

科学英語論文アブストラクトにみられる英語表現

萩原明子

1. はじめに

インターネットによる論文検索が一般的になった現在、英文で書かれた学術論文におけるアブストラクトの役割は、以前にも増して重要になってきている。論文の引用回数とその論文の重要度を測る物差しである現状において、実際に論文が読まれ引用される為には、良い研究結果を出すことがまず重要だが、それと同時に論理的でわかりやすいアブストラクトを書き、他の研究者にその論文を読んでもらう必要がある。このような観点から、効果的なアブストラクトを書く技術を身につけることは研究者にとって非常に重要なことであると考えられる。

アブストラクトは論文全体を簡潔にまとめたものだが、長い論文を限られた字数にまとめる為には英作文のある種のスキルが必要である。多くの研究者は、長い経験により恐らく無意識のうちに既にそのスキルを身につけていると考えられるが、まだ論文に触れる機会の少ない大学院生は、論文を読み、また、書くことによって少しずつそのスキルを学んでいる途上にあると考えられる。このようなアブストラクトを書くためのスキルの一つとして考えられるのは、形式的なスキーマと呼ばれる文章上の形式的な修辞（レトリック）構造に関する知識とその応用である。これは、科学英語の修辞構造が必ずしも他のジャンルの文章と同じではないということと、さらに他の言語、たとえば日本人研究者にとっては母語である日本語とも同じではないということから日本語母語話者が習得する必要のあるスキルである。日本人大学生は、長期間にわたり英語の学習を続けているが、大学の高学年にならない限り、英語で書かれた科学論文を読む経験はなく、その修辞構造に触れることもない。そのような学生が大学院生になって自分の論文のアブストラクトを書くためには、科学論文特有のレトリックを学び、それに適切な言語表現を選び使用することが必要であるが、英語教育の中でこのスキルを体系的に習得させる試みは、まだ十分にされていないのが現状である。

研究室で研究を始めたばかりの大学生や大学院生にとっては、論文のアブストラクトを読んで理解することも、実は、易しいことではない。実際、4年生で研究室に配属になった学生から、セミナーで論文を読む時に必ずしもアブストラクトをよく読んで概要を理解してから、本文を読みすすむという過程をとっていないということを聞いたことがあるが、背景知識が十分でない学生にとっては、アブストラクトを読んで理解することは、案外難しいことなのである。アブストラクトを読む為のスキルは多くの論文を読み、科学論文読解ストラテジーを磨くと共に、実際に研究をすることにより背景知識を増やす事の両方を平行して行う事が必要であるが、実際の論文読解にそれらを応用するための方法は、一般的な日本人学生のような英語の習熟度が高くない学生のための教育には、十分に取り入れておらず経験を積む以外の方法はあまり想定されていない。

一方で、情報処理技術を用いて言語学的な分析を行い既成の論文からアブストラクトを生成する技術も開発されつつある。生命科学分野において英語は共通語であり、英語の第二言語話者である多くの科学者にとっては、情報処理技術の発展により言語的なバリアがなくなっていくことは喜ばしいことではあるが、現在まだ実用に供するまでには到っていない。しかしながらテキストを機械的に分析する情報処理の手法から得られた知見は、機械翻訳や自動要約生成だけでなく、英語教育にも応用が可能である。現在の日本の英語教育はこのような情報処理技術の発展に対して、対応しているといえず、言語資料を分析する事すら、まだ十分な形で行われていない。

* 言語科学研究室

生命科学の共通語は英語であり、論文は基本的に英語で書かれるが、書き手の母語は英語であるとは限らず多くの第二言語話者が論文を英語で発表している。とりわけ、日本人研究者の占める割合は高く、国別の論文発表数はトムソンロイターの統計では世界2位 (Science Watch, 2010) であった。しかし、論文数が1位のアメリカをはじめとして英語圏からの論文発表数はそれ以外の言語に比べ圧倒的に多く、さらに被引用件数を見てみると、国あたりの平均被引用回数では、日本は4位に下がり、論文1編あたりの平均被引用回数の国別の統計では、欧米の国々が上位で、北米、ヨーロッパ以外の国で上位20位に入っているのは、オーストラリアとイスラエルのみで、日本を含むいわゆるアジア圏の国は、20位に入っていない。すなわち、学術雑誌への掲載数に比べ相対的に引用数が少ないのが、日本人研究者の書く論文であると統計上は解釈が可能なのである。もちろん、論文の引用回数には、様々な要素が関連しているため、単純に国別の統計で比較するわけにはいかないが、掲載数に比べ被引用回数が少ない理由の1つとして、研究者の母語のファクターが予想以上に大きいのではないかと考えられる。

