

日本助産学会研究助成金（学術奨励研究助成）研究報告書

助産学・周産期ケアの学術論文で用いられる英語の分析

—英語論文の理解や執筆のために—

千葉 陽子

（京都大学大学院医学研究科人間健康科学系専攻）

本報告書は、日本助産学会誌に投稿中の英語論文の和訳を改訂したものである。

研究協力者：

Neil Millar, Department of Linguistics
and English Language, University of
Lancaster, Lancaster, U.K.

Brian Budgell, Canadian Memorial
Chiropractic College, Toronto, Canada

I. 諸言

昨今の生物医学および健康分野の教育・研究・臨床実践は、国・文化・言語の違いを超えたコミュニケーションを必要とするグローバルな環境の中で行われている。スムーズなコミュニケーションは、政策実現、臨床ケア、共同研究、学生や関係者への新知識の普及などを促進する。従って、Biomedical English と呼ばれるこの分野の共通語への理解を高めることは、関係者にとってかなりの利益になり、結果的には利益がケアの受け手にもおよぶと考えられる。

しかし、Biomedical English を用いて発表された多くの研究にもかかわらず、言語についての量的研究はほとんど行われていない。そのため、生物医学および健康分野の各領域における特有の言語を分析するためのプロジェクトが開始され、この中の一研究として、助産学・周産期ケア (Midwifery and perinatal care : 以下 MPC) 分野の言語の特徴を明らかにすることを試みた。

多様な国籍の研究者が行う本研究において言語 (Language) とは即ち「英語」を指す。英語を母語とする者とそうでない者との間で言語学習負担の大きさが異なることは当然で、特に非アルファベット文化圏にいる私たち日本人特有の考察も必要になるであろう。日本の看護基礎教育における英語教育については (口元&竹内, 2009), 教

育機関ごとにその質・量が異なっており、専門学校と比べて大学において英語必修・選択科目単位数が多い。看護学生への英語教育の目的は、一般のおよび医療現場での会話能力の向上を目指す教育機関が多かったが、少ないながらも「英語論文作成に関する能力の習得」を挙げた教育機関もあり、海外看護論文の読解や英語論文作成へのニーズを認めた。園城寺ら (2006) が 1999 年から 2003 年に日本国内で刊行された雑誌の看護系英語論文を調査したところ、英語論文が採録誌全体に占める割合は 4~5%程度で、多くが日本人著者によるものであった。また、英語論文の掲載誌は主に紀要で、大学で英語論文を執筆する研究者が少しずつ増えていることを示している。しかし Ian と谷本 (2009) が国内看護系学会誌に掲載される英語抄録について行った調査によれば、多くの学会誌が英語のネイティブチェックを求めているものの、ほとんどの抄録がオンラインデータベース上に掲載されておらず、著者が英語抄録作成のために多くの労力とコストをかけても内容を世界に発信できていないことが指摘された。看護分野における英語教育や英語の使用がこのような状況である中、日本助産学会でも英語論文の受付を開始し、論文を海外に発表できる体制となった。

本研究では、学術論文の英語に焦点を絞ったが、これは論文が学生・教育者・研究者・臨床家に対して新知見を発表するための重要な手段だからである。既に、看護 : Nursing (Budgell *et al.*, 2007) や保健 : Public health (以下 PH) (Millar & Budgell, 2008) の英語や文章構造の特徴が一般英語と異なっていることは報告されており、これらに

よると、専門用語はもちろん、文法上の慣行や表現パターンでさえも異なっていることが明らかになった。一般的に、生物医学および健康分野の論文は、読みやすさの指標が低く（つまり読みにくく）、例えば文字数の多い単語の使用や、多くの受動態構文を含む複雑な文章構造で特徴づけられていた（Millar & Budgell, 2008）。特別な英語の使用はまた、各学問分野や専門家グループ内の特有の文化、例えば論文編集者や読者ですら明白に認識できないような（科学的な意味での）バイアスをも反映していると言われている。

II. 研究目的

本研究の目的は、一般英語や PH 分野の英語との比較によって、以下の点を行うことである。

1. MPC 分野の英語論文中の語彙や文章構造の特徴を明らかにすること。
2. MPC 分野における言語学習負担、特に MPC 分野にこれから入ろうとする者で一般的な英語の知識か PH 分野というような MPC 関連分野の英語の知識しか持ってない者にとっての学習負担を考察すること。
3. MPC 分野で特に母語が英語でない者の言語学習負担を考察すること。

III. 研究方法

1. コーパスの作成

MPC 分野の主要学術雑誌 5 誌の研究原著論文・論説・解説・概説で用いられた英語をもとにコーパスを作成した（表 1）。コーパスの和訳は「言語資料（松田，2007）」であり、言語的な特別な基準により選ばれ

並べ替えられた言葉の集まりで、対象分野の文献の代表的言語サンプルとして様々な目的に使用される（McEnery *et al.*, 2006）。これら 5 誌を選んだのは、電子フォームでの入手が可能で（コンピューター処理のため）、MPC 分野のものであり、インパクトファクターが付与されていたことによる。MPC 分野のコーパス（以下 MPC コーパス）は、看護や PH 分野のコーパス分析と同様、2005 年刊行の英語雑誌によって作成された。表 1 には、各雑誌から抽出された延べ語数（Tokens：言語学の専門用語で総語彙数のこと）、異なり語数（Types：異なる語彙の数で同一語彙が何回現れても 1 とカウントされる）を示した。Midwifery のみがイギリスの雑誌で、それ以外の 4 誌はアメリカ刊行のものであった。

2. コーパスの比較

まず、対象論文を XML（Extensive Markup Language）文書として保存し、ソフトウェア WordSmith 5.0（Oxford University Press, Oxford）を用いて MPC コーパスを 2 つの参照コーパスと比較した。一般英語のコーパスと、PH 分野のコーパス（Millar & Budgell, 2008）である。一般英語のコーパスは the American National Corpus（Linguistic Data Consortium, Philadelphia）の 2 次的コーパスで、2002 年刊行の New York Times で用いられた延べ語数 3,625,687 から成るコーパスである（以下 NYT コーパス）。PH 分野のコーパスは 2005 年刊行の英語主要学術雑誌 4 誌で用いられた延べ語数 1,921,278、異なり語数 47,678 からなるコーパスで、以下 PH コーパスとする。

