



Title	英語理学療法論文における特徴的 Multi-Word Expression の研究 : Corpus of Contemporary American English Full Text版を参照コーパスとして
Author(s)	八野, 幸子
Citation	大阪大学言語文化学. 2016, 25, p. 17-29
Version Type	VoR
URL	https://hdl.handle.net/11094/77730
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

英語理学療法論文における特徴的 Multi-Word Expression の研究

—Corpus of Contemporary American English Full Text 版を参照コーパスとして—*

八野 幸子**

キーワード：Multi-Word Expression、英語理学療法論文コーパス、Corpus of Contemporary American English

The purpose of this study was to investigate the characteristics of Multi-Word Expressions (MWE) in English physical therapy articles and to determine the inclusion and exclusion criteria for compiling Multi-Word Expression list of physical therapy research articles. In this study, the term MWE stands for the frequently used lexical bundles including compounds, phrasal verbs and idioms.

Various surveys on MWE have been conducted, however only a few surveys on those of physical therapy English have been done. Moreover, although the characteristics of MWE composed of two words were investigated in the preceding study of physical therapy English, the ones composed of more than three words have not been investigated. In the present study, in addition to the survey of 2-word MWE, a survey of MWE composed of three and more words was conducted.

As a measure, corpus based analysis was conducted. The data analyzed were two corpora of English medical articles - Physical Therapy Article Corpus (PT) compiled by the author and, as a reference, the sub-corpus of Corpus of Contemporary American English full-text version (COCA:Academic:Medicine) compiled by Brigham Young University. From these corpora, the frequency lists of N-grams composed of 2 and more words were generated. As a measure for statistical analysis, G-score (G) and document frequency (df) were used.

As a result, the expressions which represent the characteristics of physical therapy filed such as *activities of daily living*, *gait speed*, *spinal cord injury* and academic MWE frequently used in research articles, the ones such as *purpose of this study was to*, were extracted. In addition to these, inclusion and exclusion criteria for compiling Multi-Word Expression list of physical therapy research articles, was determined. The example of the

* Multi-Word Expressions in English Physical Therapy Articles (HACHINO Sachiko)

** 大阪大学大学院言語文化研究科博士後期課程

criteria was as follows “when compiling bigram list, the lexical bundles including *the* such as *the knee* should be excluded from the list”

1 はじめに

1.1 研究背景と Multi-Word Expression 及び単語連鎖の定義

1.1.1 研究背景

教育への応用を目的としたこれまでの英語研究で、複合語や句動詞やイディオム等を含む、複数の語の組み合わせからなる英語表現の習得の重要性は、以前から指摘されてきた。(Biber & Conrad 1999、Biber, Conrad, & Cortes, 2004、Hyland2008、小山 2010 など)

この点で、コーパス分析による理学療法分野の英語研究のうち、日本人学習者を想定し、語の連鎖に焦点を当てたものでは、理学療法論文に生起する 2 語の共起関係について調査した宮本ほか (2012) があるのみである。しかし、後述の Multi-Word Expression の先行研究の多くでは 3 語以上の表現の調査が行われ、その教育への応用の有用性が報告されているおり、理学療法分野の英語研究でも、3 語以上の表現を研究することは意義があるといえる。

また、理学療法分野の特徴語について調査をした八野 (2015) では、現役理学療法士による抽出語彙の妥当性と信頼性の検証を行ったが、特徴語を含む表現のうち、2 語の表現で宮本ほか (2012) では調査されなかった “gait speed” などの「名詞」+「名詞」の組み合わせも当該分野で有用とされたことから、2 語の表現についても宮本ほか (2012) の枠を超えた調査が必要であると考えた。

これらを踏まえ、筆者は理学療法分野の論文に特徴的に出現する単語連鎖のリスト作成を最終的な目標とし研究をしているが、本研究ではその基礎資料作成として、これまでもコロケーション抽出や特徴語抽出で用いられてきた G-score と、文書頻度¹で単語連鎖の抽出を行い、理学療法以外の医療系分野の論文との比較で、理学療法論文に特徴的に用いられる表現を検討し、最終的なリスト作成ではどのような点に留意すべきかを実際に抽出された表現の観察により検討する。

本研究で抽出を目指す表現は、理学療法分野の論文で特徴的に出現する論文表現、また一般的な文章にも現れるが、理学療法分野の研究の記述に必要と考えられるものを中心とする。これらは、英語の理学療法論文の読解や執筆に役立つものを想定している。調査対象の連鎖する語数は、2 語、3 語、4 語、5 語を議論の中心とするが、抽出可能な

