Applying Internal Medicine Corpus Analysis Findings to the Development of Pedagogical Materials

Simon FRASER
Walter DAVIES
Keiso TATSUKAWA
Institute for Foreign Language Research and Education
Hiroshima University

The analysis described in this paper forms part of a wider project being carried out in collaboration with the medical faculty of Hiroshima University to develop an EMP (English for Medical Purposes) course for third-year students. The course has evolved from an ongoing interplay between corpus analysis, word lists, and materials development. Corpora have been built in parallel with the creation of teaching materials, with insights and findings in both areas interacting with and informing each other. Materials have evolved on the basis of an increasing understanding of the medical field and students' needs, and a specialized word list (the Medical Word List) has been created to aid students in their learning of medical terms (see Fraser, Davies, and Tatsukawa, 2015, for an overview).

In background research interviews, senior members of the medical faculty placed a strong emphasis on anatomy for students at the early stages of their studies, leading to the construction of a corpus based on *Gray's Anatomy for Students* (2nd edition; 2009). This corpus has been used to identify the most frequently occurring anatomy terms, and as a resource for creating and checking teaching materials (Fraser, Davies, & Tatsukawa, 2014). The medical faculty also suggested that we focus on common diseases and symptoms, and another well-established reference was chosen for subsequent analysis: *Harrison's Principles of Internal Medicine* (18th edition; 2012).

As with the anatomy textbook analysis, we wanted to apply the corpus linguistic techniques and methodology successfully used in the analysis of medical texts (e.g., Fraser, 2013; Wang, Liang & Ge, 2008; Hsu, 2013) to a major medical textbook. Our aim in this article is two-fold: to create lists of the most useful words in internal medicine based on their frequency and range of occurrence, and to highlight ways in which the corpus findings are being used to inform the development of classroom materials that more accurately reflect the ways in which medical English is actually used.

CREATING AN INTERNAL MEDICINE CORPUS

As mentioned above, the textbook selected was *Harrison's Principles of Internal Medicine*, a well-established medical reference book, and a recommended text for our medical students. The print version of the 18th edition of the book is split into two volumes, totalling 4,012 pages and consisting of the following 18 parts (see Figure 1). Each main part is further divided into several sections.

Part 1:	General Considerations in Clinical Medicine
Part 2:	Cardinal Manifestations in Clinical Medicine
Part 3:	Genes, the Environment, and Disease
Part 4:	Regenerative Medicine
Part 5:	Aging
Part 6:	Nutrition and Weight Loss
Part 7:	Oncology and Hematology
Part 8:	Infectious Diseases
Part 9:	Terrorism and Clinical Medicine
Part 10:	Disorders of the Cardiovascular System
Part 11:	Disorders of the Respiratory System
Part 12:	Critical Care Medicine
Part 13:	Disorders of the Kidney and Urinary Tract
Part 14:	Disorders of the Gastrointestinal System
Part 15:	Immune-Mediated, Inflammatory, and Rheumatologic Disorders
Part 16:	Endocrinology and Metabolism
Part 17:	Neurologic Disorders
Part 18:	Poisoning, Drug Overdose, and Envenomation

FIGURE 1. The Main Sections of *Harrison's Principles of Internal Medicine*

The book was prepared for scanning using an electric paper cutter to remove the binding and separate the pages. Individual pages were then fed automatically into a scanner and saved in digital (PDF) format. *Adobe Acrobat* OCR recognition software was used to convert the PDF files into text format ready for editing and corpus analysis. Separate files were created for each of the main 18 sections of the book.

ANALYZING THE CORPUS

Creating Frequency Lists

AntConc 3.4.3m (Anthony, 2014) was used to produce lists of the most frequent words and word combinations, to determine the distribution (range) of these items across the different sections of the textbook, and to create concordances to provide contextual information. The corpus consists of 2,634,103 words in total (tokens), and 53,851 word types. However, these numbers include abbreviations, symbols, names, and proper nouns. Also, due to the sheer size of the corpus and the time-consuming nature of the editing process, some work is still necessary to remove misspellings and other artefacts created by the digitization process, which will have the effect of reducing the total number of word types.

The Most Important Medical Words (High Frequency and Range)

AntConc produced a list of all the words ranked according to their frequency of occurrence, together with their range across the different sections of the book. The unit of counting was the individual word form, as it was felt that the use of lemmas or word families could obscure the ways in which the words are used in the text. Table 1 lists the 100 most frequent words, excluding articles, prepositions, pronouns, and the most common conjunctions; all of these words are found in at least 16 of the 18 sections, and the vast majority appear in all sections of the book. Appendix 1 shows the most frequent 500 words, including function words, and again, most of these words occur in each section of the textbook.

TABLE 1. Top 100 Words in the Internal Medicine Corpus (Frequency/Range)

1.	PATIENTS	17102	18	51.	FUNCTION	2411	18
2.	DISEASE	10858	18	52.	TYPE	2405	18
3.	TREATMENT	5931	18	53.	BOTH	2389	18
4.	ASSOCIATED	5413	18	54.	OCCUR	2385	18
5.	CELLS	5204	18	55.	DISORDERS	2376	17
6.	THERAPY	5014	18	56.	LOW	2375	18
7.	RISK	4982	18	57.	THEY	2353	18
8.	CELL	4708	18	58.	INFECTIONS	2345	18
9.	CLINICAL	4671	18	59.	THEIR	2342	18
10.	INFECTION	3893	17	60.	E.G.	2307	18
11.	SYNDROME	3805	18	61.	INTO	2289	18
12.	COMMON	3757	18	62.	SPECIFIC	2215	18
13.	BLOOD	3721	18	63.	THOSE	2201	18
14.	MG	3675	17	64.	HEART	2199	18
15.	OFTEN	3629	18	65.	ET AL	2185	18
16.	ACUTE	3528	18	66.	FAILURE	2178	18
17.	CHRONIC	3480	18	67.	PROTEIN	2159	17
18.	DIAGNOSIS	3468	17	68.	FACTORS	2152	18
19.	INCREASED	3428	18	69.	AGENTS	2119	18
20.	NORMAL	3408	18	70.	BONE	2095	17
21.	HIGH	3365	18	71.	DOSE	2073	17
22.	SOME	3336	18	72.	DISEASES	2034	17
23.	SYMPTOMS	3168	17	73.	EFFECTS	2034	18
24.	CAUSE	3148	17	74.	PRIMARY	2019	18
25.	SEVERE	3106	18	75.	WOMEN	2003	16
26.	TABLE	3014	18	76.	WELL	2000	18
27.	LEVELS	2938	18	77.	DRUGS	1992	17
28.	RENAL	2908	17	78.	AGE	1988	18
29.	PAIN	2897	17	79.	STUDIES	1988	18
30.	USED	2895	18	80.	SKIN	1985	18
31.	INCLUDE	2882	18	81.	TUMOR	1984	16
32.	USE	2875	18	82.	RESULT	1964	18
33.	CANCER	2867	17	83.	IMPORTANT	1961	18
34.	ALL	2806	18	84.	TWO	1955	18
35.	BECAUSE	2767	18	85.	PULMONARY	1953	17
36.	THERE	2738	18	86.	FIG.	1937	18
37.	ONE	2727	18	87.	LOSS	1929	18
38.	HOWEVER	2718	18	88.	CARE	1924	18
39.	DRUG	2702	18	89.	GENE	1924	17
40.	CASES	2688	18	90.	SERUM	1914	17
41.	DUE	2675	18	91.	EARLY	1909	18
42.	NO	2623	18	92.	TISSUE	1906	18
43.	DURING	2571	18	93.	CAUSES	1905	17
44.	ALTHOUGH	2452	18	94.	SMALL	1902	18
45.	MANY	2449	18	95.	RESPONSE	1881	18
46.	YEARS	2446	18	96.	OCCURS	1876	18
47.	INCLUDING	2443	18	97.	LESIONS	1859	16
48.	PRESENT	2443	18	98.	FIRST	1847	18
49.	LIVER	2432	18	99.	RATE	1823	18
50.	ONLY	2412	18	100.	CARDIAC	1816	17