3. キーワードの抽出

本研究では学習者の前に立ちほだかる専門的語彙の抽出を目指すため、NYT コーパスや PH コーパスと比較する前にストップリスト (van Rijsbergen, 1979) と呼ばれる機能語 (前置詞, 代名詞, 冠詞) を除外した。また、MPC コーパスにおいて閾値 10 を設定し、出現回数が 10 回以下の語彙を言語学習負担にほとんど影響がない稀出語として分析から外した。これは、総語彙 1,011,870 語中 10 回 (10/1,011,870), つまり約 10 万語に 1 回 (1/100,000) より少し多い頻度ということになる。

続いて、2 つの参照コーパスとの比較により MPC コーパスのキーワードを抽出した。コーパス言語学でキーワードとは、大きな参照コーパスと比較した際に当該コーパスで偶然として期待されるよりも有意に多く出現する語のことを言う (Baker *et al.*, 2006)。異なり語 (Type) が参照コーパスと比べて MPC コーパスの中でどの程度有意に現れるかを示すキーネス (Keyness) は、対数尤度で計算される。これはカイ二乗検定に似た統計で、2 つのデータセットについて観察された値と期待される値を比較するが正規分布であることは仮定しない (Dunning, 1993)。慣例的に、閾値 15.13 を MPC コーパス中に異なり語が有意に多く現れる回数とした (Rayson *et al.*, 200)。

さらにキーワードのばらつきを、異なり語が対象 5 誌のうち何誌に、また全 312 論文中いくつの論文に有意に現れたかで示した。この分析を行ったのは、たった 1 本か数本の論文中にかなり出現した語が全コーパス中で代表的な語とされたとしても、当該分野で重要な語とは言えないかもしれないからである。MPC コーパスのから例を挙

げると、SARS (Severe Acute Respiratory Syndrome: 重症急性呼吸器症候群) という語彙が有意に出現し (出現 133 回, キーネス 389.3), 特に重要な語彙であると思われた。しかし、この語彙は 2002 年には全く用いられておらず NYT コーパス中にはなかった。しかも、SARS という語彙 133 回全てが 1 雑誌の 1 論文でのみ用いられていたため、MPC 分野全体で代表的な語彙でないことは明らかである。そのため、異なり語が全コーパスをどれくらい広くカバーしているかという「ばらつき」の概念は、異なり語が当該分野中でいかに重要な語かということを示す上で重要である。

キーワードは、以下の 2 つの外部単語リストに分類された。1 つ目は General Service List (以下 GSL) で、英語の最も一般的な 2,000 の単語家族 (word family) から成り (West, 1953), 2 つ目は Academic Word List (以下 AWL) で、学術分野の約 570 の単語家族から成る (Coxhead 2000)。単語家族とは 1 つの語源から派生した語彙のグループで、例えば woman, women, women's 一家族となる。GSL にも AWL にも現れなかった語彙は、Off-list に分類した。

4. キーフレーズの抽出

キーワードに加え、総語彙の中から頻出フレーズ (連語 3 から 7) を抽出した。

5. 読みやすさ指標の計算

5 誌それぞれから 10,000 語の語彙サンプルを抽出し、MS Word 2008 V12.1.0 (Microsoft Corporation) の統計機能を用いて読みやすさ指標を計算した。これらのサンプルは、各雑誌の 1 番目の論文から始まり 10,000 語目の単語を含む文章までをとったもので、ランダムサンプリングとは

言えないが、5誌とも同条件で抽出した。

文章の相対的な読みやすさを計算する式はいくつかあり、単語や文章の長さに基づくものが多い。本研究で用いた指標は、1) 1文あたりの平均単語数、2) 1語あたりの平均文字数、3) 受動態文の割合、4) Flesch Reading Ease (1文あたりの単語数、1単語あたりの音節数、1単語あたりの文字数によって計算したもの) (Flesch, 1948) である。1) 2) 3) のスコアが低いほど、4) のスコアが高いほど文章が読みやすいとされる。

6. 論文タイトルのキーワード・キーフレーズの抽出

対象論文のタイトルの英語のみを取り上げ、NYT コーパスと比較して有意に出現する語を抽出した。また、フレーズ (連語 2 から 5) も抽出した。

III. 結果

1. 一般英語との比較におけるキーワード

NYT コーパスとの比較の結果、MPC コーパス中より 3,590 語がこの分野で有意に出現する語として抽出された。つまりこれら 3,590 語は、一般英語中の広がりから期待される頻度よりも MPC 分野で多く出現する語である。その後、約 10 万語に 1 回以下しか現れない異なり語を除き、抽出語は 3,124 まで減少した。また 5 誌中 1 誌のみにしか現れない異なり語を除いて抽出語は 1,669 に、さらに 5% 以下の論文 (312 論文中 15 論文以下) にしか現れない語を除き、最終的に 1,108 語を抽出した (図 1)。こうして抽出された語をコアワード (Core words) とする。

これらのコアワードのうち 335 語が頻出

一般英語の上位 1000 語の中に、さらに 141 語が上位 2000 語の中にあり、合計 476 語が GSL に分類された。コアワードのうち 390 語は AWL の単語家族中にあり、残り 242 語は GSL にも AWL にも属さないため Off-list に分類された。表 2 は、MPC 分野で有意に出現するコアワードの上位 30 語を GSL・AWL・Off-list 別に示したもので、語彙は出現頻度 (Frequency) の高い順から並んでいる。既述のように、表 2 のキーネス (Keyness) は抽出された単語が一般英語と比べて MPC コーパス中で有意に多く現れるかを示し、値が高いほど有意性が強い。ばらつき (Dispersion) は、312 論文中いくつかの論文に、また 5 誌中いくつかの雑誌に当該単語が有意に出現したかを示している。全キーワードは、ウェブサイト <http://bmhlinguistics.org/joomla2/language-of-midwifery> に掲載されている。

