

平成 21 年 5 月 21 日現在

研究種目：基盤研究(B)
 研究期間：2005～2008
 課題番号：17320087
 研究課題名（和文） コーパス分析に基づく医学・看護英語リーディング・ライティング教育システムの構築
 研究課題名（英文） Developing a Corpus-based E-learning System for Teaching Reading and Writing Medical/Nursing English
 研究代表者
 横山 彰三 (YOKOYAMA SHOZO)
 宮崎大学・医学部社会医学講座英語分野・准教授
 研究者番号：60347052

研究成果の概要：本研究ではゲノム生命医学、臨床外科学、看護学、公衆衛生学の異なった4分野間の言語的差異を大規模コーパスデータの分析を通して明らかにした。それを基に医学・看護英語リーディング・ライティング用コースウェア、および検索用コンコーダンサーシステムを試作した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2005年度	2,500,000	0	2,500,000
2006年度	1,800,000	0	1,800,000
2007年度	1,100,000	330,000	1,430,000
2008年度	1,300,000	390,000	1,690,000
総計	6,700,000	720,000	7,420,000

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：ESP、コーパス言語学、医学英語、看護英語、リーディング、ライティング

1. 研究開始当初の背景

我々の目標は、ESP（職業・目的別英語教授法）理論およびコーパス言語学(corpus linguistics)に基づいて、医学・看護分野の英語ライティングとリーディング教育システムを構築することである。平成15年度～16年度科学研究費補助金(C)「ESP教授法に基づく大学専門英語教育のための効果的シラバスと教材開発の研究」(研究代表：横山彰三)において、医学・看護・工学分野のコーパス分析を実施した結果、専門教育でもっとも必要とされる英語力は「リーディング」と「ライティング」であった。さらに「専門教育を実施する上で学生の英語力に問題はあるか」との質問に対して、専門教員の7割以上が問題有りとしており、早急な手だてが必

要である。本研究では対象をEMP(English for Medical Purposes)に絞り、関連するコーパスデータの分析に基づいた英語論文の読解・作成教育システムの構築をめざした。

2. 研究の目的

日本におけるESP教育は以前に比べかなり改善され実践されてきてはいるが、いまだに大学における日々の授業や教材作成の工夫の域からなかなか抜け出せない。特に医学英語教育で英語教師ができることといえば、基本的な専門語彙を暗記させることや医学を題材にした一般的読み物やニュースの聴取り練習をさせることなどである。本研究の目的はより高度でかつ専門教育で避けては通れない医学論文の基本的な読みこなし方や

書き方を、英語教師であっても教育できる、そしてなにより学習者が自律的に学ぶことのできるシステムを日本の医学英語教育の現場に提供することである。

3. 研究の方法

コーパス言語学はますます学際的(interdisciplinary)になってきている(詳しくは Williams, 1996; Vihla, 1999; Stotesbury, 2003; Corte, 2004; Charles, 2006; Hyland, 2007 を参照)。また、科学研究論文(アカデミックライティング)に関するレトリック研究では IMRD (Introduction, Methods, Results and Discussion)とよばれる異なった修辭的段落に現れる談話機能に研究の焦点が当てられている(Vihla, 1999; Gledhill, 2000)。医学英語論文についてもコーパスデータに基づいた談話分析に関する研究が比較的多くなされているものの、医学といっても分野は細分化されており本研究では特に、ゲノム生命医学(genome biomedicine)、臨床外科学(clinical surgery)、看護学(nursing)、公衆衛生学(public health)のような異なった分野間の言語学的差異を明らかにすることを目的としている。この研究の目的のために収集および編集した4つの分野からなるコーパス(Medcorpus)を作成した。このMedcorpusはオンラインジャーナルPubMed Central (<http://www.pubmedcentral.nih.gov/>)から個別に論文を収集しそれをテキストファイルにした。すべての論文は、手作業で修辭段落ごとに分けられてテキスト情報(annotation)、段落タグ(section tag)が付加された。分野と総語数(token)は以下の通りである:ゲノム生命医学 6,910,765、臨床外科 2,786,557、看護学 1,041,480、公衆衛生学 848,095。

