



英語科学技術論文と理系大学院生の英語論文における連結語と単語連鎖の比較

小屋, 多恵子

(Citation)

Journal of Corpus-based Lexicology Studies, 3:15-30

(Issue Date)

2021-02-28

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81012538>

(URL)

<https://hdl.handle.net/20.500.14094/81012538>



英語科学技術論文と理系大学院生の英語論文における

連結語と単語連鎖の比較

小屋 多恵子(法政大学)

tkoya@hosei.ac.jp

A Comparison of Transitions and Lexical Bundles Extracted from Science Journal Corpora and a Corpus of Papers by Japanese Science and Engineering Graduates

KOYA, Taeko (Hosei University)

概要

本研究は、理系大学院生が英語で執筆した論文に見られる連結語 (transitions) と単語連鎖 (lexical bundles) の特徴を、*Nature*, *Science*, *Cell* に掲載されている論文における使用と比較することによって明らかにすることを目的とする。それぞれの論文をコーパス化し、出現する連結語と単語連鎖を対象に相関分析、 χ^2 二乗検定、コレスポンデンス分析を実施した。その結果、理系大学院生は連結語の使用が限定的であることが判明した。単語連鎖は、it is 形容詞 + to 不定詞構文、in this section, in this paper といったこれから行う研究や分析について宣言を表明する表現や、as described in materials, shown in table といった他の部分の情報を示す表現の多用が認められた。

キーワード

連結語, 単語連鎖, ESP educations

1. はじめに

英語科学技術論文では、正確に簡潔に誤解を与えずに内容を読み手に伝えることが重要である。Wallwork (2016) によると、英語の 1 文は時代と共に短くなる傾向にあり、ダーウィンが *On the Origin of Species* を出版した 1859 年では、英語を母語とする科学者の書く 1 文は 70 語以上であったのに対し、現在推奨される 1 文は 15 から 18 語程度である (p. 57)。一方、8 語文程度であればおよそ 90% の読み手が 1 読で内容を理解できるが、27 語文であると約 4% の読み手しか理解できないと言われている (p. 57)。読み手に誤解を与えない文を書くことを求められる英語科学技術論文は、専門的な内容はもとより、短い文で正確に書く書き方も求められることになる。

理系大学院生は、大学院に入ると英語による発表と論文執筆が必要である。仮説をたて、実験を重ねて新しい知見を発見し、それを英語で論文にまとめて発信する。世界中の科学者と成果を共有し、それを基にさらなる研究の成果を積み上げていく。英語で論文を書き始める大学院生は、正確で簡潔な文を書くことができず冗長な文を書いている、論文を投稿しても読んでももらえないことになりかねない。英語論文執筆初心者である大学院生に対して、大学の英語教師はどのように効果的なサポートをすることができるだろうか。

本研究は、正確で簡潔な英文作成のための効果的な指導を検討するために、理系大学院生執筆の英語論文にみられる定型表現を、英語科学雑誌論文と比較することによって特定することを目的とする。定型表現の中でも、連結語と単語連鎖に注目する。連結語と単語連鎖は、自然で論理的な英文構成に重要な役割を果たす。一流の英語科学雑誌論文の表現法と比較することにより、大学院生の書き方と目指すべき書き方との類似点・相違点が明確になり、効果的な指導につなげることができる考える。

2. 先行研究

2.1 連結語と単語連鎖

連結語とは、書き手の考えに論理的枠組みを与え、読み手に順序立てて読み進んでいけるために概念と概念との間の関係を明確にする語(句)のことであり、つなぎ言葉や接続副詞、メタ談話標識とも称される(小林 2009, 2010, 三木 2016, 日本物理学会 1999, 篠田 2007)。書き手が、これまでどのようなことを述べ、次にどのようなことを述べるのかを論理的に提示することによって、読み手に正確に書き手の意図を読み取れるシグナルを送る役割を果たす。このように、英語で科学技術論文を効果的に書くためには、連結語が不可欠であり、自然な使用を習得する必要がある。以下表1は代表的な連結語であり、大学生用のパラグラフ・ライティングテキストに提示されているものである(Zemach & Rumisek 2011)。

表 1

代表的な連結語

chronology	before, after, next, since, first, second, while, when
comparison	likewise, compared to, similarly, as...as, and
contrast	however, on the other hand, but, yet, in spite of, in contrast, although, instead
additional information	and, also, in addition, in fact, furthermore, moreover, another...is/was
examples	for example, in general, generally, for instance, specifically, in particular
cause and effect	therefore, so, thus, as a result, since, because
concluding ideas	in conclusion, in summary, finally, therefore, to conclude, to summarize

単語連鎖は、2語3語4語と連続する語の境界を越えた一連の語句を指し、ある分野のテキスト中に繰り返し出現し、英語使用が自然であるかを定める鍵といふべき言語表現である(Chen & Baker 2010, Hyland 2008a, 2008b, 小山・水本 2010)。専門分野における特徴的な英語表現を表し、書きことば、話しことばといったディスコースやレジスターを示す重要な役割を担っていると考えられている(Biber & Barbieri, 2007; Biber, Conrad, & Cortes, 2004; Cortes, 2004; Hyland, 2008b)。単語連鎖の中でも、4-gram が特徴的な英語表現を抽出するのに適していると(Cortes, 2004; Hyland 2008a)、分析を行う場合が多い。連結語も単語連鎖の一部であり、*in addition, in summary, as a result, in spite of, on the other hand* はその例である。