図 2 は、GSL・AWL の語に Off-list の語を加えることで MPC コーパスの全語彙をどれくらいカバーできるかを示している。コンピューター分析によると、GSL と AWL の単語だけで MPC コーパスの 85.55% の単語をカバーする。これらに Off-list の単語で出現回数上位 1000, 2000, 3000 に含まれる語を追加すると、それぞれ 94.84%, 96.40%, 97.23% の MPC 分野の語彙をカバーできる。つまり、もし GSL の単語家族 2000 語、AWL の単語家族 570 に加えて、3000 の Off-list 語を知っていれば、本研究のサンプルである MPC 分野の文献中 97.23% の語を読めることになる。MPC コーパスの語彙 100% をカバーするには、Off-list 語を 17,148 も習得する必要がある。一方、Off-list コアワード 242 語は、一般英

語と比べて MPC コーパスで統計学的に有意に多く出現する語であり、これらを GSL と AWL に追加することによって MPC コーパスの全語彙の 89.96% をカバーできることが示された。

MPC 分野のコアワードは、母親、子ども、ケア提供者と、これら三者の関わり合いに関する語が多く含まれるという特徴を示した。上位 100 のキーワード中、解剖学的な語は 1 語のみ (#51 breast : 胸)、病状を示す語は 3 語のみ (#14 pain : 痛み, #55 depression : 抑うつ, #94 stress : ストレス) であった。キーワードはまた、専門分野ごとの用語の質的な違いも明らかにした。MPC 分野の論文は、ケアの包括的、社会的、情緒的側面に関する語が多く、例えば family (家族), bonding (密接な結び付き), feelings (感情), anxiety (不安), quiet (静かな・穏やかな), relaxation (くつろぎ、緊張などの緩和), experience (経験) という語を認めた。

2. PH 分野の英語との比較によるキーワード

PH 分野のコーパスと比較した結果、MPC コーパス中で有意に多く現れる語として異なり語 1,841 が抽出された。10 万語に 1 回以下の頻度で出現する語を除去した結果、1,722 語となった。また 5 誌中 1 誌にしか現れなかった語を削除してキーワードは 770 語に絞られ、さらに全論文中 5% 以下の論文にしか現れない語を除外した結果、410 語のキーワードが抽出された。そのうち、205 語が GSL, 80 語が AWL, 125 語が Off-list に分類された (図 1)。PH 分野のコーパスと比べて MPC 分野で有意に多く現れなかった語 (ネガティブ・キー

ワード) は、論理、理由・根拠、科学的手法に強く関連したもので、例えば, subject (対象), costs (費用), data (データ), statistical (統計学的な), control (対照), results (結果) という語であった。このことは, Hunter (2006) が助産学は「看護者とケアの対象のそれぞれの出会いの唯一性を重んじる科学である。」と言及した見解と一致した。

3. キーフレーズ

MPC コーパスの中で 5 回以上出現したフレーズをカウントしたところ、7 フレーズが 91 セット、6 フレーズが 255 セット、5 フレーズが 891 セット、4 フレーズが 3,559 セット、3 フレーズが 12,829 セット抽出された。表 3 には、出現回数の上位 5 つのフレーズとフレーズが現れた論文数を示している。フレーズの内容を見ると、研究目的・方法を示す表現や、英語圏 (特にアメリカ) の MPC 関連団体や機関の名称が認められた。

4. 読みやすさ指標

表 4 は、5 誌それぞれから抽出した 10,000 語のサンプルについて、読みやすさ指標を計算した結果である。1 文あたりの平均単語数は 22.4 ± 1.6 、1 単語あたりの平均文字数は 5.4 ± 0.2 であった。Flesch Reading Ease 指標は 30.7 ± 5 であった。受動態構文の使用は文章の複雑さに関係し、本研究のサンプル文章の $29.4\% \pm 6\%$ が受動態であった。5 雑誌中、Journal of Midwifery and Women's Health が最も「読みやすい」という結果を得た。

5. 論文タイトルのキーワード・キーフレーズ

MPC 分野の論文のタイトルに用いられた英語のみを抽出したところ、総語彙数

3,260, 異なり語数 1130 であった。これらを NYT コーパスと比較したところ, MPC 分野で有意に出現するキーワードは 96 で, このうち機能語 9 つを除去して最終的に 87 語となった。表 5 には, これら 87 語をキネスの順に示した。キーワードの中には, *perineal* (#41: 会陰の), *vaginal* (#53: 膣の), *cervical* (子宮頸管部の), *resuscitation* (#66: 蘇生), *kangaroo* (#78: カンガルー) など, 解剖学的用語や新生児ケアについての特殊な語を認めた。また, 表 6 には 2 連語から 5 連語のキーフレーズを示した。キーフレーズの中には, アメリカの職能団体名のほか, *randomized controlled trial* や *evidence-based practice* を多く認めた。

IV. 考察

本研究では, MPC 分野の英語学術論文を代表的サンプルとして, この分野の英語の語彙および文章構造の特徴を分析した。一般英語 (NYT コーパス) との比較により, MPC 分野に新しく入る者にとって重要な言語学習負担となるコアワードを抽出した。また一部 PH コーパスとの比較結果を示した。5 誌のサンプル文章から読みやすさ指標も計算した。

MPC コーパスと外部単語リストとの比較は, MPC 分野の語彙へのアクセス度を示した。本研究では外部単語リストとして GSL (英語の最も一般的な 2,000 語の単語家族) と AWL (学術分野で用いられる 570 語の単語家族) を用いており, GSL 中のキーワードは一般英語を頻繁に使うことによって自然に学べ (Chung & Nation, 2004), 同様に AWL 中のキーワードは, 学術論文

に広くあたることによって習得しやすくなる。しかし, GSL・AWL の語の中でも MPC 分野の特別な意味・ニュアンスをもつ語には注意が必要である。GSL キーワードの例としては, *failure* (一般的には失敗, MPC 分野では *interrupted labor*: 流早産), *delivery* (一般的には配達, MPC 分野では *the childbirth*: 分娩) があつた。AWL キーワードの例としては, *formula* (一般的には計算式・式・決まり文句, MPC 分野では *milk substitute*: 母乳の代用物・粉ミルク), *labor* (一般的には労働, MPC 分野では *the process of childbirth*: 出産のプロセス) が挙げられた。これらの語の意味は辞書の最初の方に現れないため, 特に英語が母語でない者は意味を覚え使い慣れることが必要である。