これらのデータを基に、以下の点について統計的な分析を行った。

- (1) CoxheadのAcademic Word List (AWL)を用いた語彙の重なり度(AWL Coverage)をRangeソフトウェアにより計測
- (2) 各分野・修辭段落ごとの動詞頻度(粗頻度)をWordsmithソフトウェアにより計測
- (3) 上記(2)で得られた上位語に関する連語のMI scoreをWordsmithソフトウェアにより計測
- (4) 各分野・修辭段落ごとの特徴語分析をWordsmithソフトウェアのloglikelihoodにより計測
- (5) 各分野・修辭段落ごとの助動詞(modal)の頻度をWordsmithソフトウェアにより計測。さらにその連語をMI scoreで計測
- (6) 各分野・修辭段落ごとの談話構成語を多語表現(lexical chunk)の視点から定量的に計測。さらにその頻度上位300語を対象

に対応分析により修辭段落間の相関を計測

4. 研究成果

上述の分析結果(の一部)は以下の通りである。

- (1) AWLのカバー率:ゲノム生命医学 10.11%、臨床外科 9.29%、看護学 10.28%、公衆衛生学 10.78%。AWLのカバー率はいかなる専門分野においても10%以上であるということなので、Medcorpus全体で11.6%という数字は比較的高いカバー率を示している。一方医学分野における分野別のカバー率も、Surgeryで若干低いとはいえほぼ10%は達成しており、AWLの有効性が医学分野でも証明されたことになる。(詳しくは、下記[雑誌論文]の⑥を参照のこと)
- (2) 各分野の動詞粗頻度は、be, haveを筆頭に、use, show, indicate, observe, suggest, find, obtainなどが共通語彙として見いだされた。
- (3) MIスコアにより得られた連語については、比較的特殊な結びつきを示す語彙も多く見られた。
- (4) 「キーワード」のコンセプトはあるコーパスの単語リストをより無色で一般的な語彙リストと比較した際に、統計的に特に優位に頻度が多い単語のことをいい、一般的なテキストよりもある特定の分野や内容のテキストであればそのコーパスに特有の語彙が高頻度に出てくるわけで、これらがそのテキストを特徴づける単語としてキーワードと呼ばれる。Genome分野の特徴語のほとんどは一般名詞であり、cell「細胞」、gene「遺伝子」、expression「発現」、protein「タンパク質」、sequence「配列」など遺伝子に関する語彙、あるいはanalysis, results, methods, studyなど実験・研究手順に関する語彙で占められている。名詞以外では前置詞のof、be動詞の過去形(were)とともに動詞の過去分詞形(shown, identified)が確認される。これはGenome分野に存在するジャンル特性が、遺伝子解析など実験に基づく客観的データをどう解釈するかという点にあることが強くうかがえる。Surgeryコーパスも同様に一般名詞が多い。外科系の身体に関する語彙inflammatory「炎症性の」、pulmonary「肺の」、epithelial「上皮の」などに加えてcell, expression, geneなど遺伝子関連の語彙も確認される。名詞以外では前置詞のin, of、be動詞の過去形(were)とともに動詞の過去分詞形(induced, increased, associated, shown)が確認される。特に「(薬によって)誘発された」という意味で多用されるinducedは他の3分野には見られない語彙である。Nursing分野の特徴語もほとんどが一般名詞である。careがもつ

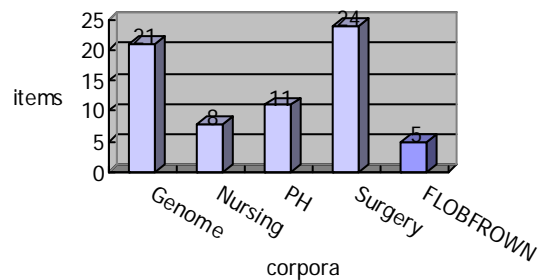
とも keyness の高い名詞としてランクされ、patient, nurse, physician, participant などの有生(animate)名詞が特徴的であり、看護(care)を通した対人的な研究が特徴的であることがうかがえる。一方、名詞以外では前置詞の of, for, be 動詞の過去形(were)とともに動詞の過去分詞形(reported, based, associated, related)が確認される。一般的に看護学の研究は質的研究が主であるといわれ、最近では、数値統計的データというよりインタビューなどによる聞き取り調査結果に基づく研究が主流である。そのような背景が動詞 reported によって現われているものと思われる。公衆衛生分野も同様に一般名詞が顕著である。health, data, population, factor, geographic, age, area などアンケート調査や統計などを想起させる語彙が多くみられる。名詞以外では前置詞の of, for, in や be 動詞の過去形(were)とともに動詞の過去分詞形(reported, based, associated)が確認され、やや Nursing コーパスの特徴語と相似した部分が確認される。この点、Nursing と PH のジャンル特性が現実世界の社会現象と関連している内容が多いことに起因するためと推測される。