2.2 単語連鎖に関する実証研究

単語連鎖は、大規模な言語データに基づいて客観的に分析できるようになった 2000 年代以降、実証的な研究が行われている。中でも、英語熟達者が使用する単語連鎖と英語初級者が使用する単語連鎖を比較分析した研究として、Cortes (2002, 2004) では、*expert writers* と *novice writers* では使用する単語連鎖が異なること、たとえ同じ単語連鎖を使用していたとしても使用頻度が異なる結果を報告している。Hyland (2008a, b) は、査読により受理された論文と大学院生の論文における単語連鎖の使用を比較したところ、大学院生の方が定型表現を多く使用していたことが認められ、定型表現を使用することによって表現力が高いことを示そうとしているのではないかと結論づけている。日本人英語学習者に対する単語連鎖使用では、小林 (2009, 2010) により学習者コーパスと母語話者コーパスを識別する特徴を抽出し、*I think* のようなコロケーションや *firstly, secondly* のような *sequencing* の過剰使用が判明している。三木 (2016) では、NICE を利用した習熟度別日本人英語学習者のメタ談話標識分析から、人称代名詞や *sequencing* の使用頻度が高いことを示している。小屋 (近刊) では、*Nature* の論文と大学院生の書いた論文を単語連鎖の観点から比較したところ、使用パターンが異なっていることを報告している。

さまざまな分野のコーパスから単語連鎖を抽出する研究としては、小山・水本 (2010) が 6 種類のコーパスで共通される単語連鎖を分析することにより、分野の違いを特定している。また、水本は、英語学術論文支援オンラインツール AWSuM を 2016 年より無料公開し、特定分野における単語連鎖を提案することによる支援を行っている。八野 (2017) では、英語理学療法論文における *Multi-Word Expression* の使用傾向を分析・考察している。

3. リサーチデザインと手法

3.1 目的

本研究は、理系大学院生執筆の英語論文にみられる連結語と単語連鎖を、英語科学雑誌論文における使用と比較することを目的とする。小屋 (近刊) を発展させ、*Nature* を含めた3つの英語科学雑誌コーパスを作成し、単語連鎖と連結語を対象として分析を試みることにする。具体的なリサーチ・クエスチョンは以下のとおりである。

RQ1: 日本人理系大学院生執筆の論文は、英語科学雑誌と比較するとどのような特徴的連結語を用いるのか。

RQ2: 日本人理系大学院生執筆の論文は、英語科学雑誌と比較するとどのような特徴的単語連鎖を用いるのか。

連結語も単語連鎖の一部ではあるが、自然な文章作成に連結語が重要であることから、連結語と単語連鎖と分けて分析を行うことにする。

3.2 対象論文

理系大学院在学中に執筆した単著論文 22 編をコーパス化した。内訳は、2007 年から 2017 年の紀要 13 編、修士論文 4 編、博士論文 5 編である。執筆者はすべて日本人大学院生とし、外国人留学生は英語使用環境や学習歴が不明のため、今回の研究対象から除外した。理系大学院生は、大学院時代から国際ジャーナルに投稿し採択されることもあるが、多くは指導教授や他の研究者との共著の場合が多く、執筆者の特定が困難なことから、対象外とした。

英語科学技術雑誌は、*Nature*, *Science*, *Cell*とした。*Nature* は、2017 年 2018 年 2019 年 1 月 5 月 9 月から各 4 編を抽出し合計 36 編、*Science* は、2017 年 2 月 6 月 10 月、2018 年 4 月 8 月 12 月、2019 年 3 月 7 月 11 月、2020 年 1 月から各 3 編を抽出し合計 30 編、*Cell* は、2016 年 4 月 8 月 10 月、2017 年 1 月 6 月 11 月、2018 年 2 月 7 月 12 月から各 3 編を抽出し合計 27 編をコーパス化した。この 3 誌は、世界的に権威がある科学総合誌であることから三大科学誌と呼ばれ、研究者から科学に興味がある人々まで幅広い読者がいる。2019 年トムソン・ロイターが提供する Journal Citation Reports (JCR) にて提供されたインパクトファクターは、*Nature*: 43.070, *Science*: 41.037, *Cell*: 36.216 を示し、影響力の強さと多くの読者に支持されていることを裏付けている。また、筆者の指導対象である理系大学院生の研究領域が、機械・電気・電子・情報・システム・生命機能・環境と幅広いため、科学総合誌、生物総合誌であるこの 3 誌を選択した。

理系大学院生執筆論文コーパス(以下、Student コーパス)、*Nature* コーパス、*Science* コーパス、*Cell* コーパスの基本情報は表 2 の通りである。

表 2

各コーパスの基本情報

	Student コーパス	<i>Nature</i> コーパス	<i>Science</i> コーパス	<i>Cell</i> コーパス
Total No. of Word Types	13,270	13,429	10,591	13,079
Total No. of Word Tokens	208,655	222,073	139,576	226,718
Type/Token Ratio	6.36	6.05	7.59	5.77

3.3 対象とした連結語と単語連鎖

連結語は、英語科学技術論文でよく使用される 84 語(句)を選定した。選定にあたっては、Hyland (2005)、篠田 (2001, 2007)、Zemach & Rumisek (2011)を参考にした。特に、大学生