¹ 文書頻度とは、コーパスを構成する複数の文書のうち、いくつの文書に渡って注目する表現が出現するか、を表す。

単語連鎖の最大値まで抽出し、リスト作成の留意点の検討に用いる。最終的なリストに含む表現は小山・水本（2010）の“freeze dry”などのように各語数の連鎖で意味が完結するものや、連鎖の後に続く形がおおむね決まっているもので、Liu（2010）を例に挙げると“in order to VP”のように3語の連鎖“in order to”の後に通常“VP”がくることが明白なものなどを想定している。

本研究で理学療法以外の医療分野の論文を参照コーパスとするのは、筆者のこれまでの研究発表で、一般的な学術文章と比較し、理学療法論文で特徴的に用いられる表現について報告したが、理学療法論文に特徴的に出現する表現の中に、医療系英語全体についても特徴的なものがあるのではないかと指摘があったためである。このような指摘はあったが、後述のLiu（2012）ほか、Multi-Word Expressionの先行研究では、“Lexical Bundles were often discipline-specific”とされ、他の医療系分野の論文にも出現するものでも、理学療法分野の論文と他の医療系分野の論文とで使用傾向が異なると考えられる。このような背景から、本研究では他分野と共通する表現でも、理学療法分野の論文に特徴的に出現する表現の抽出を試みる。

1.1.2 Multi-Word Expression と単語連鎖

本稿で用いる「単語連鎖」とはbigram、3-gram、4-gramなど、すべてのn-gramを総称した表現であり、「Multi-Word Expression（以下MWE）」は、単語連鎖のうち、小山・水本（2010）で述べられている「通常平均的に観察されるよりも多く出現する語の境界を越えた一連の語句」を意味する。この場合MWEは、小山・水本（2010）で例示されている、複合語（crystal ball, freeze dryなど）や句動詞（give up, put offなど）やidiom（kick the bucket, rain cats and dogsなど）も含む。また、MWEはMulti-Word constructions、Multi-Word Phrases、Multi-Word Unitsとも呼ばれている。（Liu 2012）

1.2 先行研究とリサーチ・クエスション

1.2.1 コーパスを用いた理学療法英語研究

上記のように理学療法分野の英語に関する研究は少ないが、コーパス分析によるものでは次のものが挙げられる。宮本ほか（2007）では理学療法論文コーパスを分析し、後述の宮本ほか（2011）の基礎となる300語の教育語彙選定が行われた。Mitsuda（2009）では一般英語から専門英語への橋渡しをする読解教材の開発が行われ、教材用語彙選定で、理学療法論文コーパス分析が行われた。頻出語彙はWest（1953）のGeneral Service List（GSL）、Coxhead（2000）のNew Academic Word List（AWL）、中村（2006）の中学校英語教科書コーパス、山添（2006）の高等学校英語教科書コーパスの高頻度語

彙と比較され、語彙表が作成された。宮本ほか（2011）では、宮本ほか（2007）で編纂された理学療法論文コーパスの再編纂と理学療法教科書コーパス編纂が行われ、抽出された特徴語と GSL、AWL が比較がされた。本研究と関連の深い宮本ほか（2012）では、宮本ほか（2011）のコーパスで、共起パターンの特性が分析された。`severely disable（先行研究表記通りに記載）`のような 2 語の共起表現の抽出が行われた。この研究では「副詞 + 動詞」、「動詞 + 名詞」、「名詞 + 形容詞」の組み合わせに限定して 2 語の共起パターンの特性が報告されているが、3 語以上の単語連鎖は分析されていない。また、他の医療分野との比較も行われていない。八野（2015）では、近年の言語研究で著者同定や、論文の分類等に用いられる Random Forests で、理学療法以外の医療系分野との比較による英語理学療法論文の特徴語抽出が行われた。