(5) 各分野とも助動詞に関しては、may, can, could が上位を占め、いわゆる認識様態的助動詞(epistemic modal)が顕著である。修辞段落の Introduction は主題に関する背景知識を与え、論文の学術的価値を披瀝する。Methods と Results は実験の方法や得られたデータや研究結果が提示される一方、Discussion では Results を受けた考察・解釈がなされる。この著者の解釈部分ではしばしば may, might, could などの法助動詞の出現率が最も高いとされているが、これらは Discussion に最も頻出し、ついで Introduction 部分に出現することが明らかになった。

(6) 4 語バンドルの総頻度データを集計した結果、20 回以上出現している evaluative な意味を有する語彙バンドルの種類は FLOB/FROWN コーパスに比べて *Medcorpus* の方がかなり多いことが明らかとなった(頻度上位 100 項目のうち Genome Bio-Science で 21 種類、Nursing で 8 種類、Public Health で 11 種類、Clinical Surgery で 24 種類に対して FLOB/FROWN では 5 種類のみ: 以上グラフ 1 を参照)。さらに *Medcorpus* の 4 つの下位分野間の語彙バンドルについて比較すると、分野ごとの位相を垣間見ることが出来る。Nursing と PH コーパスでは evaluative な語彙バンドルの頻度は比較的少なく、その代わりに「研究テーマに関する(topic-oriented)」語彙バンドルや「文章の展開に関連する(text-oriented)」語彙

バンドルによって占められている。その一方で、Genome と Surgery コーパスにおける evaluative な語彙バンドルの種類は Nursing や PH のほぼ倍以上、FLOB/FROWN と比べると 4 倍から 5 倍の種類が上位 100 項目以内に出現し、その他の語彙バンドルでは「実験手順に関する(procedure-oriented)」語彙バンドルや「データ記述に関する(description-oriented)」語彙バンドルが顕著である。

グラフ1. Items of evaluative bundles in Top 100



さらに興味深いことは Genome コーパスと Surgery コーパス間、Nursing コーパスと Public Health コーパス間には、それぞれ共通する語彙バンドル項目あるいは強い共通の傾向が確認できる点であろう。これら 2 つのグループにおける出現語彙バンドルにみられる強い類似性は、それらが属する学問分野に存在するジャンル特性(genre-specificity)によるものである可能性が高いといえる。Genome Bio-science と Clinical Surgery は両方とも実験に基づいた(experiment-based)、いわばハードサイエンスに近い研究分野であり人体または人体の一部の機能を分析し、ミクロあるいはマクロレベルでの室内実験(lab experiment)を行うことが多い。一方、Nursing と Public Health は基本的に患者/対照あるいは患者群/対照群の人間を集団とし非侵襲的な手法—アンケート調査や統計—で扱う。そのため Nursing と Public Health 分野の研究論文は、国や団体の経済状況や教育水準あるいは社会文化的な問題といった現実世界の社会現象と関連している内容が多い。Nursing と PH 論文が有するこのジャンル特性は、実験室で発見された科学的事実の積み重ねとは違って、論旨は必然的に推論的論理展開へと導かれる傾向にある。この点、社会科学系の論文にみられる傾向と類似しており、Hyland (2007) は社会科学(social science)の論文におけるレトリック特徴を “knowledge is typically constructed as plausible reasoning rather than as nature speaking directly through experimental findings”