Categorizing the Words in the Lists

As we found with anatomy (Fraser et al., 2014), the most frequent words in the internal medicine corpus will at least be recognizable to the layperson; they are not the overtly, or fully, technical words of Greco-Latin origin that we might expect, with the possible exceptions of *pulmonary, serum, lesions*, and *renal*. However, as Table 2 shows, a large number of these familiar words take on a technical meaning when they combine with other words: examples are *heart failure, blood pressure, central nervous system*, and even *muscle weakness*.

Many words can be considered to be "lay-technical" (Fraser, 2012): those terms which are obviously technical, but whose basic meaning can, nonetheless, be understood by someone without specialist knowledge in the field. Words of this type include *disease, cells, heart, infection, syndrome, cancer, blood, diagnosis, symptoms, drug, skin, lung, bone,* and *dose*.

We also find "cryptotechnical" words (Fraser, 2012): everyday words which can be said to be "cryptic" in that they have a technical sense which is likely to be obscure to a non-specialist. Examples from the word list include *risk, failure, agent, effect, response, acute, chronic, care, tissue, severe, disorders, primary, type, function, rate,* and *pressure*. Cryptotechnical words are important because of their potential for confusion; many words only become "technicalized" when combined with other words to form multiword terms.

Multiword Terms

If multiword units occur with sufficiently high frequency in the corpus, then they should be treated in the same way as single-word terms. They function in the same way, though they may consist of two or more words (often noun-noun or adjective-noun combinations). In many cases they cannot be split, and have a meaning that is often not discernable even if the individual words making up the unit are known: *growth factor* and *gram negative*, for instance. However, we also find collocations which constitute a syntactic unit, but are not "fixed" (e.g., *liver transplantation* can be replaced by *transplantation of the liver*, and *blood flow* by *flow of blood*). Most combinations of this type are examples of nominalization, which allows the treatment of processes as nouns (see, e.g., Halliday, 1985). According to Pueyo and Val (1996), nominalizations are important in science because they enable complex phenomena to be treated in just a few words.

Table 2 shows the multiword technical terms which occur most frequently in the corpus. For all of these combinations, regardless of how they are formed, or how fixed they might be, the inseparability of the words making up the unit is apparent: it is the way in which these individual words interact that gives the terms their meaning.

TABLE 2. Top 100 Multiword Terms in the Internal Medicine Corpus (Frequency/Range)

1.	ET AL	(2185/18)	10.	SIDE EFFECTS	(519/16)
2.	BONE MARROW	(623/15)	11.	RENAL FAILURE	(507/15)
3.	BLOOD PRESSURE	(580/16)	12.	LIVER DISEASE	(487/13)
4.	RISK FACTORS	(578/15)	13.	VITAMIN D	(435/12)
5.	HEPATITIS B	(564/11)	14.	CLINICAL FEATURES	(427/16)
6.	T CELL	(563/15)	15.	HEALTH CARE	(426/18)
7.	HEART DISEASE	(547/16)	16.	NERVOUS SYSTEM	(405/17)
8.	HEART FAILURE	(547/14)	17.	DIABETES MELLITUS	(380/15)
9.	T CELLS	(535/12)	18.	PHYSICAL EXAMINATION	(357/17)

19.	AUTOSOMAL DOMINANT	(352/12)	60.	LUNG DISEASE	(186/12)
20.	BREAST CANCER	(352/10)	61.	BACK PAIN	(185/11)
21.	CLINICAL MANIFESTATIONS	(346/16)	62.	CLINICAL PRACTICE	(185/14)
22.	GROWTH FACTOR	(338/16)	63.	LYMPH NODE	(183/11)
23.	SPINAL CORD	(338/13)	64.	BETA BLOCKERS	(182/11)
24.	CORONARY ARTERY	(330/15)	65.	GASTROINTESTINAL TRACT	(182/17)
25.	MYOCARDIAL INFARCTION	(328/16)	66.	PULMONARY HYPERTENSION	(181/11)
26.	HEPATITIS C	(319/10)	67.	VITAMIN K	(181/11)
27.	BLOOD FLOW	(315/14)	68.	IMMUNE SYSTEM	(180/14)
28.	CLINICAL TRIALS	(315/17)	69.	ANTI INFLAMMATORY	(178/16)
29.	DIFFERENTIAL DIAGNOSIS	(315/14)	70.	ADVERSE EFFECTS	(177/13)
30.	ABDOMINAL PAIN	(308/14)	71.	CHEST PAIN	(174/13)
31.	MM HG	(306/12)	72.	CONNECTIVE TISSUE	(174/14)
32.	LUNG CANCER	(295/14)	73.	IN VITRO	(173/14)
33.	AUTOSOMAL RECESSIVE	(281/11)	74.	RISK FACTOR	(173/14)
34.	MORTALITY RATE	(276/15)	75.	INSULIN RESISTANCE	(170/12)
35.	LIFE THREATENING	(261/15)	76.	CONGESTIVE HEART	(169/14)
36.	GRAM NEGATIVE	(250/12)	77.	CLASS 1	(167/11)
37.	STEM CELL	(250/11)	78.	CT SCAN	(166/13)
38.	MORTALITY RATES	(248/17)	79.	CONGESTIVE HEART FAILURE	(165/14)
39.	RENAL DISEASE	(247/13)	80.	PERIPHERAL BLOOD	(161/14)
40.	LYMPH NODES	(242/13)	81.	SMALL BOWEL	(161/12)
41.	FAMILY HISTORY	(240/13)	82.	AMINO ACIDS	(158/15)
42.	RADIATION THERAPY	(238/11)	83.	HEART RATE	(158/10)
43.	HIV INFECTION	(231/13)	84.	INFECTIOUS DISEASES	(158/14)
44.	CENTRAL NERVOUS	(229/18)	85.	KIDNEY DISEASE	(158/10)
45.	CENTRAL NERVOUS SYSTEM	(226/17)	86.	PROSTATE CANCER	(157/11)
46.	RENAL FUNCTION	(224/13)	87.	TNF-A	(157/12)
47.	CARDIOVASCULAR DISEASE	(223/14)	88.	END STAGE	(155/13)
48.	SMOOTH MUSCLE	(219/15)	89.	ARTERY DISEASE	(154/12)
49.	CARDIAC OUTPUT	(215/12)	90.	LUPUS ERYTHEMATOSUS	(152/11)
50.	URINARY TRACT	(213/14)	91.	METASTATIC DISEASE	(150/11)
51.	TYPE DIABETES	(207/12)	92.	MUSCLE WEAKNESS	(148/11)
52.	EPITHELIAL CELLS	(202/12)	93.	PHYSICAL ACTIVITY	(147/12)
53.	TYPE 1	(202/13)	94.	GENE EXPRESSION	(146/13)
54.	DRUG INDUCED	(201/14)	95.	NECROSIS FACTOR	(146/13)
55.	ENDOTHELIAL CELLS	(201/14)	96.	CLINICAL PRESENTATION	(145/14)
56.	LIVER TRANSPLANTATION	(196/12)	97.	FIRST LINE	(145/13)
57.	RHEUMATOID ARTHRITIS	(196/11)	98.	RENAL INSUFFICIENCY	(145/14)
58.	AMINO ACID	(195/13)	99.	MEDICAL THERAPY	(144/13)
59.	PULMONARY EDEMA	(189/13)	100.	TUMOR NECROSIS	(144/13)