1. 2. 2 Multi-Word Expression に関する研究

MWE の研究では Moon（1998）が初期の研究のうちで最も代表的なものの 1 つとされており、そのほか代表的なものでは、Biber ほか（1999）、Cortes（2004）、Biber & Barbieri（2007）、Hyland（2008）がある。歴史学と生物学の論文コーパスと当該分野の学生の学術的な英作文の MWE を比較した Cortes（2004）や、4 つの異なる専門分野の研究論文、博士論文、修士論文の MWE を比較した Hyland（2008）では、“Lexical Bundles were often discipline-specific”とされており、本研究の最終目標である、理学療法分野の英語に特化した MWE リストの作成は意義があると考えられる。Liu（2012）では British National Corpus と Corpus of Contemporary American English の学術文章のサブコーパスをデータとして、228 の高頻度 Academic Multi-Word Constructions の抽出が行われたが、この研究でも上記の Cortes（2004）、Hyland（2008）で報告された、“Lexical Bundles were often discipline-specific”という点が支持される結果が報告された。

日本人学習者を想定した ESP コーパスによる MWE 研究では、小山（2008）で、B score と MI score を用いて科学技術コーパスを分析し、bigram が検出され、その結果から MWE を確定するために使うべき指標と使用の条件について論じている。小山・水本（2010）では、6 種類のコーパスで共通する単語連鎖を比較し、分野の違いが明らかになるかについて、コレスポネンス分析による調査が行われた。小山・田中（2010）では、細分化された専門に入る前段階の学習者が習得すべき科学技術全般に関わる表現が、分類器を用いた手法で科学専門誌 Nature の論文から抽出された。田中・小山（2010）では、構文情報を考慮した ESP コーパスからの特徴語抽出が行われた。

医学、医療系分野のコーパスを用いた近年の研究では Yokoyama（2012）が挙げられ、“medical surgery”、“nursing”、“public health”の 3 分野の論文コーパスにおける、4 語と

5 語の単語連鎖が分析された。Fujieda (2015) は放射線技師養成大学の学生の英語教育への応用を目的とした研究で、当該分野の教科書や論文等の複数のコーパスを用いて、2 語、3 語、4 語の表現の抽出が行われ、分野に特化した名詞句の bigram などが抽出された。

1. 2. 3 本研究の目的と Research Question

上記の研究背景のもとに、本研究では Research Question を以下のように定める。

RQ1：他の医療系分野の論文との比較において、とりわけ理学療法分野の論文で特徴的に用いられる Multi-Word Expression とはどのようなものであるか。

RQ2：最終的な MWE リストの作成において、どのような点に留意すべきか。

2 方法

2.1 使用データ

2.1.1 英語理学療法論文コーパス

分析対象のコーパスに、著者編纂の英語理学療法論文コーパス（非公開）を用いた。本コーパスは宮本ほか（2007）、宮本ほか（2011）、宮本ほか（2012）と同様に、アメリカ物理医学リハビリテーションアカデミー発行の *Archives of Physical Medicine and Rehabilitation* とアメリカ理学療法士協会発行の *Physical Therapy* の 2 誌より、論文のみを抽出しコーパスを構築した。これら 2 誌を選択したのは、宮本ほか（2011）が示すように、理学療法分野のリーディングジャーナルであるためである。本コーパスは、総語数 1,364,005 語、異なり語数 24,437 語である。語数は R² で出した。以下このデータの表記は PT とする。

2.1.2 Corpus of Contemporary American English Full-Text 版 Academic Medicine

参照コーパスは、Brigham Young 大学の Mark Davis 氏の研究グループが構築した現代アメリカ英語の大規模汎用コーパス Corpus of Contemporary American English (COCA) の Full Text 版の Academic Medicine を用いた。本研究でこの資料を用いたのは、次のような理由からである。4 つの医学・医療系分野の論文のコーパスに出現する単語連鎖を比較した横山（2010）の例では likely を用いる表現では、公衆衛生分野（以下 PH）では likely に more や less を付加して主張をさらに弱める控えめな表現が確認され、

² 統計用言語とその実行環境を兼ね備えたフリーソフトウェア。Ver.3.2.0 を用いた。

総じて PH では「評言」の使用に関しては、ゲノム生命医学分野や臨床外科学分野に比べて、研究・調査結果の解釈に、より慎重な表現を選択する傾向がみられたと報告されており、このような傾向を「各々のディスコース・コミュニティのアカデミックライティングにおける言語的流儀、すなわちジャンル特徴であると理解することができる」としている。このような例から、似通った医学・医療系分野の論文を参照コーパスとすることにより、理学療法分野というディスコース・コミュニティのアカデミックライティングにおける言語的流儀が見えてくるのではないかと考え、それらは、今後、理学療法分野の論文読解、執筆を目指す学習者にとって役立つものとなるのではないかと考えたためである。