用テキストである Zemach & Rumisek (2011) をもとに、大学生の時に学ぶ可能性が高い連結語を中心に選抜した。単語連鎖の選定は、対象の 4 つのコーパスサイズが異なるため、まず 10 万語になるように調整頻度を算出した。そこから、4-gram を中心とした合計頻度の上位 100 位を選定し、さらにトピックに偏らない単語連鎖のみを選定し、40 語句を対象とした。

表 3

対象とした 84 の連結語

additional information	above all, additionally, as for/to..., as regards..., besides, more/ most importantly, on top of that, regarding..., what is more
cause and effect	accordingly, as a result, because, consequently, for this reason, hence, so, therefore, thus
chronology	after, before, first, firstly, lastly, next, second, when, while, secondly
comparison and contrast	also, alternatively, although, at the same time, but, by the same token, compared to, conversely, despite this, equally, even so, furthermore, however, in addition, in comparison with..., in contrast, in fact, in spite of, in the same way, instead, likewise, moreover, nevertheless, nonetheless, on the contrary, on the one hand, on the other hand, similarly, yet
concluding ideas	all in all, altogether, finally, in conclusion, in summary, on average, on balance, on the whole, therefore, to conclude, to some extent, to sum up, to summarize
examples	for example, for instance, generally, in general, in particular, mainly, particularly, primarily, principally, specifically
generalization	as a rule, basically, by and large, essentially

表 4

対象とした 40 の単語連鎖

on the basis of, at room temperature, it is necessary to, in this/the section we, in the presence of, as described in materials, be shown to, is one of the/these/those, on the other hand, under the control of, at a concentration of, be available from the, in the absence of, as shown in fig/figure, as well as the, in this/the study are, according to, the total number of, as previously described (in), in this/the study the, in this paper we, the number of, can be found in, a set of, shown in table, a wide range of, in the context of, it [BE] possible to, as a function of, as a result of, in the case of, by the addition of, on the surface of, we did not observe, we find that, used in this study, was/were added to, with respect to, at the end of, at a rate of

3.4 使用したソフトウェアと統計技法

次の 2 点のソフトを使用した。AntConc3.5.8 は、早稲田大学の Laurence Anthony 氏が開発・

発表しているフリーのコンコーダンサーである。AntConc の Word List により、頻度を抽出し、調整頻度を算出した。統計処理には、早狩進氏が開発・発表している統計のソフトウェア Seagull-Stat を使用した。Seagull-Stat を使用して、 χ^2 二乗検定、相関分析、コレスポンデンス分析を行った。

3.5 手順

1. コーパスの構築

- (1) 採取した論文は、pdf ファイルから text ファイルに変換。
- (2) Text ファイル(ansi)で保存。語彙と構文などを分析調査する目的から、タイトル、著者の氏名・所属、小見出し、目次、謝辞、参考文献は手作業で削除。文字化け、写真、図、表、数式なども手作業で削除。

2. データの収集と分析

AntConc を使用して対象とした連結語と単語連鎖の頻度を確認し表を作成。Seagull-Stat により統計量を抽出。

4. 結果と考察

4.1 RQ1: 連結語の結果と考察

英語科学技術論文でよく使用される 84 の連結語を対象に、相関係数、 χ^2 二乗検定、コレスポンデンス分析を実施した。

4.1.1 相関係数

Nature コーパス, Science コーパス, Cell コーパスと Student コーパス間の連結語の使用において、強い相関が認められた ($r = .95, .96, .94$)。この結果は、連結語はすべてのコーパスにおいて使用傾向が近いことを示唆している。

表 5

4つのコーパスにおける連結語使用の相関

	Student	Nature	Science	Cell
Student	1.00			
Nature	.95**	1.00		
Science	.96**	.94**	1.00	
Cell	.94**	1.00**	.93**	1.00

4.1.2 χ^2 二乗検定とボンフェローニ補正

Student コーパスにおける連結語が、他の 3 つのコーパスにおける連結語の使用頻度と比べ、すべてにおいて有意差があったのは、on the other hand であった。

表 6

χ^2 二乗値と p 値: Student コーパスにおいて頻度が高い連結語

連結語	vs. <i>Nature</i>		vs. <i>Science</i>		vs. <i>Cell</i>	
	χ^2	p 値	χ^2	p 値	χ^2	p 値
on the other hand	19.49	<.001	14.45	<.001	18.75	<.001

表 7 は, *Nature* コーパス, *Science* コーパス, *Cell* コーパスの 3 つにおける連結語が, Student コーパスにおける連結語と比較し有意差があったものを示している。12 の連結語の使用頻度に有意差が出る結果となった。Student コーパスと比べると, この 3 つのコーパスにおいてさまざまな連結語を使用していることがわかる。

表 7

χ^2 二乗値と p 値: 3 つの科学誌論文コーパスにおいて頻度が高い連結語

連結語	vs. <i>Nature</i>		vs. <i>Science</i>		vs. <i>Cell</i>	
	χ^2	p 値	χ^2	p 値	χ^2	p 値
after	21.03	<.001	11.87	<.001	23.88	<.001
next	14.90	<.001	7.01	<.01	16.03	<.001
first	26.38	<.001	10.79	<.01	28.60	<.001
similarly	6.93	<.01	18.07	<.001	7.34	<.01
in contrast	31.16	<.001	37.01	<.001	32.46	<.001
although	11.58	<.001	68.07	<.001	12.63	<.001
specifically	32.84	<.001	24.92	<.001	33.96	<.001
thus	35.97	<.001	28.92	<.001	38.40	<.001
despite	9.33	<.01	36.92	<.001	15.38	<.001
primarily	6.66	<.01	14.11	<.001	10.42	<.01
on average	7.24	<.01	15.88	<.001	6.69	<.01
additionally	8.60	<.01	9.14	<.01	14.70	<.001