本研究は、生命科学分野において使用される連語表現を学術論文のアブストラクトから抽出し、生命科学論文の言語的特徴を第二言語習得の観点から調査したものである。ここでいう連語とは、語と語がどんな理由であれ共起しているものをすべて指すこととする。文法、語彙の知識とは異なり、語と語のつながりは、慣用的な要素が深く関わっているため、相当な量の言語的インプットなくして連語は習得されにくいと考えられる。本稿では、まず、言語学的に見た連語の位置づけを最近のコーパス研究から論じ、次にアブストラクトを集めた2つのコーパス (日本人研究者によって書かれたアブストラクトと英語圏の研究者によって書かれたアブストラクト) で連語を分析した結果を述べ、第二言語話者である生命科学の日本人大学生、大学院生の英語教育への応用を論じる。

2. 言語習得と連語

言語の習得には多くのインプットとアウトプットの量が必要であると言われている (Richard 2002, p.44) が、第二言語話者である生命科学研究者の場合、数多くの論文を読み、自らの研究を論文上で発表するばかりでなく、ピアとして他の研究者の論文をレビューする機会や、論文集の編集を行う事もあり、言語習得という観点から考えると、論文を書くことにおいて非常に効果的な言語習得が行われている可能性がある。すなわち論文を読み、書くプロセスとは、研究者間の現実のコミュニケーション (authentic communication) のための英語使用であり、その過程において英語の習得が行われていると考えられるからである。

多くの英語教育論において、現実には目的を持ったコミュニケーションが必要であるとする理由は、言語が文法的ルールに則って使用されるだけのものではないという考え方からくるものである。生成文法的に言語習得は統語規則の習得と語彙の知識の集積によるものであると考え、現実のコミュニケーションにおける言語使用を正確に理解し習得することは困難である。実際の言語によるコミュニケーションでは、可能なすべての文法項目が使用されるわけではなく、文脈上の意味において使用可能な語彙が必ずしも使用されるわけでもない。コミュニケーションで使用される言葉は、語用論的なルールに基づき、慣用性にしたがって使用されているのである。つまり、文法の知識と語彙の知識だけでは説明の出来ない語用論的、慣用的な知識無しには、現実のコミュニケーションは成り立たない。そして、そのような知識は、現実のコミュニケーションを通じて得られるものであり、多くの英語教育理論が authentic communication の必要性を説く根拠もそこにあるのである。

このように言語を習得するためには、文法のルールを身につけ、多くの語彙を学習するだけでなく、語と語の慣用的な結びつきや、語用論的な決まり事を学ぶことが必要であり、英語で科学論文を読むためにも、英語の文法の他に、慣用化された言語のパターンに関する知識が必要である。近年、言語資料を集め分析するコーパス研究の枠組みで、連語に関する研究 (Wray 2002; Schmitt 2004) が多く行われてきている。連語の一部は、今までもたとえばオーディオリングアルメソッドに代表される古典的な言語教育法でのパターンプラクティスの様な反復学習のタスクにおいて教

材に組み込まれていたため、教育に取り入れられていなかったわけではない。特に、日本の受験英語においては、イディオムという用語で多くの慣用表現が学習項目として認知され使用されてきた。本研究のテーマである連語とは、文法的、意味的、または他の理由で結びつけられた一連の語を意味し、本来の語の意味から類推される意味と異なる意味を持つイディオムだけでなく、文法的に自然に生成されるあらゆる語と語の組み合わせのなかで何度も生起するものを指す。

連語を認知心理学的な面で捉えると、連語の処理は語と語のユニークな文法的組み合わせに比べると処理スピードが速いという言語処理上の利点があることが報告されている (Gibbs 1994; Wray 2002)。つまり、母語話者のようにその言語に触れる機会が多いと慣用的によく使われる連語が会話や文中に出てきた場合、連語全体を1つの意味のある塊として処理するため文法処理の必要がなくなり、処理の負荷が減少すると考えられている。一方、第二言語話者の場合、連語に触れる機会が限定的であるため、言語処理においての負荷が減らないばかりか、イディオムのように語の意味から全体の意味が類推できない場合は、文脈から意味を類推するほかに理解に至らないことも考えられる。幸い、科学英語の場合は、内容をより正確に伝えることが論文の目的であることから、文脈の上で多義的、両義的になる文章表現を避ける傾向があるため、イディオムが生起する可能性は少ないと考えられる。

