching–learning cycle, 33
- technical taxonomy, 27
- tense, 37, 48, 116, 117, 127
- text, **5**, 46
- text type, 5, **35**
- thematic dependent clause, 21
- thematic progression, 18, 19
- thematic structure, 17
- Theme, 17, 22, 37, 48, 120, 127
 - experiential, 17, 18, 19, 21, 22
 - interpersonal, 18
 - marked, 18, 19, 21, 49, 120
 - multiple, 18
 - textual, 17
 - T-unit, 21
 - unmarked, 22
- Vancouver Group. *See* ICMJE
- vocative, 18
- voice, 37, 48, 118, 127
- WH- element, 14
- zero article, 20