GSL から AWL へ, さらに Off-list へと続く語彙全体の連続性にもかかわらず, 一般英語と専門分野の英語との間には違いがあり, Off-list 語は特に専門的で特別な勉強なしでは習得しにくいと考えられた。従って Off-list の単語, 例えば *postpartum* (産後の), *antenatal* (産前の), *gestational* (妊娠の), *episiotomy* (会陰切開), *analgesia* (無痛) などは, この分野に入る者にとって最も大変な言語学習負担になると考えられた。

コアワード抽出のために用いた基準は, 出現頻度 10 万語に 1 回以上と, ばらつきとして全論文中 5%以上の論文に出現し, 5 誌中 2 誌以上に出現することであつた。これらは任意に設定されたが, 先行研究との比較を行えるという点で有効である (Budgell *et al.*, 2007; Millar & Budgell, 2008)。類似点がある分野, 例えば PH 分野での知識

があれば MPC 分野での言語学習負担は小さいとも考えられる。現時点では看護分野のコーパスとの比較分析は行っていない。

本研究はまた、一般英語を理解できるレベルの英語力で MPC 分野の英語を理解するために必要な語彙数についても言及した。広範囲にわたる一般英語を流暢に読み、知らない語彙の意味を推測するためには、論文中の 95%の語彙を理解する必要があると推定されており、このことは約 15,000 の異なり語の習得が必要ということを示す (Hirsh & Nation, 1992; Laufer, 1992)。しかし、MPC 分野の英語の 95%をカバーするためには、GSLの単語家族 2,000, AWLの単語家族 570 と、Off-list 語の知識があればよいことになる。出現頻度上位 1,000 の Off-list 語を加えると MPC コーパス中の語彙の 94.8%の語をカバーできるため、これは一般英語を流暢に読むために必要な語彙の閾値 (95%) とほぼ一致する。その後、さらに 1000 語の Off-list キーワードを加えたとしても、これらの語はあまり出現しないため、ただ単語力が向上するだけと考えられる (図 2 参照)。GSL・AWL の語彙と、242 の Off-list コアワードを合わせることによって、MPC コーパス中の全論文の 89.9%はカバーできるようになる。242 のコアワードは、統計学的に有意に多く出現し広く分散しているため、この分野の英語を理解できるようになるためには、まずこの 242 語に焦点を当てて勉強の方が焦点を絞らないアプローチよりかなり効果的であると考えられた。

英語を流暢に操ることは、単に個々の語彙に対する知識があればよいだけでなく、語彙が意味のある表現となるためにどのよ

うに繋がっているかを理解することも重要である (Nation, 2001)。本研究で抽出したフレーズは、英語学習者や教師にとっては必ずおさえるべき意味のある典型的表現で、特に MPC 分野に関連する組織の名前や、研究の方法に関する表現を含んでいた (表 3)。本研究対象の雑誌は英米系のものであるため、特に英語を母語としない外国人としての日本人は、頻出フレーズの意味を十分に理解しておく必要がある。例えば、アメリカ看護師助産師協会 (The American College of Nurse Midwives: ACNM) の表現に用いられる college を「大学」と訳すと誤解が生じ、アメリカ疾病予防管理センター (The Centers for Disease Prevention and Control: CDC) がどのような機関かを理解しておくことは文章の理解に有効である。研究目的や方法を示すフレーズを知っておくことは、論文を書く上でも役立つ。

MPC 分野の論文タイトルの英語は、人々が最もよく目にする語彙である。文検索の際、抄録も読まずにタイトルだけで関心のある文献かどうかを判断することもあるため、タイトルに用いられている英語を理解することは重要である。これらは、全 MPC コーパスのキーワードとして抽出された語を習得することによってほぼカバーされるが、解剖学的な語彙や新生児ケアの語彙の出現が特徴的であった。また、evidence-based practice や randomized controlled trial という用語から、臨床実践で求められるエビデンスを検証する論文や、ランダム化比較試験を行う研究があることも示唆され、MPC 分野がより科学的な学問として発展しようとする姿もうかがえた。このように、論文で取り上げられているト

ピックやそのトレンドを知る上でも、コーパス分析を活用できると考えられた。

英語文献の読みやすさに関しては、一般英語と比べて MPC 分野の英語には文字数の多い語 (1 単語あたり 5.4 ± 2 文字) や語彙数の多い文章 (1 文あたり 22.4 ± 1.6 語) が含まれていた (表 4)。Flesch Reading Ease 指数が 0 から 30 の英語の場合、これを読むにはアメリカの大学卒業程度のレベルの英語力が必要と言われており (Flesch, 1948)、Flesch Reading Ease 指数が 30.7 である MPC 分野の英語を読むにはこのレベルの英語力が必要であることが示唆された。Flesch Reading Ease が 23.2 であった PH 分野の英語文献よりは、いくらか読みやすいことが分かった (Millar & Budgell, 2008)。アメリカの大学卒業程度の英語能力を具体的に示すのは難しいが、参考として、外国人がアメリカの大学院に入学を希望する場合 (大学卒業レベルの学力がある場合) に求められる TOEFL 《Test of English as a Foreign Language》のスコアの一例は、600 (Paper based test), 250 (Computer based test), 84 (Internet based test) であり、結構高い (Boston University School of Arts 2010)。

受動態構文の多さは読みにくさに関係し、当該分野の文章中 $29.4 \pm 6\%$ が受動態で、これは PH 分野の文章 ($26.0 \pm 4\%$ が受動態) よりやや高い割合を示した (Millar 2008)。受動態での間接的言い回しが必要であったり、能動態による直接的な表現が適切でなかったりする場合もあると思われるが、これについての質的分析は今後の課題である。

最後に、英語が母国語でない者は特に語彙力数を増やすための努力が必要であるが、

本研究の結果を受けて普段から GSL や AWL の語に意識的に接する努力をするとともに、助産師教育の最初の時点でコアワードやキーフレーズの学習を取り入れるなどの工夫が必要であろう。論文を読み内容を理解するための訓練としては、5 誌のうち最も読みやすいとされた Journal of Midwifery and Women's Health の論文を用いることが適切であると示唆された。