3. 研究

3-1. 目的

本研究の目的は、コーパスデータを用いて語と語の組み合わせである連語を抽出することにより、科学論文における慣用的な語の結びつきを数的に分析し、生命科学を学ぶ若い研究者が科学英語を習得するために有益な慣用表現を英語教育に活かすための基礎データを収集することである。日本語を日常のコミュニケーションの手段として使用している日本語母語話者の研究者と英語を使用している英語圏の研究者の書くアブストラクトをそれぞれ収集し、連語の使用の違いを分析し、アブストラクトの英文において連語の使用法に違いが見られるか調べた。

3-2. 方法

生命科学に関連する学術論文アブストラクトのコーパスを編纂した。東京薬科大学生命科学部の Annual Report (2004～2007) に掲載されている学術雑誌から日本の研究機関に所属している日本人によって書かれた論文と英語圏の研究所で書かれた論文のアブストラクトを収集し、日本人研究者によるものを JS コーパスとし、英語圏で書かれたものを NS コーパスとした。論文アブストラクトを ASCII テキストに変換しコンコーダンスソフトウェアの WordSmith Tools 5.0 を使用しワードリストを作成した。その後、連語の検索を行い、語が4つ連続して生起している連語の頻度順の表を作成した。JS と NS のコーパスから得られた2つの連語表を元に、対数尤度検定を行い、生起頻度の差を求めた。

3-3. 結果

それぞれのコーパスの中に含まれる語の頻度を分析した結果は、以下の通りである。

* JS コーパス：日本人研究者によって書かれたアブストラクト 1450 編

総語数：255,996 語 / 異語数：15,962 語

* NS コーパス：英語圏の研究者によって書かれたアブストラクト 1450 編

総語数：274,038 語 / 異語数：17,067 語

NS コーパスに含まれるアブストラクトの方が、総語数、異語数ともに多い結果となった。しかし、3回以上生起する連語を調べたところ、JS コーパスの方が4連語の種類とその生起頻度共に多かった (表1)。2つのコーパスに共

通して生じた連語は467種類あり、それらの頻度は、JSのほうがNSよりも多かった。

表1. 4連語の種類と頻度

コーパス	連語数 (type)	頻度 (token)
JS	1732	8988
NS	1494	7033
共通	467	4202 (JS) 3337 (NS)

4連語のなかで両方のコーパスで出現頻度が特に多かったものは、“these results suggest that,” “in the presence of,” “in this study we,” “these results indicate that,” “here we show that,” “in the present study”の6つの連語であり、それぞれのコーパスにおける頻度はJSでは598回、NSでは305回であった。

表2. JSコーパスにおいて出現頻度の高い4連語

	JS 頻度	NS 頻度
these results suggest that	161	51 ***
in the presence of	114	85
in this study we	101	64
these results indicate that	93	28 ***
in the present study	82	33 **
the present study we	64	29 *
here we show that	60	57
an important role in	55	35
on the other hand	46	8 ***
these findings suggest that	41	14 *
for the first time	38	21
results suggest that the	37	13

*** $p < .000.1$, ** $p < .001$, * $p < .01$

表3. NSコーパスにおいて出現頻度の高い4連語

	JS 頻度	NS 頻度
in the presence of	114	85
in this study we	101	64
here we show that	60	57
these results suggest that	161	51 ***
in the absence of	28	45
an important role in	55	35
in the present study	82	33 **
a member of the	28	29
we show that the	13	29
in the regulation of	22	28
these results indicate that	93	28 ***
the present study we	64	29 *

*** $p < .000.1$, ** $p < .001$, * $p < .01$

それぞれのコーパスにおいて出現頻度の高い連語を対数尤度比を使って比較したところ、表2と表3の結果が得られた。表2はJSコーパスでの出現頻度を基準にしたものであり、表3はNSコーパスを基準にしたものである。表2

と表3からは、出現頻度の高い連語は、どちらのコーパスにおいても使用されているが、頻度の順位と比率が違うことが明らかになった。NS コーパスにおける頻度がJS コーパスにおける頻度より比率の統計において有意に多いものはなかった。

