Scope and Limits of Corpus-Based Studies of the Medical Research Article for Teaching ESL Student Researchers: Discussion and Example

Gavin Melles
Faculty International Unit, Medicine, Dentistry and Health Sciences
University of Melbourne, Australia
gmelles@unimelb.edu.au

Abstract: Medical research and knowledge is communicated among the discourse community of practitioners and researchers through spoken and written genres, including peer-reviewed research publications. Such publications are required reading for student researchers in medicine, who must learn to critically analyse and summarise this information while attending to the characteristics of this genre. English as a Second Language (ESL) students must learn to critically read such texts and manage the linguistic difficulties of medical language. With this audience in mind, corpus based applied linguistics has investigated the lexical and rhetorical specifics of the genre. These findings have contributed to our knowledge of the linguistic and rhetorical nature of texts of this genre, which can complement other instructional resources required in applied linguistic higher education contexts.

In this paper, I describe the principles of genre-based writing pedagogy, the contribution of corpus-based studies to our understanding of the medical research genre, and I describe the construction of an LSP (language for specific purposes) corpus of medical research texts for analysis and teaching purposes.

Medical Knowledge and Research Texts as Genres

Medical research and writing is not homogenous but authored and legitimated by different disciplinary cultures (Becher, 1989) although there is an overall tendency to privilege empirical ‘scientific’ methods (Lewinsohn, 1998) and random controlled trials according to the predominant bio-medical paradigm (Dawson-Saunders & Trapp, 2001). Teaching of research methods in medicine also contributes to maintain the conventional preference for quantitative experimental work (Dean, 2004). Any representative collection of medical research texts would have to reflect this diversity albeit simultaneously acknowledging the way in which the ‘scientific’ paradigm is privileged.

The concept of discourse community, partly derived from work in sociolinguistics on speech communities (Hudson, 1996), has proved useful in genre-based pedagogy as a description of the social networks (community) in academic and professional domains who communicate according to spoken and written conventions (Swales, 1990). Citing Bizzell (1992), Hyland (2000) notes that discourse community ‘seeks to locate writers in particular contexts to identify how their rhetorical strategies are dependent on the purposes, setting and audience of writing (p.9), and recognizes that it should not be seen in monolithic terms but ’implies a certain degree of interdisciplinary diversity and a degree of intradisciplinary homogeneity’ (p.10 ). Text types such as the research article are representations of such historically validated genres and corpora-based study of medical texts and language should reveal both the diversity and homogeneity in the selection and analysis of texts.

Genre-based approaches to teaching academic writing have taken up this recognition of diversity within disciplinary cultures as a basis for teaching second language students acceptable generic and specific conventions (Benesch, 1995; Hyland, 2003; Ramanathan & Kaplan, 2000). In addition to a focus on the text themselves and the social and professional conventions that limit acceptable text forms, work with second language writers has also helped to make transparent the cultural
assumptions and cross-cultural confrontations that students entering a discourse community may encounter.

Socio-cultural approaches to composition in the disciplines are far more adequate to teaching ESL students in higher education than the vocabulary and conventional grammar based texts of medical language (eg. Glendinning & Holmström, 1987; Maher, 1990). Such texts work with theme-based extracts from a range of genres, including newspapers, and prescriptive definitions of rhetorical and linguistic conventions. Corpora-based analysis of medical texts have demonstrated that language use and rhetorical structuring in the disciplines is both different and far more complex than prescriptive summaries suppose. For rhetorical analysis of research articles a conventional four-part structure of introduction, methods, results and discussion (often abbreviated as IMRaD) is used (Dubois, 1997; Skelton, 1994; Swales, 1990).

Corpus-Based Pedagogy and Language for Specific Purposes

Corpus-based linguistics harnesses the power of computers to store text corpora and extract the linguistic characteristics of domain specific texts, which are then analyzed through the use of concordancing programs (McEnery & Wilson, 1996). Text analysis is contributing to the study of register variation (Biber, 1995), and studies of language and culture (Stubbs, 1996) in sociolinguistics. Register, as a linguistic dimension, interacts with genre in that texts as the product of discourse community interactions reflect the predominant registers such communities validate in their communication. Registers are collections of linguistic features, e.g. personal pronoun use, subsets of adjectives or adverbs, etc., seen to characterize spoken and written language use defined according to situation. Register clusters of features described as conversational and scientific, are used in constructing large scale corpora like the British National Corpus (Aston & Burnard, 1998), which incorporates 100 million words of 2000 word sample texts of present-day spoken and written English. The spoken part is subdivided into two registers – task oriented and conversational, while the written section is divided into imaginative and informative registers (Leech, Wilson, & Rayson, 2001).