V. まとめ

MPC 分野の英語は、一般英語とは異なる特徴があり、PH 分野との共通点がある分野ながら、出現頻度が有意に異なる語彙もあった。MPC 分野の英語には、特徴的なキーワードとフレーズがあり、これらの語は、母親・子ども・ケア提供者間の関わり合い、出産のプロセス、包括的ケアの重要性に関する意味をもつものが多く認められた。身体や病状を示す言葉はあまり抽出されなかった。

保健医療分野において医療従事者の国家間移動を含む教育、研究、臨床実践におけるグローバル化が進んでいる中、MPC 分野の英語は焦点を絞って学習することによって比較的習得しやすくなると考えられた。つまり、一般英語とは別の学習方法によってアプローチする必要がある、当該分野のコアワードにフォーカスした英語学習プログラムを構築することによって、関係者、特に母語が英語でない者にとっての言語学習負担を軽減させることができると考えられた。

今後の課題としては、本研究結果を用いて助産師や助産師学生の実際の言語学習負担の軽減について実証していくことが挙げ

られる。また、本コーパス分析と5年後、10年後のMPCコーパス分析を比較することによって、この分野で重要な語彙やトピックの変化を言語学的見地から把握することも可能であると考えられた。

追記

本研究の一部は、第24回日本助産学会学術集会（つくば市）で発表した。

また、本報告書の末尾に「資料：MPC分野の英語の学習のために」を添付し、本研究について掲載したウェブサイトや、本文では紹介しなかった頻出動詞（過去分詞）や品詞別（副詞・形容詞）キーワードリストを紹介しているので、MPC分野の英語力向上のために活用されたい。

参考文献

園城寺康子, 六本木淑恵, 渡邊容子. (2006). 国内出版看護系雑誌にみる英語論文掲載の調査と分析: 「最新看護索引」1999-2003を対象として. *看護と情報*, 13, 59-66.

口元志帆子, 竹内久美子. (2009). 看護基礎教育における英語教育の実態調査: 全国看護系大学・短期大学・専門学校の調査結果から. *目白大学健康科学研究*, 2, 49-54.

松田徳一郎編集代表. (2007). *リーダーズ英和辞典* (第2版). 株式会社研究者.

Ian W, 谷本公重. (2009). 看護系雑誌に掲載される英語抄録のネイティブチェックとオンラインデータベース掲載に関する調査.

香川大学看護学雑誌, 13(1), 33-41.

Baker, P. Hardie, A. McEnery, T. (2006). *A glossary of corpus linguistics*, Edinburgh: Edinburgh University Press.

Boston University School of Arts. (2010). www.bu.edu/artsadmin/admissions/toefl_scores (Accessed on March 17, 2010)

Budgell, B. Miyazaki, M. O'Brien, M. et al. (2007). Developing a corpus of the nursing literature: a pilot study, *Japan Journal of Nursing Science*, 4, 21-25.

Chung, TM. Nation, P. (2004). Identifying technical vocabulary, *System: Intl J Educ Technol Appl Linguist*, 32, 251-263.

Coxhead, A. (2000). A new academic word list, *TESOL Q*, 34, 213-238.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence, *Computational Linguistics*, 19(1), 61-74.

Flesch, R. (1948). A new readability yardstick, *J Appl Psychol*, 32, 221-233.

Hirsh, D. Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.

Hunter, LP. (2006). Women give birth and pizzas are delivered: language and

western childbirth paradigms, *J Midwifery Womens Health*, 51, 119-124.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud PJ, Béjoint H. (Eds.), *Vocabulary and applied linguistics* (pp. 126-132), London: Macmillan.

McEnery, T. Tono, Y. Xiao, Z. (2006). *Corpus-based language studies: an advanced resource book*, London: Routledge.

Millar, N. Budgell, B. (2008). The language of public health – a corpus based analysis, *The Journal of Public Health*, 16(5), 369-374.

Nation, ISP. (2001). *Learning vocabulary in another language*, Cambridge: Cambridge University Press.

Rayson, P. Berridge, D. Francis, B. (2004). Extending the Cochran rule for the comparison of word frequencies between corpora, Paper presented at the 7th International Conference on Statistical Analysis of Textual Data, 10–12 March, Louvain-la-Neuve.

Van Rijsbergen CJ. (1979). *Information retrieval* (2d ed), London: Butterworths.

West, MP. (1953). *A general service list of English words: with semantic frequencies*

and a supplementary word-list for the writing of popular science and technology, London: Longmans Green.

図表

表1 研究対象雑誌とMPCコーパス

Journal (2005 impact factor)	Papers	Tokens	Types
Birth (1.836)	42	125,202	7,292
J Midwifery Womens Health (0.758)	96	276,350	12,970
J Obstet Gynecol Neonatal Nurs	90	305,599	12,705
J Perinat Neonatal Nurs (0.654)	48	137,852	9,429
Midwifery (0,746)	36	166,867	8,286
Total	312	1,011,870	23,316

* Papers: 論文数

* Tokens: 延べ語数

* Types: 異なり語数(重なる語をカウントしない場合)

表2 NYTコーパスと比較した際のMPCキーワード上位30 (外部単語リスト別、出現頻度順)