だと説明している。以上の点は、図1、図2の対応分析による4分野間の修辞段落の相関からも明らかである。

図1. 4-gramバンドル(上位300)の対応分析による4分野の修辞段落の相関

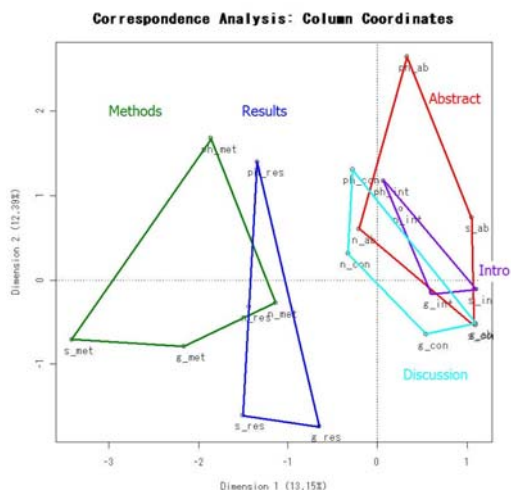
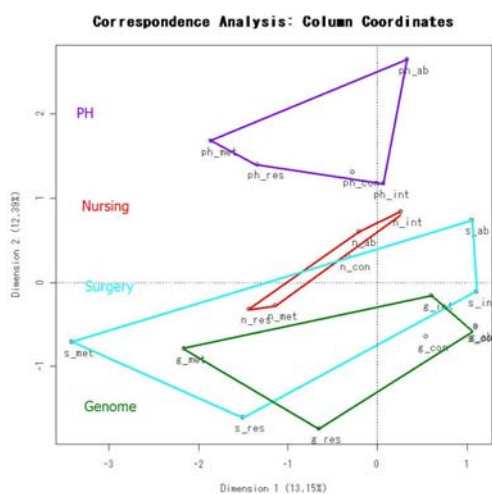


図2. 4-gramバンドル(上位300)の対応分析による4分野の相関



これら全ての統計的分析から得られた結果を基に、医学・看護英語リーディング・ライティング用コースウェア、および検索用コンコーダンサーシステムを試作した。

(1) コースウェア
医学・看護英語リーディング・ライティング教育システム<頻出動詞編>として、Moodle上にシステムのβ版を作成した。学習者に頻出動詞の用法に習熟させることを第一の目的としている。構成は修辞段落ごととし、まず<Introductionの概説>のあとにIntroductionに現れる動詞の上位20語を一

覧表で示し(図3)、そのうち共通して現れる動詞 use, find, show, increase, include について分野別の例文が示されるようにリンクを張った(図4)。

図3. 頻出動詞上位20語

4分野論文のIntroductionに現れる動詞(頻度順)				
	Genome	Surgery	Public Health	Nursing
1	be	be	be	be
2	have	have	have	have
3	use	use	use	use
4	show	show	include	find
5	contain	leave	base	include
6	suggest	increase	provide	provide
7	do	follow	increase	base
8	indicate	report	report	improve
9	express	include	associate	increase
10	find	associate	find	report
11	induce	perform	develop	associate
12	increase	find	identify	develop
13	describe	do	do	do
14	observe	undergo	relate	make
15	identify	compare	show	identify
16	include	induce	compare	relate
17	follow	suggest	make	show
18	determine	develop	examine	describe
19	require	describe	reduce	compare
20	regulate	involve	suggest	need

図4. 図3 Genome の use (赤丸部分) をクリックしたリンク先

"use" in Genome

(1) However, few laboratories now **use** this method and, in practice, novel isolates (and particularly those that presently are unculturable) are usually compared to each other, and to known taxa, by assessing the sequence similarities in their 16S rRNA genes. [source]

(2) To illustrate these points, we will **use** mainly recent examples from our own work. [source]