Specifically in the area of English for Specific Purposes (ESP) corpus-based studies of register variation aims to show how texts from different academic disciplines vary linguistically and how such texts also vary within themselves (Biber, Conrad, & Reppen, 1998, p.138). Such studies have shown, for example, that within the academic disciplines on a dimension such as impersonal : personal characterised by, for example, the use or non use of first person pronouns academic disciplines vary; such studies also show variation within sections of research articles, eg. introduction and methods, with respect to such dimensions (Biber et al., 1998, p.167). Such knowledge has complemented genre studies of the research article in the pedagogy of academic writing (Swales, 1990), where genre, eg. biology research article, connects discipline with situated text use (Lee, 2001).

Despite ongoing concerns for uptake among the teaching profession (McEnery & Wilson, 1997), corpus linguistics has influenced the production of resources for language pedagogy (Boswood, 1997; G. D. Kennedy, 1998; Meyer, 2002), including corpus-based textbooks on academic writing (Thurstun & Candlin, 1997), corpora of learner English for analysis (Granger, 1994), reference grammars of spoken and written English (Biber, 1999), and academic wordlists (Coxhead, 2000). As Biber (2001) notes corpus analyses ‘of representative corpora provide a much more solid foundation for descriptions of language use’ (p.101), than language teachers are often used to. Typically such work involves both a quantitative component in measuring frequency of word use and a qualitative approach to defining language functions. As Biber et al. (1998) note ‘a crucial part of the corpus-based approach is going beyond the quantitative patterns to propose functional interpretations explaining why the patterns exist’ (Biber et al., 1998, p.9).

In language teaching, the lexical phrase and word collocations (Burdine, 2001; Lewis, 1993; Sinclair, 1991) have become the basis for what Johns calls data-driven learning (DDL) or classroom concordancing (Johns, 1994). This is an approach which identifies the words that typically co-occur with a set of key content words, eg. a subset of key academic words used in writing, such as ‘analysis’
(Thurston & Candlin, 1997) or other high frequency academic words (McKay, 1982), and explores how they are used based on authentic text extracts; good examples of DDL in use can be seen at http://web.bham.ac.uk/johnstf/timconc.htm. As with other aspects of language use, native (and teacher) intuitions about collocations can be inaccurate (Stubbs, 1996, p.172) but can be quantified using computers and relevant statistical measures (Oakes, 1998). Although collocations and the teaching of idiomatic phrases have an established place in language pedagogy, the use of corpora with higher education students for DDL language learning have shown mixed results (eg. C. Kennedy & Miceli, 2001).

Large General Corpora (LGP) and Language for Specific Purposes (LSP)

The construction of large scale general corpora (LGP) employing text sampling criteria is well documented in the literature (eg. Aston & Burnard, 1998; G. D. Kennedy, 1998; Sinclair, 1991). The construction of LSP (language for specific purposes) corpora, however, specifically adapted to the needs and goals of academic purposes projects is less well defined. LSP corpora, which are generally smaller and more selective than other corpora, have specific inclusion and exclusion criteria. Bowker and Pearson (2002) recommend written LSP corpora of greater than 25000 words, inclusion of full texts (not extracts), authorship spread, and use of a single text type (p.54). Constructing representative corpora requires both specification of practical aims and inclusion and exclusion criteria. The corpus-base of the Academic Word List (AWL) and the MICASE (Michigan Corpus of Academic Spoken English), will be used to illustrate. Coxhead (2000) and the AWL website (http://www.vuw.ac.nz/lals/staff/averil-coxhead/awl/index.html) describes inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Table 1: AWL Corpus Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total word count</td>
</tr>
<tr>
<td>Four faculties</td>
</tr>
<tr>
<td>Genres</td>
</tr>
<tr>
<td>Total texts</td>
</tr>
<tr>
<td>Text structure</td>
</tr>
</tbody>
</table>

The on-line MICASE (http://www.isa.umich.edu/ell/micase/MICASEStats.htm) statistical overview contains four tables referring to the spread of categories used to collect and present the data and the corpus of 1.7 million words. One table is presented below. Speech events are subdivided into 16 categories and four primary discourse modes (monologic, panel, interactive and mixed). Further details can be seen at the site.
Table 2: MICASE word counts according to academic division