4.1.3 コレスポネンス分析

まず, 第 1 次元と第 2 次元における行(連結語)と列(4 つのコーパス)に与えられたアイテムの得点を基にしたコレスポネンス分析を行った。表 8 を見ると, 第 1 次元の寄与率は 55.23%, 第 2 次元の寄与率は 37.37% となり, 2 つの次元で 4 つのコーパスにおける連結語のデータが 92.6% 説明されているため, コレスポネンス分析によって適格にデータをまとめていると言える。

表 8

固有値・寄与率・累積寄与率表

	次元 1	次元 2	次元 3
固有値	0.05	0.03	0.01
寄与率 (%)	55.23	37.37	7.4
累積寄与率 (%)	55.23	92.60	100.00

図 1
 コレスポネンス分析に基づくスコア散布図
 (第 1 アイテム[コーパス])

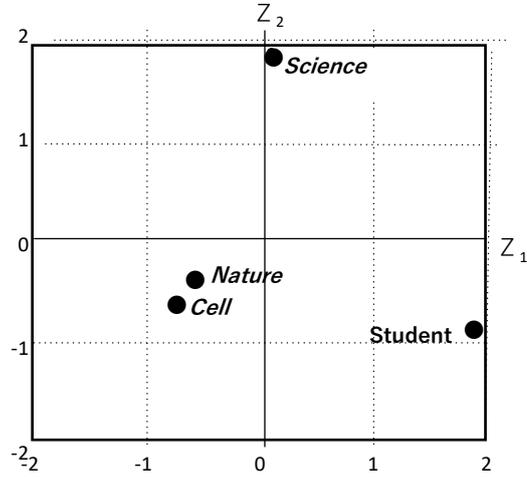
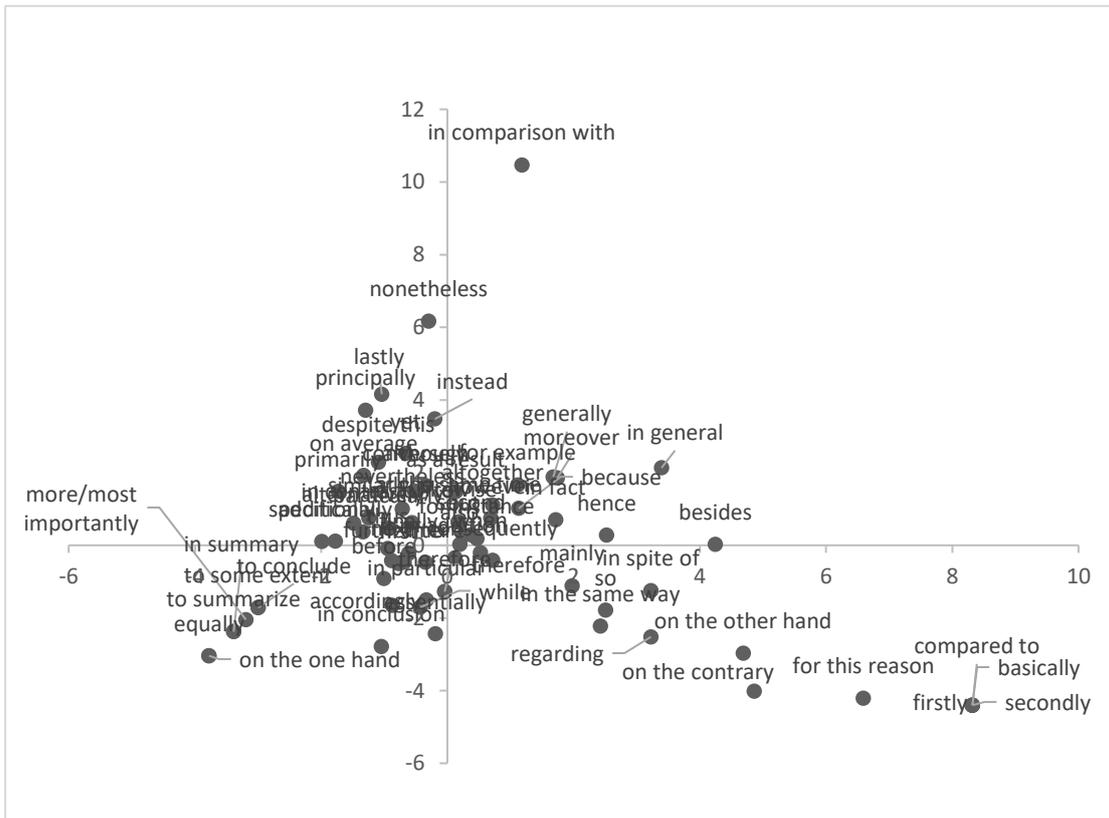


図 2
 コレスポネンス分析に基づくスコア散布図 (第 2 アイテム[連結語])