Correlating Noun Pairs

Among the multiword units shown in Appendix 2, we find an interesting category of terms that might be labelled "correlating noun pairs". Each unit consists of a pair of closely related medical terms that frequently occur together. They include *signs and symptoms, morbidity and mortality, head and neck, nausea and vomiting, manifestations and presentation*, and *diagnosis and treatment*. It is worth pointing out to learners that these words often occur together to form a new term. Sometimes, as with *signs and symptoms*, the words in the combination are so closely linked that their individual meanings have become almost indistinguishable.

Text-structuring Phrases

Many three- or four-word items are expressions that can be understood as having a role in structuring the discourse of the text. Table 3 shows the most frequently occurring discourse-structuring expressions. Most of these can be thought of in terms of Lewis' (1993) lexical chunks or Nattinger and DeCarrico's (1992) lexical phrases: prefabricated language units that can be used as wholes, rather than being composed through the use of syntactic rules.

Some of these phrases will be found in a wide variety of academic texts: *as well as, play a role*, and *on the basis of,* for instance. Other sequences are not perhaps so fixed, but the words they contain are so commonly used together in academic writing that they probably will be stored in the mind in chunks: examples are *is thought to* and *have been described*.

TABLE 3. Key Text-Structuring Collocations

as well as (the)	in some cases
in the presence of	on the basis of
in the absence of	the differential diagnosis (of)
(is) the most common	the most important
(is) associated with	has been reported
in addition to	the extent of
a number of	as a consequence (of)
a variety of	the most common cause (of)
as a result (of)	the fact that
is characterized by	factors such as
in the setting of	is thought to
in response to	have been described
based on (the)	play a role (in)
at the time (of)	(the) majority of patients
in contrast to	on the other hand

Another type of expression we are considering as belonging to this category has a rather different function. We see multiword items such as *in the presence of, in the absence of*, and *the differential diagnosis of*, which appear to be highly specific to the discipline. These are not really text-structuring expressions as such, but they are found frequently in descriptions of clinical procedure or when making observations in medicine. The importance of identifying this type of multiword item lies in the fact that although clearly belonging to the domain, it will not necessarily be found in existing medical dictionaries or glossaries.

Appendix 2 shows that the modal verbs *should, can,* and *may* have an important role to play in the discourse of medical texts. It is useful to know that these words occur frequently in the following expressions:

Should:
should be considered; should be performed; should be used; should be avoided
Can:
can be used; can lead to; can result in; can occur in
May:
may lead to; may be associated with; may result in; may be present; may result from; may be necessary; may contribute to; may present with; may be required; may be helpful

The Importance of Field Knowledge

Our findings have implications for teachers as well as learners. It has often been stated that it is not the job of the language teacher to teach technical language (e.g., Hutchison & Waters, 1987). However, other writers, including Bell (1996), raise the importance of subject-content knowledge in the teaching of English for Academic Purposes. In a similar vein, Davies, Fraser, and Tatsukawa (2016) posit that to teach medical English effectively, it is necessary for the instructor to have sufficient "field competence", or understanding of the medical field. The linguistic competence resulting from a knowledge of the different categories of words central to medical discourse, their characteristics, and their frequency and patterns of occurrence will go a long way toward achieving this.

APPLYING THE FINDINGS TO MATERIALS DEVELOPMENT

Comparison of Units of Material with the Relevant Sections of *Harrison's*

Each unit of the teaching materials has already been checked using the *Gray's Anatomy* corpus, leading us to rewrite and amend the materials to include important positional terms such as *inferior, superior, anterior,* and *posterior,* and to improve the discourse by using more appropriate terms and expressions (Fraser et al., 2014). In a similar way, the *Harrison's Principles of Internal Medicine* corpus can also be used as a resource in the creation and checking of materials. The textbook is divided into 18 sections, many of which are broadly equivalent to the key medical areas identified by the medical faculty. Each unit of the medical materials can be checked using the relevant section of *Harrison's*; for example, Part 10 (The Cardiovascular System) can be used to review the Heart unit, and check for usage of terms as well as important omissions. (See Davies et al., 2016, for examples of the materials.) Figure 2 shows the way in which the different units can be monitored:

Unit of materials	Section of *Harrison's*
The Skeletal System	N/A (check with *Gray's Anatomy*)
The Heart	Part 10: Cardiovascular System
The Brain	Part 17: Neurologic Disorders
The Digestive System	Part 14: Gastrointestinal Tract
The Endocrine System	Part 16: Endocrinology and Metabolism
The Pulmonary System	Part 11: Disorders of the Respiratory System
The Urinary System	Part 13: Disorders of the Kidney and Urinary Tract

FIGURE 2. Units of Materials and the Corresponding Sections of *Harrison's*

Work has begun on modifying the Heart and Digestive System units, and in the following sections we examine these units in some detail, making comparisons in terms of the frequently occurring lexical items found in the relevant sections of the corpus.

The Heart Unit

The words below occur frequently in Part 10 of *Harrison's* (i.e., with a frequency of 100 or more), and were all found in this unit of the materials:

heart, disease, pressure, risk, blood, pulmonary, associated, myocardial, artery, arterial, clinical, acute, severe, patient, symptoms, chronic, function, infarction, diagnosis, cardiovascular, angina, stenosis, venous, ventricle, aorta, atrium

It was encouraging to discover that these important words had been incorporated into the materials.