	GSL				AWL				OFF-LIST			
	Key word	Freq.	Keyness	Disp.	Key word	Freq.	Keyness	Disp.	Key word	Freq.	Keyness	Disp.
1	women	8094	16215.5	266\2	significant	913	1397.71	212\5	pregnancy	2116	5807.89	195\5
2	care	5290	10947.4	298\5	factors	882	1890.79	187\2	infant	1552	4216.41	170\4
3	study	3283	6065.87	251\5	period	791	967.25	189\5	midwifery	1488	4353.94	103\2
4	support	2100	3129.95	232\3	outcomes	786	2205.63	191\5	postpartum	1475	4299.82	129\2
5	risk	1492	2082.18	216\4	process	751	696.66	165\3	maternal	1284	3635.81	171\5
6	hospital	1412	2482.38	182\3	positive	696	1123.65	145\5	breastfeeding	1020	2968.86	62\4
7	reported	1336	1864.82	214\2	intervention	665	1691.4	118\5	pregnant	815	1900.46	136\3
8	nurses	1291	3358.63	167\5	available	619	427.61	185\5	midwife	703	2030.18	98\2
9	studies	1217	2146.28	215\5	professional	599	775.51	150\5	childbirth	568	1625.69	108\5
10	education	1201	1693.91	187\3	status	574	913.23	160\2	n	533	70.84	106\3
11	delivery	1103	2622.42	183\3	issues	557	377.67	176\5	newborns	468	1307.83	59\5
12	treatment	988	1505.09	142\4	identified	531	811.22	176\5	vaginal	423	1042.26	89\5
13	based	969	471.13	245\5	primary	499	777.77	159\5	diagnosis	415	991.04	95\3
14	age	964	1090.48	200\2	specific	498	727.91	185\2	scores	414	710.16	81\5
15	groups	949	1042.35	189\2	focus	485	514.77	149\2	cancer	404	207.56	41\2
16	experience	923	1148.85	184\5	appropriate	470	828.99	186\5	antenatal	401	1173.01	64\2
17	patient	904	1815.12	138\3	assessment	464	972.31	135\5	acnm	373	1091.09	29\2
18	screening	895	1955.24	87\5	access	449	422.08	133\2	emotional	355	522.56	92\3
19	important	886	638.43	232\5	professionals	439	899.84	122\3	postnatal	349	1020.88	65\2
20	related	877	1444.3	230\5	significantly	433	756.41	133\5	interviews	332	461.61	70\4
21	work	852	16.91	181\3	community	431	186.88	114\5	abuse	329	206.96	41\3
22	high	843	71.3	223\3	potential	431	307.83	170\3	counseling	301	619.42	66\3
23	weight	838	1514.95	94\5	communication	400	886.36	93\5	questionnaire	299	798.03	79\5
24	pain	835	1543.15	94\5	survey	387	497.53	100\5	african	278	230.49	50\2
25	findings	818	1790.53	198\5	strategies	382	710.53	108\4	physician	276	554.29	94\5
26	results	809	1066.46	214\5	approach	379	277.1	132\2	adolescents	256	694.81	31\4
27	low	787	522.15	179\3	variables	371	1004.83	97\5	reproductive	255	536.31	68\5
28	associated	772	1438.53	197\4	perceived	369	721.71	94\5	childbearing	246	707.08	81\5
29	number	758	344.24	219\5	labour	355	1038.44	29\4	therapy	240	157.16	59\2
30	life	729	32.26	171\2	identify	350	533.68	152\5	versus	232	441.42	88\4

表3 MPCコーパスからのキーフレーズの抽出

7-grams	Freq.	Texts
the purpose of this study was to	35	28
the American college of obstetricians and gynecologists	32	26
the purpose of this article is to	30	29
the centers for disease control and prevention	18	14
between the ages of # and #	16	13
6-grams	Freq.	Texts
the American college of nurse midwives	85	37
the American college of nurse midwifery	36	7
department of health and human services	29	21
there were no significant differences in	19	11
the first # months of life	18	8
5-grams	Freq.	Texts
# p less than #	134	39
ranged from # to #	82	38
between # and # weeks	43	25
the findings of this study	37	23
the second stage of labor	34	8
4-grams	Freq.	Texts
in the United States	332	131
at the time of	137	75
# to # weeks	111	47
on the basis of	102	61
at # weeks postpartum	93	10
3-grams	Freq.	Texts
as well as	379	174
the use of	369	140
in this study	340	110
health care providers	230	91
in table #	228	107

*#は数字を示す

*Texts: フレーズが現れた論文の数

表4 読みやすさ指標のスコア

Journal	Words/Sentence	Characters/Word	Passive	Flesch Reading Ease
Birth	23,7	5,5	30%	25,8
J Midwifery Womens Health	19,9	5,1	19%	36,7
J Obstet Gynecol Neonatal Nurs	23,7	5,3	34%	33,5
J Perinat Neonatal Nurs	21,8	5,5	29%	26,7
Midwifery	22,9	5,5	35%	30,6
Mean (S.D.)	22.4 (1.6)	5.4 (.2)	29.4% (6%)	30.7 (5)

表5 NYTコーパスと比較して有意に出現する論文タイトルの英語の抽出

N	Key word	Freq.	Keyness	N	Key word	Freq.	Keyness
1	WOMEN	62	342.86	45	NURSING	7	49.40
2	MIDWIFERY	24	332.99	46	MIDWIFE	4	47.84
3	MIDWIVES	19	255.65	47	FEEDING	6	47.40
4	CARE	41	236.71	48	RESOURCES	9	47.30
5	PERINATAL	17	235.83	49	SAFETY	10	45.65
6	PREGNANCY	23	235.60	50	ADOLESCENTS	4	42.03
7	BIRTH	27	227.49	51	LABOR	9	41.69
8	POSTPARTUM	15	200.60	52	LABOUR	3	41.61
9	CESAREAN	10	138.71	53	VAGINAL	5	41.33
10	CHILDBIRTH	11	138.04	54	CLINICAL	6	37.88
11	BREASTFEEDING	10	132.00	55	COMPETENCIES	3	37.11
12	HEALTH	29	131.27	56	INTERVENTIONS	4	36.43
13	NEONATAL	10	124.67	57	GENETIC	6	36.10
14	MATERNAL	11	122.10	58	RESEARCH	11	35.95
15	CHILDBEARING	9	118.33	59	ETHICAL	5	34.37
16	PRACTICE	20	116.77	60	EVALUATION	4	34.26
17	INFANT	12	112.54	61	CERVICAL	3	32.05
18	NURSE	12	106.50	62	NUTRITION	4	31.96
19	EXPERIENCES	12	98.00	63	ACUTE	4	31.96
20	OBSTETRIC	7	91.06	64	CONTROLLED	6	31.41
21	INFANTS	9	90.49	65	EVIDENCE	8	31.18
22	NEWBORN	8	88.79	66	RESUSCITATION	3	31.03
23	PRETERM	6	83.22	67	IMPROVING	5	31.00
24	PRENATAL	7	80.80	68	METHOD	5	30.13
25	SUPPORT	18	77.85	69	INTERVENTION	4	29.97
26	STUDY	17	73.18	70	SURVEY	6	29.81
27	PATIENT	11	70.73	71	TRIAL	8	29.70
28	POSTNATAL	5	69.34	72	PAIN	6	29.58
29	ANTENATAL	5	69.34	73	SKIN	6	29.46
30	SCREENING	10	68.91	74	REPRODUCTIVE	4	28.94
31	MOTHERS	9	66.90	75	PREMATURE	4	28.74
32	SOCIAL	13	66.35	76	EFFECTS	6	28.42
33	WOMEN'S	12	64.71	77	PREVALENCE	3	28.13
34	NURSES	8	63.66	78	KANGAROO	3	28.13
35	OUTCOMES	6	63.04	79	HOSPITAL	7	27.81
36	RANDOMIZED	6	62.07	80	UPDATE	4	27.52
37	IMPLICATIONS	7	58.95	81	PRACTICES	6	27.51
38	PREGNANT	8	58.08	82	DISPARITIES	3	27.08
39	MATERNITY	6	57.46	83	ATTITUDES	4	26.61
40	SYNDROME	6	56.26	84	INDUCTION	3	25.79
41	PERINEAL	4	55.47	85	PERCEPTIONS	3	25.41
42	DELIVERY	8	55.22	86	BREAST	5	25.27
43	DEPRESSION	8	50.72	87	ASSESSMENT	4	24.27
44	FACTORS	8	50.72				