コレスポネンス分析に基づく散布図は図の見やすさを考慮し、今回は第 1 第 2 アイテムを別の散布図で示している。図1のコーパスの散布図では、第 1 次元のプラス得点方向には、Student コーパスが位置しており、マイナス得点方向には *Nature* コーパス、*Cell* コーパスが位置している。*Science* コーパスは第 1 次元ではゼロ付近である。

次に図 2 の第 2 アイテムがプロットされた散布図を確認する。この図 2 は図 1 の第 1 アイテムの散布図と対応しているため、Student コーパスの方向に大学院生が良く使用する *basically*, *firstly*, *secondly* といった連結語が見られる。一方左側には *Nature*, *Cell* に顕著に現れる連結語が見られるはずであるが、対象連結語の数が多く密集していることや、図 1 と図 2 では軸のスケールが異なることから、具体的な連結語の分布スコアを示した表 9、表 10 を見ていくことにする。

表 9

第 1 次元におけるプラス・マイナス得点上位 10 位までの連結語

プラス得点上位 10 位		マイナス得点上位 10 位	
<i>firstly</i>	8.32	<i>equally</i>	-3.78
<i>secondly</i>	8.32	<i>on the one hand</i>	-3.78
<i>basically</i>	8.32	<i>in summary</i>	-3.39
<i>compared to</i>	8.32	<i>to summarize</i>	-3.39
<i>for this reason</i>	6.59	<i>more/most importantly</i>	-3.19
<i>on the contrary</i>	4.86	<i>to some extent</i>	-3.00
<i>on the other hand</i>	4.69	<i>to conclude</i>	-3.00
<i>besides</i>	4.24	<i>specifically</i>	-2.00
<i>in general</i>	3.39	<i>additionally</i>	-1.77
<i>in spite of</i>	3.23	<i>primarily</i>	-1.73

表 10

第 2 次元におけるプラス・マイナス得点上位 10 位までの連結語

プラス得点上位 10 位		マイナス得点上位 10 位	
<i>in comparison with</i>	10.47	<i>firstly</i>	-4.40
<i>nonetheless</i>	6.17	<i>secondly</i>	-4.40
<i>lastly</i>	4.16	<i>basically</i>	-4.40
<i>principally</i>	3.71	<i>compared to</i>	-4.40
<i>instead</i>	3.48	<i>for this reason</i>	-4.20
<i>yet</i>	2.53	<i>on the contrary</i>	-4.01
<i>despite this</i>	2.29	<i>equally</i>	-3.04
<i>in general</i>	2.13	<i>on the one hand</i>	-3.04
<i>on average</i>	1.92	<i>on the other hand</i>	-2.97
<i>generally</i>	1.88	<i>in conclusion</i>	-2.78

表 9 は、第 1 次元におけるプラス・マイナス得点上位 10 位までの連結語であるが、プラス得点上位 10 位は日本人大学院生が比較的使用する表現であり、マイナス得点上位 10 位は使用頻度が低い表現に分けられている。プラス得点には、*firstly*, *secondly* のような *sequencing* が見られるが、大学院生を含め、多くの場合は *first*, *second* を使用する。*firstly*, *secondly* の使用は、

Nature, *Science*, *Cell* コーパスでは、ほぼ認められない。Firstly よりも first を推奨する意見 (日本物理学会 1999) もある中、大学院生が firstly, secondly を使用したのは、Hyland (2008a, b) が指摘しているように、表現力があることを示したいためであるかもしれないが、断定できるほどのデータがないため推測の域を出ない。また、理由を表す for the reason, 前後の文で比較対照する表現の on the contrary, on the other hand, in spite of は、大学院生には馴染みの深い表現であると言えよう。

第 2 次元におけるプラス・マイナス得点上位 10 位までの連結語であるが、プラスの得点方向に *Science* があり、マイナスの得点方向に *Student* がある。寄与率が 37.37% であるので、解釈が難しい。マイナス得点が第 1 次元におけるプラス得点上位 10 位とほぼ同じであることから、こちらも日本人大学院生による使用率の高い低いで分けられそうである一方、*Nature*, *Cell* もマイナス得点方向にあることから、この点をどのように説明するのか、また、*Science* と *Nature/Cell* にはどのような差があるのか、この点についてはうまく説明がつかないため、分析方法を精査して次回分析につなげるべき課題である。

4.2 RQ2: 4-gram 中心とした単語連鎖の結果と考察

4.2.1 相関係数

表 11 により、*Nature* コーパス、*Science* コーパス、*Cell* コーパスと *Student* コーパス間の単語連鎖において弱い相関、中程度の相関が認められた ($r = .09, .33, .45$)。全体的に使用傾向の類似性は限定的であることを示している。一方、*Science* コーパスと *Cell* コーパスには強い相関が確認された ($r = .84$)。

表 11

4 つのコーパスにおける単語連鎖使用の相関

	<i>Student</i>	<i>Nature</i>	<i>Science</i>	<i>Cell</i>
<i>Student</i>	1.00			
<i>Nature</i>	.09**	1.00		
<i>Science</i>	.33**	.45**	1.00	
<i>Cell</i>	.45**	.48**	.84**	1.00

4.2.2 χ^2 二乗検定とボンフェローニ補正

表 12 は、*Student* コーパスと他の 3 コーパスとの間に有意差があった単語連鎖を示している。it is 形容詞+to 不定詞構文, frame markers と呼ばれるセクションはじめの宣言文のような表現 (in this paper we, in this section we) や endophoric markers と呼ばれる本文中の他の部分の情報指し示す表現 (shown in table, as described in materials) (Hyland, 2005) に有意差があることが見て取れる。