A check could also be made to determine whether the words we considered to be key terms, and have highlighted in the materials, are actually deserving of our attention. Each unit of the materials includes a "word box" containing the most basic anatomical terms. The words in the Heart unit are shown below, and, again reassuringly, all were found to occur frequently in the corpus, and in the Cardiovascular System section particularly.

right ventricle	left atrium	aorta	inferior vena cava	systemic circulation
tricuspid valve	pulmonary vein		pulmonary circulation	
superior vena cava	pulmonary artery		right atrium	
mitral valve	pulmonary valve	aortic valve	left ventricle	

Also, in each unit students are given a word-matching task to introduce the key vocabulary (Figure 3):

Vocabulary

Match the words in the box with their definitions.

consume	sternum	sweating	anxiety	obesity
germ	fatigue	edema	angina pectoris	
fainting	pounding	palpitation		

1. the production of liquid (mainly water) on the skin to cool the body

_____

2. the condition of being fat or overweight

_____

~~~
~~~

appearing in the Cardiovascular System section of the corpus but not the materials are *atrial, syndrome, ischemia, tachycardia, coagulation, murmur, calcium, arrythmia, atherosclerosis, atherosclerotic, cholesterol,* and *pulse*.

The high-frequency two-word combinations in this section of the corpus that occur in the materials are: *associated with, heart disease, blood pressure, myocardial infarction, risk factors, aortic valve, mitral valve, left ventricle,* and *right ventricle*. As with single-word items, however, a number of high-frequency combinations do not appear in the materials, including *(congestive) heart failure, coronary artery, pulmonary hypertension, beta blocker(s), blood flow, smooth muscle, SA (sinoatrial) node, cardiac output, atrial fibrillation, arterial pressure, contribute to,* and *myocardial ischemia*.

The following are particularly important words that are central to any description of the cardiovascular system, and their omission needs to be addressed in our revision of the materials: *cardiac, heart failure, coronary artery, atrial fibrillation, SA node, tachycardia, hypertension, smooth muscle, beta blocker, systolic,* and *diastolic*. Some terms do not appear in the Heart unit, but are of less concern as they are covered elsewhere in the materials: *therapy, treatment, dysfunction, stroke, flow,* and *diabetes mellitus*.

The Digestive System Unit

The following words occur frequently in the Gastrointestinal Tract section of the corpus, and are all found in the medical materials:

gastric, bowel, intestinal, symptoms, associated, pain, diarrhea, bile, colon, ulcer, diagnosis, abdominal, stool, intestine, severe, acid(s), absorption, cancer, pancreatic, esophagus, esophagitis, appendicitis, gut

A number of high frequency items, however, are not found:

esophogeal, therapy, treatment, obstruction, mucosal, CD, bleeding, chronic, risk, endoscopy, endoscopic, colonic, colitis, GI, disorders, inflammatory, UC, mucosa, syndrome, celiac, bacterial, secretion, IBD, gastrointestinal, anal, perforation, gastritis, sphincter

With regard to multiword units, only *small intestine* is found in the materials; the following terms are all absent: *abdominal pain, gastric acid, bowel disease, celiac disease, bile acids, small bowel, H. pylori, inflammatory bowel disease,* and *irritable bowel syndrome*.

Interestingly, whereas in the corpus we find *pancreatic head* and *pancreatic tail*, in the materials we have *pancreas (head)* and *pancreas (tail)*. Also, in this unit (and elsewhere in the materials), different forms of certain words are found; for instance, the adjective *inflammatory* does not appear in the materials, whereas the noun *inflammation* does. Findings such as these will be of great help in the fine-tuning of the materials.

The Pulmonary System and Urinary System Units

Space does not permit a detailed comparison of the treatment of the other systems in the materials with the appropriate sections of *Harrison's*, but it is worth noting the following key terms which are missing in each of the units:

Pulmonary System:

COPD (chronic obstructive pulmonary disease), airway(s), pulmonary fibrosis, syndrome, pressure, gas, interstitial, CF (cystic fibrosis), airflow obstruction, dyspnea, emphysema, transplantation, aspiration, bronchitis

Urinary System:

glomerular, GFR (glomerular filtration rate), dialysis, failure and hypertension, calcium, infection, interstitial, proximal tubule, distal tubule, potassium, reabsorption, channel, associated with, renal failure, kidney disease, renal disease, urinary tract, renal function, interstitial nephritis, collecting duct, blood pressure

Using the Corpus to Monitor the Materials: The Heart Unit

On the basis of the corpus findings, it has been possible to make the following amendments to the Heart unit. Most of these changes are minor, but we feel that the resulting text is a more accurate reflection of the ways in which medical conditions are actually described, and that it contains the most appropriate terminology to use in these descriptions. In Figure 4 below, we see that the text can be revised relatively easily to include a good proportion of the terms that have been identified as being important in cardiovascular medicine: *SA node, coronary artery, cardiac, flow, systole, diastole, hypertension, cholesterol, heart failure, stroke, atrial fibrillation, contribute to,* and *pulse*. The items *contributed to* and *nausea and vomiting* have also been included to reflect their high frequency in the corpus. This monitoring procedure is being repeated with the other units of teaching materials.

Original Text	**Revised Text**
Anatomy section, heart diagram No "SA node" in diagram No "coronary artery" in diagram	 "SA node" added to diagram "coronary artery" added to diagram
Circulatory System passage From it the blood *passes* into other arteries… The blood returns to the heart through venules and veins, entering the heart through the two venae cavae.	 From it the blood *flows* into other arteries… The blood returns to the heart through venules and veins, entering the heart through the two venae cavae. The contraction phase of the *cardiac* cycle is known as *systole*, and the relaxation phase as *diastole*.
Cardiovascular Medicine passage This was a major advance in the understanding of the function *of the heart* and the *movement* of the blood.	 This was a major advance in the understanding of *cardiac* function and the *flow* of the blood.

There are a number of ways in which the heart can suffer from problems, including the following: myocardial infarction, endocarditis, chronic mitral valve regurgitation, and aortic valve stenosis.	There are a number of ways in which the heart can suffer from problems, including the following: myocardial infarction, endocarditis, chronic mitral valve regurgitation, and aortic valve stenosis. Conditions such as *kidney disease* and *diabetes* can also contribute to *heart failure* or *stroke*.
…sweating, *nausea, vomiting*, abnormal heartbeats, and anxiety.	…sweating, *nausea and vomiting*, abnormal heartbeats *(atrial fibrillation)*, and anxiety.
There are a number of risk factors *associated with* myocardial infarction, the biggest being smoking, obesity, and lack of exercise.	There are a number of risk factors *contributing to* myocardial infarction, the biggest being smoking, obesity, lack of exercise, high blood *cholesterol*, and *hypertension*.
Doctor-patient conversation Doctor: I see. Have you noticed anything else?	Doctor: I see. Let me check your *pulse*. Have you noticed anything else?