<table>
<thead>
<tr>
<th>Academic Division</th>
<th>Speech Events</th>
<th>Speakers</th>
<th>Words</th>
<th>% of Total Corpus</th>
<th>% Male</th>
<th>% Female</th>
<th>% Faculty*</th>
<th>% Students*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanities &amp; Arts</td>
<td>36</td>
<td>349</td>
<td>434,669</td>
<td>26</td>
<td>56</td>
<td>44</td>
<td>63</td>
<td>29</td>
</tr>
<tr>
<td>Social Sciences &amp; Education</td>
<td>35</td>
<td>452</td>
<td>420,347</td>
<td>25</td>
<td>37</td>
<td>63</td>
<td>44</td>
<td>55</td>
</tr>
<tr>
<td>Biological &amp; Health Sciences</td>
<td>32</td>
<td>257</td>
<td>325,456</td>
<td>19</td>
<td>41</td>
<td>59</td>
<td>55</td>
<td>42</td>
</tr>
<tr>
<td>Physical Sciences &amp; Engineering</td>
<td>36</td>
<td>314</td>
<td>358,776</td>
<td>21</td>
<td>55</td>
<td>45</td>
<td>44</td>
<td>52</td>
</tr>
<tr>
<td>Other/NA</td>
<td>13</td>
<td>199</td>
<td>156,292</td>
<td>9</td>
<td>37</td>
<td>63</td>
<td>20</td>
<td>41</td>
</tr>
<tr>
<td>Totals</td>
<td>152</td>
<td>1,571</td>
<td>1,695,540</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The corpus is being used for applied linguistic research and discourse analysis as well as for direct pedagogical purposes. It is a spoken corpus and highly detailed in terms of searchable categories and uses. As is clear from the above table, an attempt has been made to have a representative spread of texts, speech events, and academic division. Medical research texts form a much more specific written genre.

Corpus-Based Studies of the Language of Medical Texts: A Brief Review

Corpus analysis of medical texts has an established history in medical informatics (eg. Baud, Lovis, Rassinoux, Michel, & Scherrer, 1998; Happe, Pouliquen, Burgun, Cuggia, & Le Beux, 2003), and cognitive psycholinguistics (eg. Chiarello, Shears, & Lund, 2000; Desmet, Brysbaert, & De Baecke, 2002). With few exceptions (eg. Busch-Lauer, 1995) such studies do not draw pedagogical conclusions on the relevance of findings for teaching purposes. Fewer corpus-based studies of medical research texts exist within applied linguistics, and although they provide important information for EAP teachers this implication is not always clearly articulated in the relevant study. The main findings have been about linguistic and rhetorical properties of texts based on word analysis. In what follows, I highlight the nature, content, findings and pedagogical implications of published corpus-based studies of medical language.

Adams Smith (1984), in an early study to determine the difference between objective statement of fact and subjective opinion, examined a selection of papers from the British Medical Journal for linguistic markers of (subjective) author comment. They found a variety of verbal and non-verbal modals and attitudinal markers used which varied across text types, such as editorials, clinical case notes. In research papers, they note the objective nature of methods and results section and a greater degree of subjective comment in introductions and particularly discussion sections. They conclude that such findings should be reflected in teaching materials for students of medical English.

Salager-Meyer et al. (1989) examine the communicative function of 17 grammar variables, eg. verb tenses, in 51 medical texts across three text types – case reports, editorials and research papers. They provide three ‘axes’ contrasting present/future tense and past tense/passive voice, modals, and use of gerunds. These three dimensions are used to distinguish the three text types analysed. Advice about the preparation of teaching materials is given at the conclusion of the study. This study can be compared to the multidimensional analysis of Biber and Finegan (1994) discussed below. Salager-Meyer (1994) takes up the communicative function of hedges in medical texts in the different rhetorical
sections of research papers and case reports. The author analyses a corpus of 15 articles drawn from 5 medical journals to analyse hedge and found that discussion and comment sections are the most hedged sections and in specific sections there is a tendency for specific verbs to be used for hedging. The author points out the importance of his findings for EAP teaching purposes, including translation exercises. This study can be compared to the work of Vartalla (1999) reported below.