このデータから理学療法分野以外の論文のみを抽出し、conclusion より後の部分を削除した。これはこの部分に、reference や URL などの不必要な情報が多く、それらを取り除かず行ったパイロット抽出で、上位の表現の多くが数値の連続や、URL など、抽出を目指すものではなかったためである。また、PT は reference 等の情報は含んでいないためである。これらの処理後の総語数は 1,753,922 語、異なり語数 39,682 語である。以下このデータに関する表記は CM とする。

2.2 分析方法

R で Gries (2009) の方法をもとに word list を作成した³ 後、そこから、N-gram (2 語、3 語、4 語、5 語等の単語の連鎖) を抽出し、G-score で p 値が 0.01 より小さく、文書頻度が 4 以上のものを選んだ。この際、数字のみの組み合わせ、数字または数を表す語と and, or, but, to などの接続詞や前置詞のみの組み合わせは考察の対象外とした。また、2 語に関しては数字や数を表す語以外の語と and, but などの接続詞との組み合わせ、冠詞と数字または数を表す語の組み合わせも削除した。2 語でこの削除をしたのは、これらの連鎖は 2 語で意味が完結せず、後に来るものがおおむね特定できないためである。文書頻度を 4 とした理由は Cortes (2004) などでは、文書頻度 5 以上の MWE が抽出されているが、本研究の資料で文書頻度 5 と 4 を比較すると、文書頻度 5 では muscle strength and power のような理学療法分野で有用と考えられる表現のうちいくつかは抽出できず、文書 4 では抽出され、本研究では、Cortes (2004) などでも用いられる文書頻度に限りなく近いが、文書頻度 1 のように偶然的出現ではない、文書頻度 4 を採用した。

³ この方法では、大文字が小文字に変換され、文字と数字以外の記号、たとえばピリオドなどが削除される。詳細は pp.106-110 を参照されたい。なお、N-gram の抽出はコーパスを構成する複数の文章それぞれに対して行った。つまり、ファイルをまたいでの表現の抽出は行われていない。

3 結果と考察

本研究では bigram から 16-gram までの N-gram を求めた。最長の 16-gram では “available as a supplement to the online version of this article at <http://www.ajph.org>” が抽出され、15-gram から 9-gram まではこの表現の一部分の連鎖または意味をなさない数値の連鎖であった。そのため表 1 に 16-gram および 8-gram から bigram までについて、抽出された N-gram 数を示す。8-gram から 6-gram にも上記の 16-gram の表現の断片が複数抽出されているが、その他の表現も抽出されているため、表 1 に示す。

表 1 抽出された N-gram 数

	PT	CM		PT	CM
bigram	3964	2990	6-gram	24	20
3-gram	1690	801	7-gram	7	15
4-gram	369	123	8-gram	1	12
5-gram	83	39	16-gram	0	1

N-gram のリストは紙幅の都合上 bigram から 5-gram の上位の 20 項目のみを示す。また (PT) は PT の、(CM) は CM の表現を表す。

最長の 16-gram では “available as a supplement to the online version of this article at <http://www.ajph.org>” (CM) が抽出されたが、これは論文の追加資料の入手先に関するものである。8-gram では “to our knowledge this is the first study” (PT) や “the centers for disease control and prevention cdc” (CM) などが抽出されたが、コンコーダンスを確認したところ、“to our knowledge, this is the first study” のように 3-gram と 5-gram の組み合わせや、“the centers for disease control and prevention cdc” (CM)、“the centers for disease control and prevention (cdc)” のように略語を含んだ表現であったため、これらはリスト作成時には () 内の略語を含まないより N の小さい表現の中に含めることとした。7-gram で抽出された表現のうち 2 つは “the aim of this study was to” (PT) のような研究の目的を示す論文表現であった。その他 “the centers for disease control and prevention” (CM) などの施設名と “the extended icf core set for stroke” (PT) のような国際生活機能分類を表す表現であった。これらの表現については固有名詞であるため、表現リストから除外する。6-gram では “approved by the institutional review board”、“was approved by the institutional review” のように同じ 6-gram 内で重複するものが見られた。この表現のコンコーダンスを確認したところ、前者の表現が後者のように受動態の一部として用いられているだけでなく、直前の名詞の後置修飾に用いられるものも

あった。このような表現は、前者をリストに加え、受動態の一部としても用いられることを、注などで記すこととした。リスト上の表記としては“be/ 名詞 +approved by the institutional review board”などを想定している。