FIGURE 4. Examples of Changes Made to the Heart Unit

CONCLUSION

This paper has investigated some of the ways in which a corpus analysis of *Harrison's Principles of Internal Medicine* is informing the development of teaching materials and word lists for use on a third-year medical English course at a Japanese university.

Several units of material have already been developed, along with a course-specific word list. These materials have been carefully constructed on the basis of feedback and advice from senior medical faculty and students, aided by the authors' experience in materials design. However, as teachers/applied linguists, our knowledge of the medical field is limited; in the initial creation of medical discourse, there will, inevitably, be important terms and text-structuring phrases which have been overlooked or used inappropriately. Corpus analysis is making it possible to check for such omissions and discrepancies, allowing us to increase the number of key terms embedded in each unit of the materials. We have also identified categories of regularly occurring terms that can help us to interlink the different units.

Of course, it is not possible, or even desirable, to incorporate all potentially useful terms in a relatively brief and necessarily concise unit of text. However, an effort should be made to include as many of the very high frequency words and multiword units as possible, and this will be addressed in subsequent revisions of the materials. Any terms with very high frequency and range which we are not able to write into the materials should be considered for inclusion in the Medical Word List.

REFERENCES

Anthony, L. (2014). *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University; available from<http://www.ant-lab.sci.waseda.ac.jp/>

Bell, T. (1996). Do EAP teachers and learners require knowledge of their students' specialist academic subjects? *The Internet TESL Journal.*

Davies, W., Fraser, S., & Tatsukawa, K. (2016). Analysing Medical English materials development at a university using a communicative competence model. *Hiroshima Studies in Language and Language Education*, 19, 93-108.

Drake, R., Vogl, W., & Mitchell, A. (2009) Gray's Anatomy for Students (2nd ed.). London: Churchill Livingstone.

Fraser, S. (2012). Factors affecting the learnability of technical vocabulary: Findings from a specialized corpus. *Hiroshima Studies in Language and Language Education*, 15, 123-142.

Fraser, S. (2013). Building corpora and compiling pedagogical lists for university medical students. *Hiroshima Studies in Language and Language Education*, 16, 65-88.

Fraser, S., Davies, W., & Tatsukawa, K. (2014). A corpus analysis of an anatomy textbook: Preliminary findings and implications for medical English development. *Hiroshima Studies in Language and Language Education*, 17, 179-194.

Fraser, S., Davies, W., & Tatsukawa, K. (2015). Creating a corpus-informed EMP course for medical English undergraduates. *Professional and Academic English*, 45, 16-20.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Hutchison, T. & Waters, A. (1987). *English for Specific Purposes*. Cambridge: Cambridge University Press.

Longo, D., Fauci, A., Kasper, D., Hauser, S., Jameson, J., & Loscalzo, J. (2012). *Harrison's Principles of Internal Medicine* (18th ed.). New York: McGraw-Hill Education.

Lewis, M. (1993). *The Lexical Approach: The State of ELT and a Way Forward.* Hove: Language Teaching Publications.

Nattinger, J. R. & DeCarrico, J. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Pueyo, I. G. & Val, S. (1996). The construction of technicality in the field of plastics: A functional approach towards teaching technical terminology. *English for Specific Purposes*, 15 (4), 251-278.

APPENDIX 1. The 500 most frequent lexical items (including Multiword Units; range >14/18)

1. THE
2. OF
3. AND
4. IN
5. TO
6. A
7. IS
8. WITH
9. OR
10. ARE
11. FOR
12. BE
13. PATIENTS
14. AS
15. BY
16. THAT
17. MAY
18. IN THE
19. DISEASE
20. AN
21. CAN
22. HAVE
23. FROM
24. ON
25. NOT
26. THIS
27. AT
28. THESE
29. SHOULD
30. HAS
31. MOST
32. TREATMENT
33. BUT
34. OTHER
35. ALSO
36. BEEN
37. IT
38. ASSOCIATED
39. MAY BE
40. CELLS
41. TO THE
42. SUCH
43. WHICH
44. MORE
45. THERAPY
46. RISK
47. CELL
48. CLINICAL
49. SHOULD BE
50. PATIENT
51. CAN BE
52. ASSOCIATED WITH
53. SUCH AS
54. THAN
55. INFECTION
56. USUALLY
57. IF
58. SYNDROME
59. AFTER
60. COMMON
61. BLOOD
62. MG
63. OFTEN
64. ACUTE
65. CHRONIC
66. WHEN
67. DIAGNOSIS
68. INCREASED
69. NORMAL
70. HIGH
71. SOME
72. SYMPTOMS
73. CAUSE
74. SEVERE
75. TABLE
76. LEVELS
77. RENAL
78. PAIN
79. USED
80. WHO
81. INCLUDE
82. USE
83. CANCER
84. ALL
85. BECAUSE
86. THERE
87. ONE
88. HOWEVER
89. DRUG
90. CASES
91. DUE
92. NO
93. DUE TO
94. DURING
95. IN PATIENTS WITH
96. TO BE
97. ALTHOUGH
98. MANY
99. YEARS
100. INCLUDING
101. PRESENT
102. LIVER
103. ONLY
104. FUNCTION
105. TYPE
106. BOTH
107. OCCUR
108. DISORDERS
109. LOW
110. THEY
111. INFECTIONS
112. THEIR
113. HAS BEEN
114. E.G.
115. INTO
116. SPECIFIC
117. THOSE
118. HEART
119. ET AL
120. FAILURE
121. PROTEIN
122. FACTORS
123. HAVE BEEN
124. AGENTS
125. BONE
126. DOSE
127. DISEASES
128. EFFECTS
129. PRIMARY
130. WOMEN
131. WELL
132. DRUGS
133. AGE
134. STUDIES
135. SKIN
136. TUMOR
137. RESULT
138. IMPORTANT
139. TWO
140. PULMONARY
141. RISK OF
142. FIG
143. LOSS
144. CARE
145. GENE
146. SERUM
147. EARLY
148. TISSUE
149. CAUSES
150. SMALL
151. RESPONSE
152. OCCURS
153. LESIONS
154. ITS
155. FIRST
156. RATE
157. CARDIAC
158. WILL
159. INCREASE