*同一単語家族のものをまとめていない。

表6 論文タイトルのフレーズの抽出

5-grams	Freq.
THE AMERICAN COLLEGE OF NURSE-MIDWIVES	3
CURRENT RESOURCES FOR EVIDENCE-BASED PRACTICE	3
4-grams	
THE AMERICAN COLLEGE OF	6
A RANDOMIZED CONTROLLED TRIAL	4
IN PERINATAL AND NEONATAL	3
CURRENT RESOURCES FOR EVIDENCE-BASED	3
AMERICAN COLLEGE OF NURSE-MIDWIVES	3
RANDOMIZED CONTROLLED TRIAL OF	3
RESOURCES FOR EVIDENCE-BASED PRACTICE	3
3-grams	
AMERICAN COLLEGE OF	6
THE AMERICAN COLLEGE	6
PERINATAL AND NEONATAL	5
RANDOMIZED CONTROLLED TRIAL	5
A RANDOMIZED CONTROLLED	4
SOCIAL SUPPORT IN	4
OF POSTPARTUM DEPRESSION	4
2-grams	
SOCIAL SUPPORT	9
MIDWIFERY PRACTICE	8
PATIENT SAFETY	7
AMERICAN COLLEGE	6
CONTROLLED TRIAL	6
HEALTH CARE	6
POSTPARTUM DEPRESSION	6

* 5 and 4-grams: many other phrases appeared twice.

* 3-grams: many other phrases appeared three times.

* 2-grams: many other phrases appeared five times.

	NYTコーパスとの比較	PHコーパスとの比較
参照コーパスと比べてMPC分野とより関連している語 (有意に出現する語)	3,590 types	1,841 types
↓	↓	↓
稀出語 (MPCコーパス中10回以下の出現 $10/1,001,870 \approx 1/100,000$)	↓	↓
つまり10万語に1回程度しか現れない単語を除去	↓	↓
	3,124 types	1,722 types
↓	↓	↓
稀出語 (5雑誌中1雑誌のみに出現) を除去	↓	↓
	1,669 types	770 types
↓	↓	↓
稀出語 (5%以下の出現 <math>< 312</math> 論文中15論文以下に出現) を除去	↓	↓
Core keywords	1,108 types	410 types
Core keywords の内訳		
GSL	476	205
AWL	390	80
Off-list	242	125

図1 MPCコーパスからのキーワードの抽出

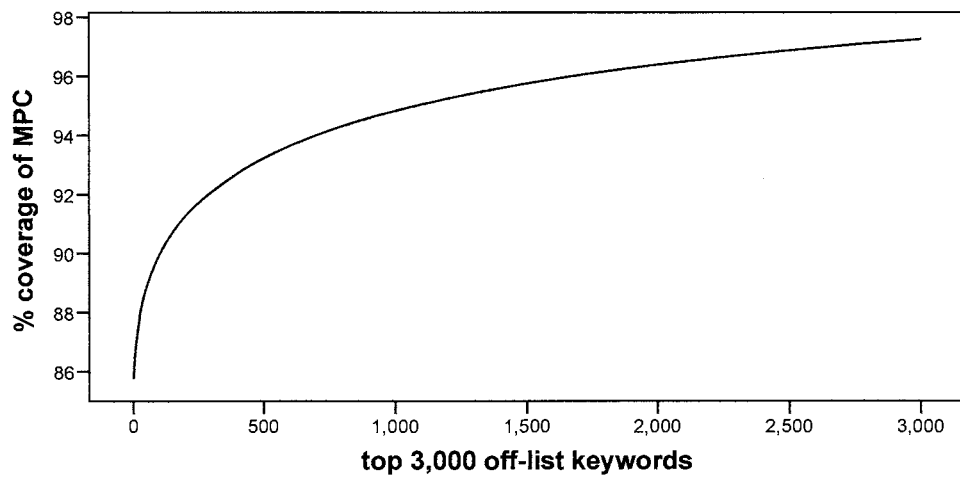


図 2 Off-list キーワード上位 3,000 語と MPC コーパス中の語彙のカバー範囲との関係

資料 MPC 分野の英語学習のために

資料 1 ウェブサイト Home

- 本研究の概要を Biomedical and Health Linguistics のウェブサイトに掲載した。
<http://bmhlinguistics.org/joomla2/language-of-midwifery>
- 「keywords」「n-grams」をクリックすることによって、生データにアクセスできる。
- Keywords : NYT コーパスとの比較において MPC コーパスの中で有意に現れる語
 - Freq. (Frequency) : MPC コーパス中での単語の出現回数
 - RC Freq. : 参照コーパス (Reference corpus : RC) の中での単語の出現回数
 - Keyness : 語彙がどれだけ MPC 分野と深く関連しているかを示す
 - List : 1. The 1,000 most common English words
2. The second thousand most common English words
(Together 1 and 2 = General Service List)
3. The Academic Word List of Coxhead (570 word families)
4. Off-list
- Text : 語彙が有意に出現している論文の数 (312 論文中)
- Journals : 語彙が有意に出現している雑誌の数 (5 誌中)

The screenshot shows the homepage of the BMHLinguistics website. The header includes the logo and name 'BMHLINGUISTICS Centre for Biomedical and Health Linguistics' and a search bar. The navigation menu contains: Home, About BMH Linguistics, Biomedical Language, Learning Resources, TEbMP, Useful Links, Contact, and News. The main content area is titled 'The Language of Midwifery and Perinatal Care' and includes a list of journals: Birth, The Journal of Midwifery and Women's Health, The Journal of Obstetric, Gynecological and Neonatal Nursing, The Journal of Perinatal and Neonatal Nursing, and Midwifery. A login section is visible on the left side of the page.