これまでに見てきた 16 語から、6 語の連鎖について総合的に判断すると、16-gram での“available as a supplement to the online version of this article at [http www ajph org](http://www.ajph.org)”は論文の追加情報の入手先であり、8-gram で抽出された“the centers for disease control and prevention cdc”のような施設名を表す固有名詞や、7 語の“the centers for disease control and prevention” (CM) のような施設名、“the extended icf core set for stroke” (PT) のような国際生活機能分類を表す表現などは、施設名や評価指標等をあらわす固有名詞である。これらも論文の読解や執筆に役立つ表現となる可能性はあるが、論文の読解力や、執筆力を高める表現であるとは考えにくい。本研究は、教育への応用を目的とした語彙表作成を目的としているため、これらの表現は表現リストから除外する。

5 語の表現について、表 2 を概観すると、PT では論文表現や固有名詞が目立っていた 6 語までの連鎖と比較して、“at the time of injury”や“at the end of treatment”のよう

表 2 5-gram の上位 20 項目

PT	CM
1 of this study was to	for disease control and prevention
2 at the time of injury	centers for disease control and
3 results of the present study	u s environmental protection agency
4 item short form health survey	disease control and prevention cdc
5 the results of the present	the centers for disease control
6 short form health survey sf	protection agency u s epa
7 test retest reliability of the	environmental protection agency u s
8 classification of functioning disability and	s environmental protection agency u
9 international classification of	ref 1 00 1 00
functioning disability	
10 medical outcomes study 36 item	the u s environmental protection
11 aim of this study was	at http www ajph org
12 purpose of this study was	in the past 30 days
13 at the end of treatment	article at http www ajph
14 in the present study the	of this article at http
15 36 item short form health	online version of this article
16 compared with the control group	the online version of this
17 used in the present study	this article at http www
18 were excluded if they had	version of this article at
19 the 6 month follow up	to the online version of
20 this study was to compare	a supplement to the online

な理学療法分野と関連が深いと考えられる表現が、顕著に出現している。CM では 5-gram でもより語数の多い N-gram で抽出された表現と重複するものがまだ多く見られる。

4-gram の表現では CM ではここでも施設名などの表現が多く含まれているが、PT では論文表現に加えて、“activities of daily living” などの理学療法らしさの現れる表現が抽出されている。この表現は本研究に先立って行った文書頻度 5 による抽出と文書頻度 4 による抽出で、文書頻度 5 の場合でも PT の上位表現（2 位）に抽出された表現である。また PT では “between the 2 groups”、CM では “between the two groups” と表現されているところも、同じ医療系論文でありながら、表記に異なりがあるという点で興味深い。

3 語の表現では、PT では “people with sci” のような「人 +with+ 傷病名」の構造がみられ、4 語ではこれらを含む表現、2 語ではこれらの一部であると推測可能な表現が抽出されており、実際にコンコーダンスを確認したところ、3 語の表現にそれらの完全な形が多く含まれていることがわかった。また、4 語の表現で「人 +with+ 傷病名」の構造が PT、CM の表現にどれくらい含まれるかを確認したところ、PT では 11 表現、CM では 1 表現が確認され、医療系文脈で用いられる可能性の高い表現でありながら、PT

表 3 4-gram の上位 20 項目

PT	CM
1 in the present study	in the united states
2 activities of daily living	the study reported here
3 the subjects in the	centers for disease control
4 between the 2 groups	disease control and prevention
5 this study was to	for disease control and
6 of this study was	s environmental protection agency
7 of the present study	u s environmental protection
8 subjects were instructed to	the centers for disease
9 subjects were asked to	control and prevention cdc
10 p less than 05	between the two groups
11 in people with pd	in the placebo group
12 the se of measurement	agency u s epa
13 muscle strength and power	protection agency u s
14 in the current study	environmental protection agency u
15 in persons with sci	of students with ebd
16 has been shown to	ref 1 00 1
17 spinal cord injury sci	the u s environmental
18 participants were instructed to	percent of the respondents
19 1 year after discharge	statistically significant p 0
20 of the sf 36	u s department of