160. WITHIN
161. MUTATIONS
162. LUNG
163. PRESSURE
164. BETWEEN
165. LESS
166. HEPATITIS
167. ACID
168. INDIVIDUALS
169. FACTOR
170. IS NOT
171. RELATED
172. DEFICIENCY
173. OF THESE
174. WITHOUT
175. RESULTS
176. TIME
177. ACTIVITY
178. EFFECTIVE
179. RECEPTOR
180. THERE IS
181. TREATMENT OF
182. TUMORS
183. WAS
184. SEVERAL
185. MUSCLE
186. LARGE
187. EVIDENCE
188. OVER
189. DEVELOP
190. PRESENCE
191. FEATURES
192. SYSTEM
193. DEVELOPMENT
194. HISTORY
195. ABOUT
196. THROUGH
197. USE OF
198. AMONG
199. LIFE
200. LONG
201. MAJOR
202. LEVEL
203. ANY
204. PLASMA
205. PARTICULARLY
206. GENETIC
207. SEEN
208. HEALTH
209. DAYS
210. MANAGEMENT
211. BASED
212. CAUSED
213. HYPERTENSION
214. CHANGES
215. UP
216. FOUND
217. PRESENCE OF
218. SYSTEMIC
219. EVEN
220. FIGURE
221. INFLAMMATORY
222. LOWER
223. MULTIPLE
224. ONSET
225. FEVER
226. SEE
227. ELEVATED
228. CONTROL
229. DO
230. CAUSED BY
231. REDUCED
232. FORM
233. AS WELL
234. MEDICAL
235. CALCIUM
236. PER
237. EITHER
238. DIABETES
239. NUMBER
240. GROWTH
241. MEN
242. THE PRESENCE OF
243. COMMONLY
244. TREATED
245. VASCULAR
246. DIAGNOSIS OF
247. SECONDARY
248. SURGERY
249. RATES
250. DOSES
251. INJURY
252. ADDITION
253. RESPIRATORY
254. ABNORMALITIES
255. ESPECIALLY
256. HUMAN
257. NEW
258. CONSIDERED
259. INSULIN
260. MOST COMMON
261. WEEKS
262. POSITIVE
263. THUS
264. RARE
265. LEFT
266. EXPOSURE
267. DISORDER
268. BEFORE
269. GENES
270. MORTALITY
271. ORAL
272. TEST
273. DIAGNOSTIC
274. PRODUCTION
275. BODY
276. DEATH
277. LEAD
278. FLUID
279. WHILE
280. MONTHS
281. LIKELY
282. MANIFESTATIONS
283. FREQUENTLY
284. BECAUSE OF
285. HORMONE
286. EXAMINATION
287. AVAILABLE
288. OF PATIENTS WITH
289. AS WELL AS
290. EFFECT
291. BRAIN
292. CD
293. MUST
294. YEAR
295. GIVEN
296. NUMBER OF
297. IMMUNE
298. IS USUALLY
299. INCIDENCE
300. TYPICALLY
301. VITAMIN
302. DOES
303. SURVIVAL
304. IN ADDITION
305. GENERALLY
306. PERIPHERAL
307. TERM
308. RESISTANCE
309. CONDITIONS
310. PATIENTS WHO
311. TESTS
312. VENTRICULAR
313. GLUCOSE
314. FURTHER
315. CONTRAST
316. CORONARY
317. SIGNIFICANT
318. WERE
319. SIGNS
320. SIMILAR
321. ANEMIA
322. ADMINISTRATION
323. ADULTS
324. INDUCED

325. INITIAL
326. SURGICAL
327. COMPLICATIONS
328. PROTEINS
329. CHILDREN
330. SHOWN
331. KG
332. NEGATIVE
333. HIGHER
334. TRANSPLANTATION
335. BACTERIAL
336. ANTI
337. WEIGHT
338. PHYSICAL
339. PRODUCE
340. THREE
341. THE MOST COMMON
342. DEVELOPMENT OF
343. DYSFUNCTION
344. DO NOT
345. ROLE
346. EVALUATION
347. VERY
348. BIOPSY
349. VOLUME
350. TESTING
351. DECREASED
352. MARROW
353. USEFUL
354. DIFFERENT
355. ARTERY
356. ABSENCE
357. ACTIVATION
358. EACH
359. LEAD TO
360. CT
361. EXAMPLE
362. OBSTRUCTION
363. FAMILY
364. INVOLVEMENT
365. BELOW
366. MILD
367. POSSIBLE
368. DNA
369. THYROID
370. CHEST
371. KNOWN
372. ANTIBODIES
373. IL
374. IRON
375. RIGHT
376. NERVE
377. FINDINGS
378. SINGLE
379. STAGE
380. FOLLOWING
381. DAILY
382. INCREASES
383. DAMAGE
384. REQUIRE
385. AGAINST
386. CHARACTERIZED
387. POPULATION
388. BLEEDING
389. RADIATION
390. CAUSE OF
391. TRACT
392. IN SOME
393. LEVELS OF
394. RARELY
395. REPORTED
396. LOSS OF
397. RECOMMENDED
398. HEPATIC
399. ACTIVE
400. COMPLEX
401. INFLAMMATION
402. ABSENCE OF
403. UNDERLYING
404. KIDNEY
405. REDUCE
406. REQUIRED
407. POTENTIAL
408. SIDE
409. SYNDROMES
410. METABOLIC
411. INHIBITORS
412. PROCESS
413. IS ASSOCIATED WITH
414. COURSE
415. DAY
416. ABDOMINAL
417. AREAS
418. CENTRAL
419. TRIALS
420. GROUP
421. THE RISK OF
422. COMBINATION
423. WATER
424. FORMS
425. SINCE
426. UP TO
427. INCREASE IN
428. REDUCTION
429. INVOLVED
430. AGENT
431. AND ARE
432. IDENTIFIED
433. RECEPTORS
434. STUDY
435. PREGNANCY
436. PREVENTION
437. DIARRHEA
438. BENEFIT
439. FLOW
440. ARTHRITIS
441. CAUSES OF
442. LIMITED
443. CHEMOTHERAPY
444. RESULTING
445. LEAST
446. PROGRESSIVE
447. SITE
448. ML
449. CHARACTERIZED BY
450. OCCURS IN
451. PROVIDE
452. BASED ON
453. RAPID
454. URINE
455. HIV
456. ILLNESS
457. USING
458. MASS
459. APPROACH
460. PERFORMED
461. SECOND
462. SCREENING
463. INTESTINAL
464. BECOME
465. URINARY
466. MYOCARDIAL
467. CERTAIN
468. MRI
469. RED
470. ABNORMAL
471. ASSOCIATION
472. CARDIOVASCULAR
473. WEAKNESS
474. AFFECTED
475. USED TO
476. RESPONSE TO
477. THE USE
478. THEN
479. VIRAL
480. MEMBRANE
481. NEUROLOGIC
482. APPROPRIATE
483. HISTORY OF
484. DATA
485. HOST
486. OCCUR IN
487. COUNTRIES
488. MAY ALSO
489. THE USE OF

490. VIRUS
491. MECHANISMS
492. FOR EXAMPLE
493. BEING
494. BINDING
495. BREAST
496. LIKE
497. PART
498. RECURRENT
499. MUTATIONS IN
500. MANAGEMENT OF

APPENDIX 2. Top 300 Text Structuring Multiword Units in the Internal Medicine Corpus (Range>12)