資料 2 サーチエンジン

- The Language of Midwifery and Perinatal Care のページの「search engine」をクリックすると、以下の画面が立ち上がり、関心のある語彙がどのような文章の中で用いられているかをチェックすることができる。

CorpusSearch - version 4.0

Please enter the search parameters below and press the "Search" button.

(Once you press "Search" please be patient waiting for your results. If no matching records are found then you will receive a message to that effect. Pressing the "Search" button again will not speed things up.)

Which Corpus?	Midwifery
Maximum Sentences to return:	20
Leading/Trailing Characters:	40
Search for:	
	<input type="button" value="Search"/>

Sep., 2008 v1 - first development release, deployed to temporary server
Feb., 2009 v2 - added client code for corpus selection, and feedback on number of rows returned, not deployed.
Jun., 2009 v3 - revised to clean up structure, deployed to bmhlinguistics.org
Sep., 2009 v4 - implemented server side logging and ant, bug fixes

- Which corpus?: Midwifery
- Maximum sentences to return: 1 ページに現れる文の数
- Leading/Trailing Characters: 1 文あたりの文字数の設定
- Search for: 探したい語彙、例えば「maternal mortality」を入力すると下記画面が現れる。

Which Corpus?	Midwifery
Maximum Sentences to return:	20
Leading/Trailing Characters:	40
Search for:	maternal mortality
	<input type="button" value="Search"/>

The risk of [maternal mortality](#) for those with cor pulmonale and deterioration in 2001 (2 137.8; P < .005), whereas [maternal mortality](#) associated with abortion decreased from health seeking. postpartum hemorrhage; [maternal mortality](#); safe motherhood; birth preparedness; health care; direct and indirect estimates show [maternal mortality](#) remains unacceptably high. Evidence from 14 and 15 Regional disparities confirm the role of hemorrhage in [maternal mortality](#) in India. 28 In view of the high estimates of [maternal mortality](#) for rural India, particularly for the more than a decade of efforts to reduce [maternal mortality](#). Recommended as the key strategy to reduce [maternal mortality](#). The medical conditions to which most [maternal mortality](#) is ascribed. Intervention to emergency obstetric care reduced [maternal mortality](#) by 34%. Further questioned, the poststudy decline in [maternal mortality](#) in part of the Matlab comparison group. At the time of the study, the national [maternal mortality](#) ratio was estimated at 120/100,000 live births. In Lam Dong province, where [maternal mortality](#) ratios (212/100,000 live births) were high. The conditions associated with most [maternal mortality](#). Finally, improving essential obstetric key intervention recommended to reduce [maternal mortality](#). It is noted to have one of the highest infant and [maternal mortality](#) rates in the world. The disproportionately high infant and [maternal mortality](#) rates in Afghanistan, the US Department of Health and Human Services. Afghanistan; [maternal mortality](#); infant mortality; pregnant women; health care. It is noted to have one of the highest infant and [maternal mortality](#) rates in the world. It is alarming when compared with infant and [maternal mortality](#) rates in developed nations.

- 管理者は「maternal mortality」をクリックすると原文にリンクできるような設定を試みているが、E-journal の著作権や購読代金の関係もあり、大学などで E-journal を購読していても現時点ではアクセスは無理である。

資料 3 : 頻出動詞 (過去分詞形)

- 出現頻度順の動詞の過去分詞形。受動態になったり、形容詞的用法になったりするため、文章の意味を理解する上で重要な語彙である。特に日本人の学習のために上位 50 語をリストアップした。

Verb: past participle	Freq.	Verb: past participle	Freq.
1 used	812	26 offered	114
2 associated	342	27 determined	109
3 found	308	28 presented	109
4 conducted	278	29 born	107
5 considered	274	30 noted	106
6 given	258	31 taken	105
7 asked	255	32 done	102
8 made	239	33 established	101
9 needed	221	34 measured	100
10 related	219	35 limited	100
11 reported	217	36 recruited	92
12 provided	207	37 encouraged	91
13 based	203	38 recommended	90
14 identified	202	39 informed	90
15 obtained	176	40 diagnosed	89
16 seen	175	41 assessed	87
17 performed	163	42 examined	85
18 described	158	43 supported	84
19 known	156	44 observed	84
20 developed	143	45 compared	82
21 required	142	46 tested	81
22 included	142	47 addressed	79
23 shown	133	48 discussed	79
24 defined	133	49 designed	78
25 collected	131	50 excluded	78

資料 4 : 副詞・形容詞

- キーワードの中には名詞が目立つが、日本人の品詞別学習のため特に副詞・形容詞のキーワードをリストアップした。

Adverb	Freq.	Adjective	Freq.
1 CLEARLY	146	1 POSTPARTUM	1475
2 GENERALLY	138	2 PREGNANT	815
3 FURTHERMORE	134	3 SPECIFIC	498
4 DIRECTLY	108	4 POTENTIAL	431
5 REGULARLY	53	5 VAGINAL	423
6 MAINLY	51	6 ANTENATAL	401
7 RANDOMLY	50	7 EMOTIONAL	355
8 CORRECTLY	49	8 POSTNATAL	349
9 EMOTIONALLY	47	9 AFRICAN	278
10 CONSEQUENTLY	33	10 MENTAL	226
		11 GESTATIONAL	216
		12 DEMOGRAPHIC	211
		13 ACTIVE	199
		14 INTRAPARTUM	179
		15 SPONTANEOUS	179
		16 PSYCHOLOGICAL	173
		17 SUCCESSFUL	158
		18 RURAL	139
		19 RELEVANT	138
		20 FLUID	126
		21 CERTIFIED	118
		22 PSYCHOSOCIAL	117
		23 ACUTE	114
		24 VULNERABLE	111
		25 ONGOING	111