Biber & Finegan (1994) investigated linguistic variation in the IMRD sections of a corpus of 20 medical research texts published in 1985; 10 texts taken from New England Journal of Medicine and 10 from the Scottish Medical Journal. They note several linguistic peculiarities of the total corpus, such as present tense use being much higher in introductions and discussions and that past tense predominates in the other two (methods and results); agentless passives (passives with no ‘by-phrase’) are also dominant in the methods section. The authors also construct several ‘functional’ dimensions, such as ‘involved versus informational production’ and ‘narrative versus non-narrative discourse’ using specific linguistic indices to distinguish article sections. They note, for example, that introduction sections are more narrative like than other sections of the article. Some observations are also made on rhetorical variation between the two journals. Although they make no reference to it, this multidimensional linguistic and functional analysis of medical texts provides important information for EAP teachers about linguistic conventions.

Williams (1996), using 8 texts – four clinical and four experimental, examined lexical verb use in medical research reports; texts totalled 25000 words. 28 verbs were found in both groups of article and classified functionally as to whether they were used for reporting, observation, defining, relations, cause and effect, change and growth, and methods. Significant differences were found for five of these verbs in terms of their frequency in clinical and experimental articles. The article includes qualitative explanations of examples in context. Williams concludes, in that ‘differences in the communicative purpose and its textual realization between medical research types may be greater than previously assumed’ (Williams, 1996, p.195), with consequences for teaching.

Nwogu (1997) in a rhetorical structure analysis of research introductions by Swales (1990) used a corpus of 35 texts from five internationally refereed journals (most are included in my corpus, see below). Comparing his analysis to Swales (1990) four-part move analysis, he enumerates eleven potential rhetorical moves in such texts all bounded by linguistic indicators. Although Nwogu sees his contribution in terms of linguistic confirmation of what the medical discourse community knows by experience; also see Nwogu (1991) for a similar study of popularised medical writing. No pedagogical claims made but his conclusions have obvious applications to EAP pedagogy both of reading and writing in terms of teaching move.

Vartalla (1999), with a corpus of 15 texts from Scientific American and NEJM, explores the use of hedging devices in medical texts. The author concludes not only that hedging (using verbs, eg. suggest, modal auxiliaries, eg. might, and adverbs, eg. perhaps, etc.) is common in both popular and specialist texts there are differences in communicative function between the two genres and this corresponds to features of audience and purpose. He claims that this ‘idea has been largely unaccounted for in ESP textbooks and research, but its implications for ESP research and teaching merit careful consideration’ (Varttala, 1999, p.190).

Gledhill (2000) and colleagues designed a large pharmaceutical sciences corpus of 150 cancer research articles to examine collocations in introductions of articles. They enumerate specific high frequency collocations in such articles and also insist that the construction of representative LSP medical corpora is essential. The analysis focused on the left and a right collocation of high frequency grammatical words in research article introductions, such as been, have, and is, attempting to provide functional explanations for word use. They enumerate the methodological stages required in the compilation of the corpus and the criteria for selection. For learners, they conclude that ‘the phraseological units we have identified are formulated by previous discourses and must be acquired or learnt by the community’ (p. 131). They also note the formulaic characteristics of science texts which such analyses foreground needs to be taught alongside the potential for variation in writing that is part of the dynamic nature of scientific genres.
Ferguson (2001) using a 100,000 word corpus examines the use of *if* conditionals in medical discourse. Three genres (two written and one spoken) are used for comparison: research articles, journal editorials and doctor-patient consultations. A total of 177 such conditionals were found and differences in use between the two mediums. Conditionals in spoken interaction are used for demonstrating politeness and sensitivity while in the methods section of research papers they are used for giving operational definitions. The findings have implications for teaching ESL learners about genre and medium specific uses of grammar structures.

Salager-Meyer (Salager-Meyer, Alcaraz Ariza, & Zambrano, 2003) has recently extended his earlier synchronic studies of language variation to consider cross-cultural differences and historical change in the medical discourse of English, French and Spanish medical texts. Using a corpus of historical texts (1930-1995), the authors find more critique, authority and ‘passion’ in the commitment of French and Spanish writers to textual claims compared to their Anglo-Saxon colleagues. They note, however, a convergence of Spanish writers towards the more neutral hedging style of Anglo-Saxon writers during the last decade of the twentieth century. They draw conclusions about the implications this has for how science has been increasingly marketed as a competitive professional global market.