1. IN PATIENTS WITH
2. THE PRESENCE OF
3. OF PATIENTS WITH
4. AS WELL AS
5. THE MOST COMMON
6. THE RISK OF
7. THE USE OF
8. THE DEVELOPMENT OF
9. THE ABSENCE OF
10. THE TREATMENT OF
11. THE DIAGNOSIS OF
12. IN THE ABSENCE OF
13. ASSOCIATED WITH A
14. IN ADDITION TO
15. FOR PATIENTS WITH
16. BECAUSE OF THE
17. IS THE MOST
18. A NUMBER OF
19. A VARIETY OF
20. ARE ASSOCIATED WITH
21. SHOULD BE CONSIDERED
22. BE ASSOCIATED WITH
23. AS A RESULT
24. IS CHARACTERIZED BY
25. THE INCIDENCE OF
26. AN INCREASE IN
27. INCREASED RISK OF
28. MUTATIONS IN THE
29. A RESULT OF
30. CAN BE USED
31. MAY ALSO BE
32. BEEN SHOWN TO
33. ONE OF THE
34. IN A PATIENT
35. APPEARS TO BE
36. SUCH AS THE
37. BE USED TO
38. IN PATIENTS WHO
39. OF THE DISEASE
40. CAN LEAD TO
41. AS A RESULT OF
42. A HISTORY OF
43. BEEN ASSOCIATED WITH
44. DUE TO THE
45. IS CAUSED BY
46. PATIENTS WHO HAVE
47. THE MANAGEMENT OF
48. THE MAJORITY OF
49. HAS NOT BEEN
50. PATIENTS WITH A
51. IN THE SETTING
52. THE SETTING OF
53. OF THE PATIENT
54. IN THE SETTING OF
55. WITH OR WITHOUT
56. IS THE MOST COMMON
57. THE TIME OF
58. FOR THE TREATMENT
59. THE NUMBER OF
60. THE ONSET OF
61. AND MAY BE
62. FOR THE TREATMENT OF
63. IN RESPONSE TO
64. BASED ON THE
65. IN SOME PATIENTS
66. REFERRED TO AS
67. PATIENTS WITH CHRONIC
68. THE RATE OF
69. IN THE PRESENCE
70. THE PATIENT IS
71. MAY LEAD TO
72. AT THE TIME
73. IN CONTRAST TO
74. IT IS IMPORTANT
75. THE COURSE OF
76. IS DUE TO
77. SHOULD NOT BE
78. IN THE PRESENCE OF
79. THE SEVERITY OF
80. IN ASSOCIATION WITH
81. MAY BE ASSOCIATED
82. IN THE TREATMENT
83. IS IMPORTANT TO
84. RISK FACTORS FOR
85. MORE COMMON IN
86. MAY BE ASSOCIATED WITH
87. AT THE TIME OF
88. IT IS IMPORTANT TO
89. CAN ALSO BE
90. IN SOME CASES
91. HAVE BEEN REPORTED
92. MORE LIKELY TO

93. PATIENTS WITH SEVERE
94. IN THE TREATMENT OF
95. ASSOCIATED WITH THE
96. PART OF THE
97. SHOULD BE PERFORMED
98. THE DEGREE OF
99. APPEAR TO BE
100. IF THE PATIENT
101. MAY RESULT IN
102. BE CONSIDERED IN
103. IT IS NOT
104. THE PREVALENCE OF
105. THE PATHOGENESIS OF
106. IN A PATIENT WITH
107. THE LEVEL OF
108. THE ROLE OF
109. COMMON CAUSE OF
110. IT MAY BE
111. THE BASIS OF
112. AND SHOULD BE
113. ASSOCIATED WITH AN
114. ONE THIRD OF
115. PATIENTS WHO ARE
116. DEPENDING ON THE
117. THE DIAGNOSIS IS
118. YEARS OF AGE
119. ARE THE MOST
120. ON THE BASIS
121. HAS BEEN SHOWN
122. IN COMBINATION WITH
123. ON THE BASIS OF
124. THE NEED FOR
125. AND CAN BE
126. THE DIFFERENTIAL DIAGNOSIS
127. MAY NOT BE
128. A COMBINATION OF
129. A CONSEQUENCE OF
130. BE TREATED WITH
131. LIKELY TO BE
132. SHOULD BE USED
133. THE SITE OF
134. BE DUE TO
135. THE POSSIBILITY OF
136. RELATED TO THE
137. THE CAUSE OF
138. IS ASSOCIATED WITH A
139. THAT CAN BE
140. MAY BE A
141. A REDUCTION IN
142. SIGNS AND SYMPTOMS
143. CAN BE USED TO
144. SHOULD BE TREATED
145. THE RESULT OF
146. AN INCREASED RISK
147. IN THE FIRST
148. PATIENTS SHOULD BE
149. IN WHICH THE
150. HAVE BEEN IDENTIFIED
151. AND TREATMENT OF
152. SOME PATIENTS WITH
153. THE FREQUENCY OF
154. PATIENTS WITH ACUTE
155. OCCURS IN OF
156. MAY BE PRESENT
157. THE MOST IMPORTANT
158. FOR TREATMENT OF
159. HAVE NOT BEEN
160. HAS BEEN REPORTED
161. IS BASED ON
162. DUE TO A
163. FOUND IN THE
164. BE CAUSED BY
165. CELLS IN THE
166. MORBIDITY AND MORTALITY
167. A ROLE IN
168. IN OF PATIENTS WITH
169. THE LIKELIHOOD OF
170. HAS BEEN SHOWN TO
171. IT HAS BEEN
172. MOST PATIENTS WITH
173. THE EXTENT OF
174. AS A CONSEQUENCE
175. KG PER DAY
176. IN CASES OF
177. OF THESE PATIENTS
178. DEPENDS ON THE
179. OCCUR IN THE
180. AN INCREASED RISK OF
181. AT RISK FOR
182. CONTRIBUTE TO THE
183. MAY BE THE
184. MOST COMMON CAUSE
185. SHOULD BE GIVEN
186. INCREASE IN THE
187. SHOULD BE AVOIDED
188. THE DURATION OF
189. CHANGES IN THE
190. HAS ALSO BEEN
191. IN THESE PATIENTS
192. THAT MAY BE
193. IS MORE COMMON
194. PRESENCE OF A
195. ROLE IN THE
196. ALL PATIENTS WITH
197. MOST OF THE
198. THE EFFECTS OF
199. CAN RESULT IN
200. THE DISEASE IS
201. OF THE MOST
202. ONE HALF OF