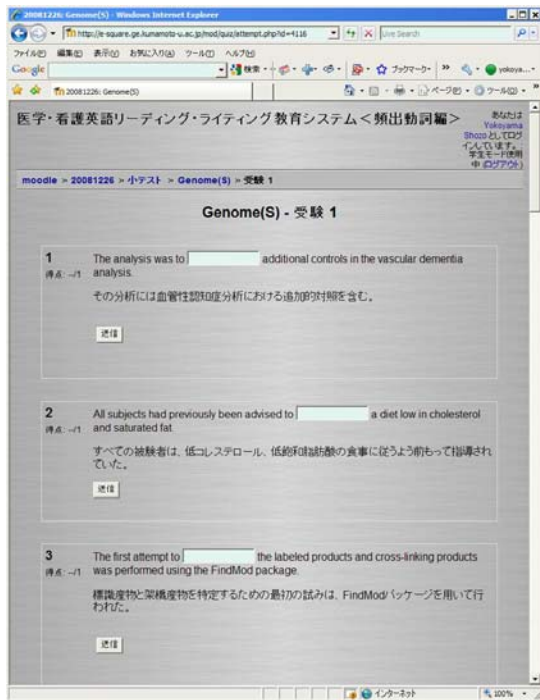
(3) Several rational approaches have been **used** to prevent EAE, including passive transfer of monoclonal antibodies that interfere with the recognition of the MHC, TCR and MBP peptide complex [8 , 9], antibodies against CD4 [10] and T regulatory cells [11 - 14]. [source]

(4) We also show, **using** immunohistochemistry, that the CSF1R protein is expressed in the superficial epithelium of the ileum and colon. [source]

[ページトップへ移動する]

次に<Introductionの練習問題>として学習者がオンライン上で受験回ごとに問題順がシャッフル提示されるテストに臨み(図5)、その場で採点可能なシステムを作成した。教師の介入により学習者は個別に学習を進めることが可能である。

図 5. 練習問題部分



以下同様に、Materials and Methods、Results、Discussion 別に頻出動詞について学習するようデザインされている。

(2) コンコーダンスシステム (Data-driven Learning:DDL システム)

DDLによる教育のためのコンコーダンスシステムを作成し、学習者の自律的学習をサポートする (このConcordancerはMedcorpusのデータを使って、大羽良氏(研究協力者)がFree CSS Templatesを使ってプログラムした) (図 6)。

図 6. ウェブコンコーダンス (検索結果表示)



ここでは 4 分野と全ての修辭段落を選択して検索をした結果(太い点線で囲んだ部分)を

表示した。この場合右上の Options (細い点線で囲んだ部分)で必要なセクションにチェックを入れることで、分野と修辭段落の選択が可能になる(図 7、図 8)。

図 7. ウェブコンコーダンス (分野選択)

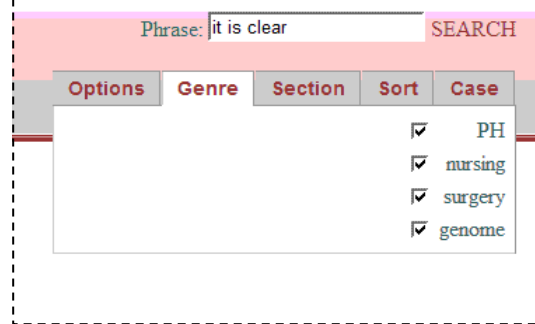
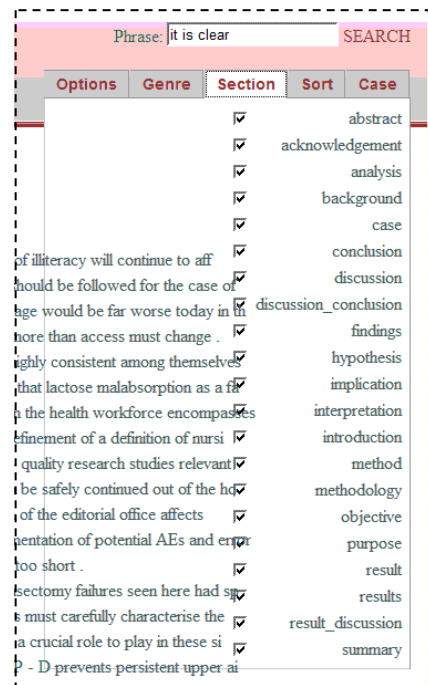


図 8. ウェブコンコーダンス (修辭段落選択)



ウェブコンコーダンスもまだ試作段階であるが、共起語の左右ソートなど今後改良を加えてさらにより使い勝手のよいものにしていけば、医学英語論文教育の方法も大きく改善される可能性があるといえる。