表 12

χ^2 二乗値と p 値: Student コーパスにおいて頻度が高い単語連鎖

単語連鎖	vs. <i>Nature</i>		vs. <i>Science</i>		vs. <i>Cell</i>	
	χ^2	p 値	χ^2	p 値	χ^2	p 値
on the other hand	19.49	<.001	14.45	<.001	19.49	<.001
in this paper we	22.46	<.001	13.10	<.01	22.46	<.001
shown in table	20.23	<.001	9.49	<.01	20.23	<.001
as described in materials	16.90	<.001	9.76	<.01	16.90	<.001
it is necessary to	9.85	<.01	8.43	<.01	9.85	<.01
in this section we	14.23	<.001	9.76	<.01	14.23	<.001
it is possible to	7.55	<.01	8.17	<.01	7.55	<.01

一方, *Nature* コーパス, *Science* コーパス, *Cell* コーパスの3つに有意差があった単語連鎖はなかった。個々に Student コーパスと比較をすると, *Nature* コーパスに有意差があった単語連鎖は, on the basis of, the number of, at a concentration of, we find that, is one of the, be available from the, *Science* コーパスに有意差があった単語連鎖は, on the basis of, according to, as a function of, the number of, we didn't observe, as well as, *Cell* コーパスに有意差があった単語連鎖は, according to, at room temperature, the number of, be shown of, in the context of であった。専門内容をサポートするような表現 (on the basis of, the number of, at room temperature) がより多いと言える。

4.2.3 コレスポネンス分析

表 13

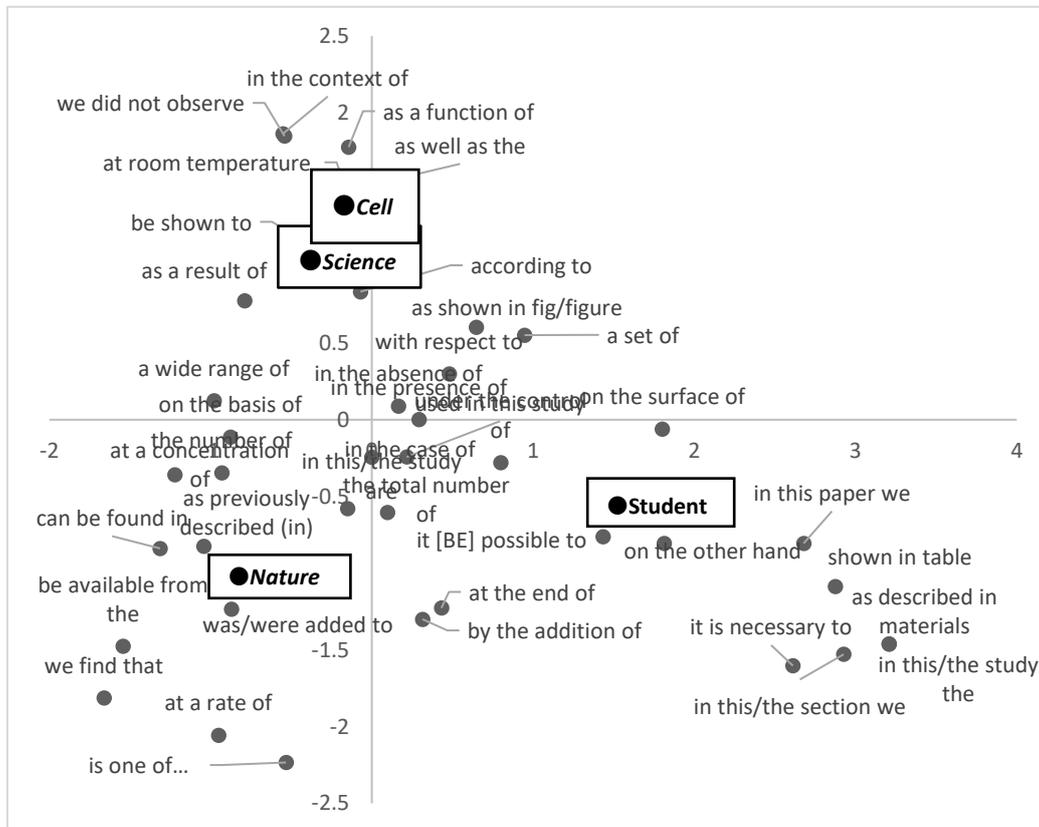
固有値・寄与率・累積寄与率表

	次元 1	次元 2	次元 3
固有値	0.26	0.17	0.05
寄与率 (%)	56.50	34.15	9.34
累積寄与率 (%)	56.50	90.66	100.00

第 1 次元と第 2 次元における行 (単語連鎖) と列 (4つのコーパス) に与えられたアイテムの得点を基にしたコレスポネンス分析を行った。表 13 が示しているように, 第 1 次元の寄与率は 56.50%, 第 2 次元の寄与率は 34.15% となり, 第 2 次元までの累積寄与率が 90.66% となっているため, 分析対象は第 2 次元まででうまくデータをまとめていると言える。