203. THE FACT THAT
204. TO REDUCE THE
205. UP TO OF PATIENTS
206. IS AN IMPORTANT
207. AS WELL AS THE
208. FACTORS SUCH AS
209. HAS BEEN USED
210. MAY OCCUR IN
211. WELL AS THE
212. A DECREASE IN
213. IS THOUGHT TO
214. THE PATIENT HAS
215. THOUGHT TO BE
216. IS A COMMON
217. THE EVALUATION OF
218. THE MOST COMMON CAUSE
219. IN MOST PATIENTS
220. IN THE COURSE
221. MOST COMMON CAUSE OF
222. ONE OR MORE
223. THE PATIENT WITH
224. BE SEEN IN
225. HAS BEEN ASSOCIATED
226. THE ABILITY TO
227. THE FORMATION OF
228. THE RELEASE OF
229. THE PRESENCE OF A
230. A DIAGNOSIS OF
231. ACTIVATION OF THE
232. IN THE LIVER
233. THE PATIENT AND
234. DIFFERENTIAL DIAGNOSIS OF
235. FAMILY HISTORY OF
236. IN INDIVIDUALS WITH
237. IT CAN BE
238. SHOULD BE CONSIDERED IN
239. HAS BEEN ASSOCIATED WITH
240. OF THE SKIN
241. THE PRODUCTION OF
242. IN THIS SETTING
243. IS PRESENT IN
244. AS A CONSEQUENCE OF
245. FOR THE DIAGNOSIS
246. HAVE BEEN DESCRIBED
247. MAY RESULT FROM
248. AGENTS SUCH AS
249. THE CASE OF
250. CAN BE SEEN
251. QUALITY OF LIFE
252. RESPONSIBLE FOR THE
253. SIMILAR TO THOSE
254. FOR AT LEAST
255. MAY BE USED
256. WITH AN INCREASED
257. CAN OCCUR IN
258. PRESENT IN THE
259. THE ADMINISTRATION OF
260. THE AGE OF
261. THE GENERAL POPULATION
262. ARE FOUND IN
263. DO NOT HAVE
264. HAVE ALSO BEEN
265. SOME OF THE
266. ARE MORE LIKELY
267. ASSOCIATED WITH INCREASED
268. IN MOST CASES
269. IN THE CASE
270. MANY OF THE
271. MAY BE NECESSARY
272. THE AMOUNT OF
273. IN THE CASE OF
274. MAY PRESENT WITH
275. NEED TO BE
276. WHICH MAY BE
277. OF CASES OF
278. THE IMPORTANCE OF
279. CAUSED BY A
280. DURING THE FIRST
281. INVOLVED IN THE
282. MAY CONTRIBUTE TO
283. DISEASES SUCH AS
284. HAVE SHOWN THAT
285. INVOLVEMENT OF THE
286. NAUSEA AND VOMITING
287. PLAY A ROLE
288. THE GASTROINTESTINAL TRACT
289. THERE MAY BE
290. MAY BE REQUIRED
291. OF PATIENTS AND
292. OF THE WORLD
293. AT HIGH RISK
294. BE CONSIDERED FOR
295. EARLY IN THE
296. IN PATIENTS WITH CHRONIC
297. INCREASED RISK FOR
298. MAJORITY OF PATIENTS
299. OF PATIENTS WHO
300. OTHER CAUSE

ABSTRACT

Applying Internal Medicine Corpus Analysis Findings to the Development of Pedagogical Materials

Simon FRASER
Walter DAVIES
Keiso TATSUKAWA
Institute for Foreign Language Research and Education
Hiroshima University

This article investigates some of the ways in which a corpus analysis of *Harrison's Principles of Internal Medicine* is informing the development of pedagogical materials for use on an EMP (English for Medical Purposes) course at Hiroshima University. The course, designed for third-year students, has evolved from an ongoing interaction between corpus analysis, word lists, and materials development. Corpora and teaching materials have been built in parallel, with materials evolving on the basis of an increasing understanding of the medical field and students' needs. A specialized word list (the Medical Word List) has been compiled to aid students in their learning of medical terms (Fraser, Davies, & Tatsukawa, 2015).

In background research interviews, senior members of the medical faculty placed a strong emphasis on anatomy for students at the early stages of their studies, leading to the construction of a corpus based on *Gray's Anatomy for Students* (2nd edition; 2009). This corpus has successfully been used to identify the most frequently occurring anatomy terms, and as a resource for creating and checking teaching materials (Fraser, Davies, & Tatsukawa, 2014). The medical faculty also suggested that we focus on common diseases and symptoms, and another well-established reference was chosen for subsequent analysis: *Harrison's Principles of Internal Medicine* (18th edition; 2012).

In this paper, we document the creation of lists of the most useful terms and expressions in internal medicine based on their frequency of occurrence and range across the different sections of the textbook. We also examine the characteristics of these items and the particular contexts in which they occur. Our findings have implications for EMP materials development, and examples are given here of how the corpus is being used to monitor and amend each unit of medical materials. In the initial creation of medical discourse, important terms and text-structuring phrases will inevitably be overlooked or used inappropriately by materials writers lacking expert knowledge of the field; corpus analysis is making it possible to check for such omissions and discrepancies, allowing us to improve accuracy and increase the number of key terms embedded in each unit of the materials.

要　約

オリジナル医学英語コーパスの分析知見の教材開発への援用

サイモン・フレイザー
デービス・ウォルター
達　川　奎　三
広島大学外国語教育研究センター

本論文の目的は，*Harrison's Principles of Internal Medicine* (18th edition; 2012) という書籍のコーパス分析が，広島大学における医学生を対象とした専門英語教育で用いる教材作成に，いかに結びついたかを検討・報告することである。医学部 3 年生を対象とした特別授業は，現在も進行中であるコーパス分析・語彙リスト作成と教材開発を行きつ戻りつしながら, 展開されている。コーパス作成と教材開発は，著者陣の医学分野に関する理解と学生のニーズ把握の深まりと相まって行われた。そして，医学生の学習支援を目指して，精選された医学英単語リストが，構築された (Fraser, Davies, & Tatsukawa, 2015)。

プロジェクトを進めるに際しての医学部教授陣への聞き取りでは，学部の早い段階での「解剖学」の勉強の重要性が強調された。それ故，*Gray's Anatomy for Students* (2nd edition; 2009) を用いてコーパスを構築した。このコーパスは最も一般的に用いられる解剖学用語の特定に有益であった (Fraser, Davies, & Tatsukawa, 2014)。さらに，医学部側からは一般疾病やその症状にも重点を置くべきであるとの指摘があり，さらなる分析のために評価の高い *Harrison's Principles of Internal Medicine* を用いることとした。

本論では，出現度数や教材での汎用範囲に基づいて，内科に関する最も有益な術語や表現リストを記述する。また，これらの語彙・表現の特徴や出現しやすい文脈などを検討する。考察を通して医学英語教育の教材開発における示唆が得られ，構築したコーパスがどのように教材の各ユニットにおいて用いられ修正されたかを例示する。初期の医学的談話（会話や文章）の作成（執筆）においては，医学の専門的知識が十分でないと，重要な術語や定型表現などを見逃したり不適切な使い方をしたりする。それ故，コーパス分析をすることによって，このような見逃しや不一致を確認し，使用の正確性を向上させ，教材の各ユニットに組み込むべき重要語句の数を増やすことが可能となる。