では「人+with+傷病名」の構造が好んで用いられることが示唆された。

2語の連鎖に関して、PTでは“walking speed”、“gait speed”などの専門分野をよくあらわす複合語が顕著に抽出されている。なお、抽出方法や、参照コーパスが医療系論文でないことなどがあげられるため、参考程度にとどめるが、Fujieda (2015)ではx ray, breast cancerなどが抽出されており、2語の連鎖には分野ごとに大きな異なりがありそうである。リスト作成の留意点では、PTの表現の“the knee”のように、2語で役立つ表現とは呼べないものも多く含まれていることがわかった。これらも表現リストから除く必要があると考えられる。ただし、theを1語目に持つ2語の表現では、本論の上位の表現には出現していないが、the+序数などは2語の表現と考えられるため、リスト作成に当たってはこれらを区別しなくてはならない。

全体的に見て、Cortes (2004)、Hyland (2008)、Liu (2012)などが報告したようにMWEはDiscipline specificであることが本研究においても示唆され、さらにそのような表現は、bigramから5-gramまでの上位表現に顕著に見られることも唆された。

表4 3-gramの上位20項目

	PT	CM
1	in people with	u s e pa
2	subjects in the	percent of the
3	the present study	students with ebd
4	for people with	p lt 001
5	compared with the	the united states
6	between the 2	p 0 05
7	people with sci	et al 2000
8	in this study	and co authors
9	of people with	p 0 001
10	in persons with	et al 2008
11	children with cp	et al 2009
12	in the present	can not be
13	in subjects with	et al 2007
14	strength and power	between the two
15	persons with sci	blood lead levels
16	test retest reliability	in the united
17	for subjects with	et al 2003
18	the 2 groups	the two groups
19	the current study	et al 2004
20	acquired brain injury	at least one

表5 bigramの上位20項目

	PT	CM
1	people with	et al
2	subjects with	e g
3	with sci	p 0
4	physical therapy	public health
5	physical therapists	i e
6	subjects were	u s
7	the subjects	environmental health
8	walking speed	lt 001
9	in people	percent of
10	subjects in	african American
11	after stroke	p lt
12	p 001	exposure to
13	persons with	students with
14	upper limb	drinking water
15	physical therapist	e coli
16	gait speed	food safety
17	the knee	with ebd
18	muscle strength	samples were
19	compared with	s e pa
20	all subjects	the school

またこの MWE から見られる理学療法論文における構文的特徴では、with を用いた後置修飾、とりわけ「人+with+ 傷病名」の構造が他の医療系分野との比較においても PT の特徴として顕著に見られ、これらは理学療法分野の英語の学習者にとって習得または復習すべき項目であることが示唆された。リスト作成の留意点として、N-gram による抽出では“approved by the institutional review board”、“was approved by the institutional review”のような部分的な重複が見られるため、それらの集約のために前後の N-gram の参照や、コンコーダンスの確認が有効であることが示唆された。

4 おわりに

本研究では、英語理学療法論文の特徴的 Multi-Word Expression について、「RQ1: 他の医療系分野の論文との比較において、とりわけ理学療法分野の論文で特徴的に用いられる Multi-word Expression とはどのようなものであるか」「RQ2: 最終的な MWE リストの作成において、どのような点に留意すべきか。」という点に関して調査と検討を行った。RQ1 に対しては理学療法分野で有用であるとされる“activities of daily living”のような表現が抽出され、「人+with+ 傷病名」の構造は理学療法分野ではとりわけ顕著に用いられることなどが明らかとなった。さらに、上位の表現には論文表現のみならず、分野をよくあらわす表現が多く含まれ、Cortes (2004)、Hyland (2008)、Liu (2012) などが報告したように MWE は理学療法英語においてもこの結果を支持するものとなった。また、RQ2 では、重複の取り扱いが課題となり、最終的にリスト化するには、コンコーダンス及び前後の N-gram の参照により、より有用なものをリストに含むべきであることがわかった。しかしながら、この方法は手動による方法であるため、今後は更なる文献調査も行い、よりよい重複の集約方法について考えてゆきたい。また 2 語の単語連鎖で“the knee”のような表現とは呼べない「the+ 語」の構造が抽出されていることから、注意は必要であるが、これらをリストから除くべきであるという点、その他上記の考察では触れていないが、“compared with the”のような表現は、実際の教育での利用では、Liu (2012) のように“compared with (the) + 名詞”のような形で示す必要があることが分かった。

今後の課題は以下のようなものである。本研究は最終的なリスト作成の基礎資料の作成であり、複合語、イディオム、句動詞、単語連鎖などの分類は行わないが、上記のリストを概観すると現段階では単語連鎖が多く、次いで複合語が顕著に見られる。最終的なリストでは、このような分類も視野に入れたい。また「the+ 語」などを削除したリストを作成し、理学療法士による妥当性と信頼性の検証を行いたい。