次に, コレスポネンス分析に基づく散布図を図3で見えていく。第 1 次元のプラス得点方向には, Student コーパスが位置しており, マイナス得点方向には *Nature* コーパスが位置している。*Science* コーパス・*Cell* コーパスは第 1 次元ではゼロ付近である。各コーパス付近に特徴的な単語連鎖がプロットされている。

図 3
 コレスポネンス分析に基づくスコア散布図 (単語連鎖)



具体的に、第 1 次元の得点に基づいて単語連鎖をプラス・マイナス得点の上位 10 位までを表 14, 表 15 で見ていくと、大学院生が良く使用する *it is* 構文とセクションのはじめの宣言を示す *frame markers* に加え、連結語でも使用が認められた *on the other hand* も見られる。これらは、 χ^2 二乗検定の結果とほぼ同一である。一方、マイナス得点に現れる単語連鎖は、科学的な内容に関わる *at a rate of*, *at a concentration of*, *a wide range of*, *the number of* や、前の記述を想起させる *endophoric markers* の *as previously described in*, 方法や結果を表す *was added to*, *we find that*, *be available from the*, *can be found in* が見られる。これは、 χ^2 二乗検定の *Nature*, *Science*, *Cell* コーパスに有意差があった表現と共通している。内容や書体において端的に表現するために使用されるフレーズと言えるため、科学技術論文に慣れている書き手が使用する表現がゼロ付近からマイナス付近で見られると考えられる。

第 2 次元におけるプラス・マイナス得点上位 10 位までの単語連鎖であるが、プラスの得点方向に *Science* と *Cell* があり、マイナスの得点方向に *Nature* と *Student* がある。*Science* と *Nature* では、単語連鎖に共通した使用が認められると言える一方で、*Nature* とは単語連鎖の使用傾向が異なることを示唆している。小屋(近刊)によると、*Nature* と *Student* では単語連鎖の使用が異なる傾向になり、*Student* は *Brown Corpus family* に近い使用傾向があることがわかっている。

今回の分析では、*Nature*と*Student*がマイナスの得点方向に共に位置するため、*Science*と*Cell*の単語連鎖の使用は、*Nature*と*Student*の使用よりもさらに異なる使用傾向があると言える可能性がある。表から専門性をサポートする表現として、プラスの得点方向では、*as a function of*, *at room temperature*, マイナスの得点方向には特に示されていないと言えそうであるが、解釈はかなり困難であるため、連結語のコレスポネンス分析の結果と同様、コーパスサイズを増やしたり、他のコーパスを加えて分析したりすることにより、解釈を試みるべきであると考えられる。

表 14

第1次元におけるプラス・マイナス得点上位10位までの単語連鎖

プラス得点上位10位		マイナス得点上位10位	
as described in materials	3.21	we find that	-1.66
in this/the study the	3.21	be available from the	-1.54
in this/the section we	2.93	can be found in	-1.31
shown in table	2.87	at a concentration of	-1.22
in this paper we	2.68	as previously described (in)	-1.04
it is necessary to	2.61	a wide range of	-0.98
on the other hand	1.81	at a rate of	-0.95
on the surface of	1.80	the number of	-0.93
it [BE] possible to	1.43	on the basis of	-0.88
a set of	0.95	was/were added to	-0.87

表 15

第2次元におけるプラス・マイナス得点上位10位までの単語連鎖

プラス得点上位10位		マイナス得点上位10位	
in the context of	1.86	is one of the/these/those	-2.23
we did not observe	1.85	at a rate of	-2.06
as a function of	1.78	we find that	-1.81
as well as the	1.57	it is necessary to	-1.60
at room temperature	1.08	in this/the section we	-1.53
be shown to	1.01	be available from the	-1.48
according to	0.83	as described in materials	-1.46
as a result of	0.78	in this/the study the	-1.46
as shown in fig/figure	0.60	by the addition of	-1.30
a set of	0.55	was/were added to	-1.23

5. まとめ

まず、連結語については、3つの科学雑誌コーパスがさまざまな表現を使用しているのに対し、理系大学院生は限定的な使用に留まっていることがわかった。日本物理学会(1999)では、「日本人は、連結語を使用する技術に欠けるため、論理的な文章を書くことが不得意である」というアメリカ人物理学者のコメントを引用し、連結語使用の技術を高めるべきとの主張している(p. 130)。実際に、中学校・高等学校の学習指導要領において、論理的な文章記述のために連結語の習得を目標にすると記載があったり、科学技術論文を書く人に向けたマニュアル本が多く出版されたり、

科学技術論文に必要な表現集を搭載する電子辞書が出てきたりしているが、まだその成果は出ていないことになる。また、特に **sequencing** の連結語は、日本人英語学習者の過剰使用が指摘され、それは論理的添加について表面的な理解に留まり、本質的な理解の欠如を示唆する(小林 2009, 2010, 三木 2016) 報告や、英語の習熟度レベルが低い日本人英語学習者ほど **sequencing** に頼るといふ報告もある(三木 2016)。しかしながら、理系大学院生が書く科学技術論文は、まだ使用を推奨する段階と言えるが、連結語の使用推奨と共に、論理的で効果的な使用の指導も重視すべきであると考えらる。