出現の比率において統計的に有意な差が出たものに、アブストラクトの結論部分の表現に使用される連語がある。アブストラクトの最後の部分では、研究の結果がどのような意味を持つかを述べることが多いが、“these results suggest that,” “these results indicate that,” “these findings suggest that” は、非常に頻繁に使用される表現であり、JS コーパスでの使用の比率が有意に高かった。特に高頻度の“these results suggest that”のパターンで副詞が入るものを抽出してみたところ、strongly が suggest の前に挿入されているケースを調べたところ JS コーパスで頻度が19回のところ NS コーパスでは、3回だけであった。

アブストラクトの最後に使用される“these results suggest that”のパターンと同じような連語を調べたのが、表5である。同じ文型で使用されるモダリティの違う動詞をJSでの頻度を基準に並べてある。動詞 suggest, indicate という断定の程度が低い動詞が高頻度で使用され、その傾向はJSにおいて顕著であった。

表5. JS コーパス, NS コーパスにみられる結論部分の表現の違い

	JS 頻度	NS 頻度
these (results/data/findings, etc.) suggest that	237	89
these (results/data/findings, etc.) indicate that	125	37
these (results/data/findings, etc.) demonstrate that	13	23
these (results/data/findings, etc.) show that	8	22
these (results/data/findings, etc.) imply that	3	0

4. 考察

論文において一番読まれる可能性のある部分がアブストラクトだが、連語の分布を調べると、日本人科学者が書いたものと英語圏の研究者が書いたものに違いが見られた。しかし、具体的な連語を調べると、日本人研究者は、アブストラクトで頻繁に使用される連語を多く使用していることがデータから明らかになった。アブストラクトというジャンルの文章の特徴を第二言語話者として習得している様子がうかがえる。これは、多くのアブストラクトを読んだり書いたりした経験から身につけたスキルであると考えられる。

連語の頻度を比較した場合、英語圏の研究者が書いたアブストラクトに使用されている連語に比べ、日本人研究者の方が、特定の連語を多用している傾向が見られた。これは、第二言語話者に共通に見られる連語の過剰使用 (Wray 2002) であると考えられる。第二言語話者は、数の少ない使い慣れた安全な表現を多用することでコミュニケーション上の必要性を満たすということだが、日本人研究者がアブストラクトを書くときも同様に使い慣れた表現を選んでいると解釈できる。アブストラクトの結句として使われる表現には、その傾向がより強く見られた。NS コーパスでも suggest の生起頻度は高いが、連語としてJS コーパスでの使用例は多く、断定の意が弱いいため、安全な動詞として使用されているようである。一方、NS は、より語気の強い動詞を使用する頻度が高く、例は少ないものも suggest を strongly といった副詞を使って修飾するよりは、show や demonstrate のような断定の意の強い動詞を選んで使用している例が見られた。

日本人研究者が第二言語で研究発表をしなければならない点は、言語的な負担が多いことから、研究上不利であると考えられるが、現実には、研究に携わっているうちに、必要な英作文能力を身につけ、実際に使用している点がデータからある程度読み取ることができた。しかし、連語の使用については、第二言語話者の特徴がみられ、使い慣れた安全な言語表現を多用し、結果として、研究成果を位置づける重要な表現が語用論的にみると研究者の意図よりも弱く

なっていることが指摘できる。使用される動詞の日本語訳での意味にとらわれて動詞を選ぶよりも、同様の機能を持つ連語のレパートリーを増やし、連語としての用法として使い分けることが重要であろう。

「習うより慣れよ」という表現があるが、言語運用は、慣用によってのみ学べるものである。しかし、経験だけで学ぶには、非常に長い時間が必要である。生命科学の研究者を目指す学生に効果的な英語教育を行うには、文法、語彙とともに連語や語用といった慣用的な言語表現を明示的に取り入れる必要がある。

引用文献

- Gibbs, Raymond W. 1994. *The poetics of mind: figurative thought, language, and understanding*.
- Richards, Jack C. 2002. Accuracy and Fluency Revisited. In *New Perspectives on Grammar Teaching in Second Language Classrooms*. Hinkel, Eli, Fotos, Sandra (ed. and introd.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schmitt, Norbert. 2004. Formulaic sequences in action: An introduction. In *Formulaic Sequences: Acquisition, Processing and Use*. Norbert Schmitt (ed.), 1-22. Amsterdam: John Benjamins Publishing Company.
- Science Watch. 2010. Retrieved from <http://sciencewatch.com/dr/cou/2010/10decALL/>
- Wray, Alison. 2001. Formulaic sequence in second language teaching: Principle and practice. *Applied Linguistics*. 21, 463-489.
- Wray, Alison. 2002. *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.