参考文献

- 小山由紀江・田中省作 (2010) 「科学技術英語の特徴表現：科学専門誌 *Nature* の論文を対象とした分析『ESP コーパスからの特徴表現抽出』統計数理研究所共同研究レポート 239 統計数理研究所 pp.31-42
- 小山由紀江・水本篤 (2010) 「単語連鎖にみる、科学技術分野と他分野の英語表現の比較」『ESP コーパスからの特徴表現抽出』, 統計数理研究所共同研究レポート 239 統計数理研究所 pp.1-12
- 小山由紀江・田中省作 (2010) 「構文情報を考慮した ESP コーパスからの特徴表現の抽出」『ESP コーパスからの特徴表現抽出』統計数理研究所共同研究レポート 239 統計数理研究所 pp.13-30
- 中村純作 (2006) 「教科書コーパスから何が見えるか：方法論と中学校英語教科書の場合」『立命館言語文化研究』17 (4) 立命館大学言語文化研究所 pp.143-166
- 八野幸子 (2014) 「コーパス分析による英語理学療法論文と一般学術文章における慣用語句使用傾向の比較」 大阪大学言語文化学会 第 45 回大会発表
- 八野幸子 (2015) 「Random Forests による英語理学療法論文からの特徴語抽出 - Corpus of Contemporary American English Full-Text 版を参照コーパスとして」 *Asphodel* 50 号 pp.97-114 同志社女子大学英語英文学会
- 宮本祥子・宮本謙三・宅間豊井上佳和・竹林秀晃・岡部孝生・滝本幸治 (2007) 「理学療法教育における英語文献読解のための教育語彙選定：独自のコーパス分析を通して」『理学療法学』34 (6) pp.260-266
- 宮本祥子・五百蔵高浩・宮本謙三・宅間豊・井上佳和・竹林秀晃・岡部孝生・滝本幸治 (2011) 「理学療法分野における英語専門語彙 (ESP 語彙) の抽出とその特性」『理学療法学』38 (6) pp.421-435
- 宮本祥子・五百蔵高浩・宮本謙三・宅間豊・井上佳和・竹林秀晃・岡部孝生・滝本幸治 (2012) 「理学療法分野における英語表現－共起パターンの特性」『理学療法学』39 (2) pp.90-101
- 山添孝夫 (2006) 「教科書コーパスから何が見えるか：高等学校英語教科書の場合」『立命館言語文化研究』17 (4) 立命館大学言語文化研究所 pp.167-189
- 横山彰三 (2010) 「ESP の理論と実践」大学英語教育学会 (監)『英語教育学大系第 4 巻 21 世紀の ESP - 新しい ESP 理論の構築と実践』第 2 章 pp.17-43 大修館書店
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university and written registers, *English for Specific Purposes*, 26, 263-286. doi:10.1016/j.esp.2006.08.003

- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 397-423. doi:10.1016/j.esp.2003.12.001
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238. doi:10.2307/3587951
- Fujieda, M. (2015) Multi-word units in specialized corpora: What pedagogical information can they give us? *Extraction of the Domain Specific Expressions from ESP Corpora and Their Pedagogical Applications The Institute of Statistical Mathematics Cooperative Research Report* 338, 59-58
- Gries, S.T. (2009) *Quantitative Corpus Linguistics with R: A Practical Introduction* New York :Routledge.
- Hyland, K. (2008). Academic clusters: text patterning in published and post graduate writing. *International Journal of Applied Linguistics*, 18, 41-62 doi:10.1111/j.1473-4192.2008.00178.x
- Liu, D. (2012) The most frequently-used multi-word constructions in academic written English: A multi-corpus study. *English for Specific Purposes* 31, 25-35 doi:10.1016/j.esp.2011.07.002
- Mitsuda, S. (2009). Developing ESP reading material for physical therapist. *On Language and Language Education*, 2, 71-98
- West, M. (1953). *A General Service List of English Words*, London : Longman Green & Co.
- Yokoyama, S. (2012) A Cross-sectional analysis of lexical bundles: “stance” in written medical discourse. *Journal of medical English education*, 11 (3), 59-67
- Corpus of Contemporary American English Full-Text (2015 年 9 月 25 日閲覧)
<http://corpus.byu.edu/full-text>