また、最近では学習者の自律性(learner autonomy)が語学教育、とくに英語教育ではキーワードとなっている。DDL を利用した学習の開始直後は上述のように教師介在が基本であるが、自律的学習者に対しては誘導的学習(inductive learning)により教師から「学ぶ」ことと併せて、学習者自らが研究者(learner as a researcher)となり、言語的事実=ジャンル特徴を「発見」していくようなカリキュラムを考えることも重要なポイント

ントになるであろう。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 7件)

- ① Shozo YOKOYAMA, From Needs Analysis Towards Materials Development: Exploring and Designing an EMP Course for Japanese Medical Students, 5th Language for Specific Purposes International Seminar, CD-ROM, 1-21, 2006, 査読有
- ② 横山彰三、鈴木千鶴子、安浪誠祐、Kathleen Brown、川北直子、医学・看護英語教育システム構築に向けたコーパス分析、第45回 JACET 全国大会要綱、68-69、2006、査読有
- ③ 横山彰三、玉田吉行、Hands-on EMP プログラム：宮崎大学医学部の取り組み、第9回 医学英語教育学会抄録集、17、2006、査読有
- ④ Chizuko SUZUKI et al.、Pedagogical implications of a collaborative e-learning environment with a tutorial system、International Journal of Pedagogies and Learning、2(1)、89-98、2006、査読有
- ⑤ Chizuko SUZUKI et al.、Implementing a tutorial system into a BBS communication site、The Journal of the Language and Culture Center, Nagasaki Junshin Catholic University、2、1-14、2007、査読無
- ⑥ 横山彰三、鈴木千鶴子、安浪誠祐、Kathleen Brown、川北直子、医学・看護英語教育システム構築に向けたコーパス分析：高頻度語彙・特徴語・AWL Coverage、ESP の研究と実践、6、2-56、2007、査読無
- ⑦ Shozo YOKOYAMA、Corpus-based Analysis on Medical English: From a Perspective of English for Medical Purposes、Proceedings of 2007 PKETA International Conference、1巻、4-9、2007、査読有

[学会発表] (計 6件)

- ① Shozo YOKOYAMA、Chizuko SUZUKI、Seisuke YASUNAMI、Naoko KAWAKITA、Kathleen BROWN、The discourse of academic writing in medicine and nursing: cross-sectional analysis based on a corpus of EMP、10th International Pragmatics Conference、2007年7月、Goteborg University (スウェーデン)
- ② Shozo YOKOYAMA、Chizuko SUZUKI、Seisuke YASUNAMI、Naoko KAWAKITA、Kathleen BROWN、Corpus-based cross-sectional analysis of academic papers in medicine:

investigating collocation patterns of sub-technical terms、Corpus Linguistics 2007、2007年7月、University of Birmingham (英国)

- ③ Shozo YOKOYAMA、Corpus-based Analysis on Medical English: From a Perspective of English for Medical Purposes、PKETA International Conference、2007年10月、Pusan (韓国)
- ④ Shozo YOKOYAMA、A Cross-sectional Analysis of Lexical Bundles in Written Medical Discourse、American Association for Corpus Linguistics、2008年3月、Brigham Young University (米国)
- ⑤ Shozo YOKOYAMA、Genre Specificity in Academic Corpora: A Cross-sectional Analysis of Distinctive Words in Medical Research Articles、Kumamoto University International Symposium on Corpora and English Education、2008年10月25日、熊本大学 (熊本市)
- ⑥ Shozo YOKOYAMA、A Corpus-based Analysis of Stance in the Japanese Medical Students' Essays、International Language for Specific Academic Purposes Conference、2009年2月7日、クレタ (ギリシャ)

6. 研究組織

(1) 研究代表者

横山 彰三 (YOKOYAMA SHOZO)

宮崎大学・医学部・准教授

研究者番号：60347052

(2) 研究分担者

鈴木 千鶴子 (SUZUKI CHIZUKO)

長崎純心大学・人文学部・教授

研究者番号：10123837

安浪 誠祐 (YASUNAMI SEISUKE)

熊本大学・大学教育機能開発総合研究センター・准教授

研究者番号：00290833

ブラウン キャスリーン (BROWN KATHLEEN)

久留米大学・外国語教育研究所・准教授

研究者番号：70421310

川北 直子 (KAWAKITA NAOKO)

宮崎県立看護大学・看護学部・准教授

研究者番号：80405589