In addition to the above studies, Thompson (1994) describes the construction of a medical corpus – MEDERA (Medical Experimental Research Articles) – aimed specifically at advanced learners of English in medical faculties. Using a genre-analysis approach the materials are described as self-access materials to aid students in constructing their own research texts. No current information was available. Finally, Vihla (1998) reports on the construction of a corpus of American medical texts (397,311 words) including a range of professional and popular texts, including textbook samples and newspaper articles. The corpus, intended for her PhD research, is intended to be generally representative of American medical discourse; no specific teaching purpose is mentioned nor could I discover whether the corpus had become publicly available.

**Limitations of Existing Corpus Based Studies**

Thus, existing corpus-based and genre specific studies of medical research texts have identified a number of key lexical, phraseological, and rhetorical features of this genre. Such findings have implications for the pedagogy of higher education writing and can complement or substantially modify existing approaches to teaching the genre. In many cases, however, no detailed explanations for text choice have been given, however, and multiple genres, eg. editorial and case study, have been combined rather than focusing on specific genres. In addition, although analyses have clear pedagogical implications these are not always addressed by authors. In educational contexts, ultimately all criteria are subordinate to ‘the goals of your project’ (Bowker & Pearson, 2002, p.53) although practical considerations are not often carefully addressed in the literature (but see Tribble, 1997).

From the perspective of genre as a social process, even such sophisticated analyses of the genre as linguistic products excludes the social interaction and processes of learning that constitute the learning of disciplinary discourses (even though some corpus-based studies refer to their linguistic findings as evidence of discourse conventions). Also, in the case of medical research texts tables, graphs, and photographic images are an essential element of the texts as genre and are usually excluded from text-based analyses. The exclusion of images and other elements of the ‘staging’ (Brown & Yule, 1983, p.125-152) of academic texts in the analysis of academic writing for second language learners is also an unfortunate omission (eg. Belcher & Braine, 1995) that reduces the text to a form which is a poor representation of text as genre. Finally, a danger with corpora within strict time limits, which are then made publicly available, is that such closed corpora may falsely reflect a static image of language use in disciplinary contexts.

To address some of the limitations mentioned above, to add to existing studies, and to complement my own work with medical research students, I have worked towards the construction of an open
corpus of medical research texts. Following Bowker and Pearson (2002), the corpus has been constructed to match eight criteria outlined in the table below.

Table 3: Key dimensions in constructing an LSP (after Bowker & Pearson, 2002)

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Considerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>Determined by project needs, availability of data, and time available; a small well-designed corpus will be better than a larger less well defined collection. Generally LSP corpora will be smaller than an LGP.</td>
</tr>
<tr>
<td>Text extracts or full texts</td>
<td>Full texts are preferred rather than sampling procedures, eg. random 2000 word selections. This is especially important since the IMRD structure of the text is a key rhetorical structure for analysis and spread of lexical items. In addition, from previous studies, it has been noted that specific sub-sections of texts are often lexically and grammatically specific.</td>
</tr>
<tr>
<td>Number of texts</td>
<td>If the intention is to have a good representation of text types and language from the relevant LSP then a good range of authors is better than multiple texts by a single author. In addition, subject or topic spread should be important also and this is important in medical research.</td>
</tr>
<tr>
<td>Medium</td>
<td>Spoken or written</td>
</tr>
<tr>
<td>Subject</td>
<td>Texts should be specific to the subject you are analysing, eg. biochemistry or scientific research articles</td>
</tr>
<tr>
<td>Text Type</td>
<td>Should be relevant to the purposes of the study and may, therefore, be more or less restricted. The AWL, for example contains a range of text types; this corpus does not.</td>
</tr>
<tr>
<td>Authorship</td>
<td>Texts should be gathered from acknowledged experts in the field. This criterion is met in the medical research corpus by only using original published peer reviewed research articles</td>
</tr>
<tr>
<td>Language</td>
<td>Monolingual or bilingual corpus</td>
</tr>
<tr>
<td>Publication date</td>
<td>Age range of texts depends on purposes. In this case texts, with few exceptions are kept within a narrow age range.</td>
</tr>
</tbody>
</table>

The total corpus currently stands at 104 texts. Only original research articles were included and other genres/text types within the journals, eg. editorials, short communications, etc. were excluded. This was to ensure consistency across texts and match the text genre – original research articles – that medical research students are expected to review; texts were included that matched the research domains and projects of senior undergraduate students I teach in a research intensive year at the University of Melbourne. To achieve balance texts from journals in specific domains, eg. preventative medicine, were equally sampled alongside bigger collections of generic publications, eg. BMJ. Some texts originally collected (147) were excluded to match the balanced number spread and because they duplicated topics or authorship.