次に、単語連鎖については、*Nature* コーパス、*Science* コーパス、*Cell* コーパスと比較した場合、**it is** 形容詞+**to** 不定詞構文や **frame markers**, **endophoric markers** が多く使用されていることが認められた。**it is** 形容詞+**to** 不定詞構文については、*Nature*, *Science*, *Cell* の使用と比べると、学生の使用頻度が圧倒的に高い。この表現は中学校で学習するため馴染みが深い。主語が形式主語の **it** では、内容の把握が即座にできず時間がかかる。キーワードになる名詞を主語にすることによって回避できることから、**it is** 形容詞+**to** 不定詞構文を使わずに効果的に表せるように指導すべきである。篠田(2001)によると “It is easy to detect rust stains.” を例にあげ、 “Rust stains are easy to detect.” にすれば、理解のスピードが増すと指摘している (pp. 1-3)。このように、コーパスから得られる客観的なデータから、大学院生の書く論文の特徴と *Nature*, *Science*, *Cell* のような雑誌の特徴の違いを明らかにし、量的かつ質的に明示的指導を行っていく方がよいだろう。

frame markers の **in this paper/section/study** という表現は、*Nature*, *Science*, *Cell* ではほぼ使用が認められない。これは、特に明記する必要がなく、論文中では当たり前のことであるため、余剰性の多い表現であると考えられる。しかしながら、大学院生の論文ではよく使用される定型表現であり、**in this paper/study** はさまざまなセクションで使用され、**in this section** はセクションの冒頭で使用されることが多い。例えば、以下の(1)(2)は理系大学院生が実際に書いた記述である。

- (1) In this section, we describe a laboratory session for differential geometry of curves.
- (2) *In this section, first it explains the JavaME platform, and then it explains the J2ME Wireless Toolkit that is the simulation tool to execute the application, finally it explains about the function of the platform on the mobile device side.
→ In this section, we first explain the JavaME platform, ...

(1)は、**method** のあるセクションにおける冒頭の 1 文である。読めばわかるにもかかわらず、宣言してから書き始めている。(2)も同じく、**method** のあるセクションの冒頭の 1 文である。こちらは **sequencing** を使用して3点を順番に説明していくことを宣言しているが、主語を **we** に直す必要がある文である。野口・松浦(2008)、パケット(2004)では、このような句は、科学技術論文における

定型表現であるとされているため、大学院生が頻繁に使用するのではないかと考える。*Nature, Science, Cell* では、このような表現を使用しないでどのようにセクションを初めているかを実際に見て学ぶことが重要だと考える。また、(2)の主語 *it* は、明らかに誤りであることから、文法的な間違いも併せて指導することが必要ではある。

endophoric markers については、他の雑誌でも見られるが特に学生コーパスに多く見られることから、多用しすぎずに自然に読み手を導くような使い方を習得すべきであろう。

最後に、4つのコーパスを比較するのに、相関係数、 χ^2 二乗検定、コレスポネンス検定を行ったが、コレスポネンス分析結果の解釈がかなり難解であった。1つの理由として、理系の領域における4コーパスを比較したため、違いが思ったほど明確にならなかったことが考えられる。今回は理系のジャンルに絞り、トピックに偏らない高頻度な連結語と単語連鎖に着目したが、別の分野、例えば英語教育分野、文学、法学、文学等、さまざまな分野を含めて分析をした方が、解釈が容易になる可能性がある。また、コーパスサイズが限定的であったため、論文を増やすことで解釈にプラスになる可能性も考えられる。いずれにしても、客観的なデータに質的な視点を入れた分析を融合されて分析を進め、大学院生の論文執筆に効果的なサポートを目指したいと考える。新たな分析については、稿を改めて論述することにした。

引用文献

- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263-286.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.
- Chen, Y., & Baker, P. (2010). Lexical Bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30-49.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 397-423.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics, 18*(1), 41-62.
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.
- 石川慎一郎・前田忠彦・山崎誠(2010) (編著)『言語研究のための統計入門』くろしお出版。
- 小林雄一郎(2009)「NS/NNS テキスト分類モデルに基づく日本人英作文の特徴抽出」『人文科学とコンピュータシンポジウム論文集ーデジタル・ヒューマニティーズの可能性』, 261-268.
- 小林雄一郎(2010)「テキストマイニングによる学習者作文における談話能力の測定と評価」. *STEP Bulletin, 22*, 14-29.

- 小屋多恵子(近刊)「英語科学誌 *Nature* と理系大学院生の英語論文における表現の比較－単語連鎖に着目して－」『小金井論集』(法政大学) 15.
- 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比較」『統計数理研究所共同研究レポート』vol.239, 1-12.
- 三木望(2016)「第 34 回:コーパスで学習者の英作文の特徴を探る－『メタ談話標識』とは何か－」研究者 WEB マガジン *Lingua*, リレー連載:実践で学ぶコーパス活用術, <http://www.kenkyusha.co.jp/uploads/lingua/prt/15/MikiNozomi1607.html>.
- 日本物理学会編(1999)『科学英語論文のすべて』丸善出版.
- 篠田義明(2001)『科学技術の英語－IT時代の英語の読み方・書き方』早稲田大学出版部.
- 篠田義明(2007)『伝える英語の発想法』早稲田大学出版部.
- Wallwork, A. (2016). *English for writing research papers*. Springer.
- 八野幸子(2018)『英語理学療法論文における Multi-Word Expression－コーパス研究から学習及教育への応用を念頭に－』(博士論文 大阪大学大学院言語文化研究科言語文化専攻).
- Zemach, D., & Rumisek, L. (2011). *Success with College Writing*. Macmillan Language House.