Bibliographic details, including abstracts, of texts chosen were uploaded into Endnote 7.0™. This allowed for review of texts for inclusion and exclusion to achieve a reasonably balanced albeit open corpora. All texts were saved in two formats – Acrobat PDF™ and text format. Texts were saved in PDF format to conserve the integrity of the text formatting and images as more accurate portrayals of the texts as genre. Articles were also downloaded in html format and NoteTab Light™ Freeware Version 4.95 was used to strip html documents of coding, images. Texts included had to be accessible in electronic form through subscribed electronic journals available from the University of Melbourne. An attempt was also made to include texts of relatively similar size in terms of word numbers but this was not always possible. The table below outlines the general characteristics of the corpus.
Table 4: Medical corpus characteristics

<table>
<thead>
<tr>
<th>Journals</th>
<th># texts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generic research journals</strong></td>
<td></td>
</tr>
<tr>
<td>British Medical Journal (BMJ)</td>
<td>14</td>
</tr>
<tr>
<td>The Lancet</td>
<td>18</td>
</tr>
<tr>
<td>Journal of the American Medical Association (JAMA)</td>
<td>16</td>
</tr>
<tr>
<td>New England Journal of Medicine (NEJM)</td>
<td>6</td>
</tr>
<tr>
<td><strong>Domain-specific journals</strong></td>
<td></td>
</tr>
<tr>
<td>Diabetes Research and Clinical Practice (DRCP)</td>
<td>5</td>
</tr>
<tr>
<td>Blood Cells, Molecules, and Diseases (BCMD)</td>
<td>5</td>
</tr>
<tr>
<td>Cancer Detection and Prevention (CDP)</td>
<td>5</td>
</tr>
<tr>
<td>Trends in Microbiology (TM)</td>
<td>5</td>
</tr>
<tr>
<td>Obstetrics &amp; Gynaecology (OG)</td>
<td>5</td>
</tr>
<tr>
<td>Preventive Medicine (PM)</td>
<td>5</td>
</tr>
<tr>
<td>Social Science &amp; Medicine (SSM)</td>
<td>5</td>
</tr>
<tr>
<td>The Journal of Paediatrics (JP)</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Surgical Research (JSR)</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Emergency Medicine</td>
<td>5</td>
</tr>
<tr>
<td><strong>Total texts</strong></td>
<td>104</td>
</tr>
</tbody>
</table>

**Discussion**

Increasingly applied linguists are working in discipline specific areas such as law (Bruce, 2002), engineering (Lax, 2002), and architecture (Swales, Barks, Ostermann, & Simpson, 2001) where they must simultaneously teach language and academic conventions to NESB students and acquire sufficient knowledge of the disciplines to give informed advice. Engaging with academic disciplines requires sustained observation, discussion and analysis on the part of such teachers to develop a knowledge of the discipline specific discourse conventions. The construction of a corpora of medical research texts to inform my work with second language students involved in a research intensive year at University of Melbourne seemed sufficient motivation.

In addition to contributing to existing linguistic analyses of medical texts, the corpora will serve several other pedagogical functions. Full texts with images will be used with students cohorts to highlight conventions within the sections of the research text and also to compare published advice about lexical and rhetorical characteristics with actual use. An analysis of the texts as genres, including the function of visual elements and other ‘staging’ devices will complement concordance based analysis of lexical frequencies and collocations. Finally, the corpus remains, in principle, open to being supplemented and updated. Future studies will report on these developments.

**Reference List**


Scope and Limits of Corpus-Based Studies of the Medical Research Article for Teaching ESL Student Researchers: Discussion and Example


Scope and Limits of Corpus-Based Studies of the Medical Research Article for Teaching ESL Student Researchers: Discussion and Example


**Brief Author Details**

**Dr. Gavin Melles** currently works as a lecturer in the Faculty International Unit of Medicine, Dentistry and Health Sciences at the University of Melbourne. His specific teaching focus is with undergraduate (3rd year) and postgraduate research students. He has a Masters in Linguistics and a Doctorate in Education. His research and publications are in the area of discipline specific writing and teaching.