Kobe University Repository : Kernel

PDF issue: 2025-05-21

コロケーションとその構成単語の難度の関係性についての研究

石井，康毅

(Citation)
Journal of Corpus-based Lexicology Studies,5:39-53

(Issue Date)
2023-03-10

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100479384

(URL)
https://hdl.handle.net/20.500.14094/0100479384

英語コーパス学会 語彙研究会（Japan Association for English Corpus Studies, Lexicology SIG）
Journal of Corpus-based Lexicology Studies, Vol. 5 (2022)
ISSN: 2434-169X

コロケーションとその構成単語の難度の関係性についての研究

石井　康毅（成城大学）
ishii@seijo.ac.jp

Research on the Relationship between Difficulty Levels of Collocations and Their Constituent Words

ISHII, Yasutake (Seijo University)

Abstract

While the difficulty levels of words have been extensively researched and many resources have been developed, there has been little research on the difficulty levels of collocations. As a preliminary study for future research on the levels of collocations, the relationships of CEFR levels between collocations and their constituent words were investigated using data from the English Vocabulary Profile and the CEFR-J Wordlist. It was found that most collocations have higher levels than their constituent words, and some possible reasons are suggested.

Keywords

collocations, lexical difficulty, English Vocabulary Profile, CEFR-J Wordlist

1. 研究の背景と目的

単語同士の慣習上の自然な結びつきであるコロケーション（collocation）の重要性は、英語教育の研究・実践において広く認識されるようになって久しく（Granger, 1998；Howarth, 1998；Lewis, 2000；Ellis, 2003；Nesselhauf, 2005；堀, 2009；Chen & Baker, 2010）、大規模コーパスに基づいて編集されたコロケーション辞典（McIntosh, 2009；Rundell, 2010；Mayor, 2013；塚本, 2013；八木他, 2015）やコロケーションリストなどのリソース（Davies & Gardner, 2010；Ackermann & Chen, 2013）も、学習者や教育関係者に広く利用されるようになってきた。なお、「コロケーション」という術語は、(1) 一定の言語慣習により結びつきの相性がよいとみなされる複数の単語が結びつく事象、(2) ある特定の単語から見て結びつく相手の語、(3) 結びつきによってできたフレーズ、の 3 つの対象を指し得るが、本稿では専ら上記（3）の結びつきによってできたフレーズの意味で「コロケーション」を用いる。また、本稿でコロケーションとそれ以外のフレーズを区別する際に使用した基準は 2.2.1 で述べる。

単語については、頻度や難易度を示す情報の付いたリソースが数多く開発され利用され

てきている。例えば New General Service List（Brown et al., 2013）、JACET 8000（大学英語教育学会基本語改訂特別委員会, 2016）、CEFR-J Wordlist（東京外国語大学投野由紀夫研究室, 2020）などがよく知られ、利用されている。（New General Service List と JACET8000 は頻度に基づくバンド情報を示していて、CEFR-J Wordlist は 2.1 で述べるように難易度情報の一種と考えられる CEFR レベルの情報を示している。）。しかしながら、コロケーションについては、2.1 で述べるように限定的な数のコロケーションに言語能力レベルの規定として世界で広く利用されている CEFR（Common European Framework of Reference）（Council of Europe, 2001）のレベルを表示している English Vocabulary Profile（Cambridge University Press, 2015）（以降「EVP」）などの小規模な例を除き、コロケーションの習得難度レベルを一貫した方法で判定し、網羅的なリストを作成するという研究・実践は、筆者の知る限りではこれまでに行われていない。コロケーションの難易度情報があれば、学習者にとってのみならず、教材作成・評価を行う教育関係者にとっても非常に有効なリソースとなるはずである。

コロケーション、あるいはより一般的に複数語のまとまりであるフレーズ単位の語彙リストの作成が難しい要因はいくつか存在する。第一の要因として、フレーズは単語と違って独立した単位の認定が自明ではないため、何をフレーズ項目として認定するかという基準が必ずしも明確ではない。第二の要因として、フレーズ項目を特定したとしても、フレーズは単語よりも頻度が小さいため、コーパス中での頻度を基に習得難度を判断するという、単語の場合に使える手法をそのまま利用することが難しい。第三の要因として、学習者を意識した習得難度レベルをフレーズに付与する場合には、（望ましくはインプットとアウトプットの両方のモードで）習熟度レベル別のコーパスを利用する必要があるが、習熟度レベル別のコーパスは数と規模の両方の点で限定されている。第四の要因として、フレーズ全体の習得難度レベルが、その構成要素である単語（以降「構成単語」と呼ぶ）の習得難度レベルとは必ずしも一致しない場合が多い。本研究では、この第四の要因に特に注目して、コロケーションのフレーズ全体としての習得難度と、コロケーションの構成単語の習得難度の関係を調査し、将来的にコロケーションの習得難度についての包括的な研究を行うための基礎的知見を得ることを目指す。

コロケーションの習得難度とその構成単語の習得難度の関係について、2.1 で述べる EVP に基づいて例を挙げる。EVP では、a burning issue/question (‘a subject or question that must be taken care of or answered quickly’)というフレーズの CEFR レベル（最初級の A1 から最上級の C2 までの 6 段階）は C1 とされているが、その構成単語のレベルを見ると、burning (‘ON FIRE’)と issue (‘SUBJECT’)は B1、question (‘PROBLEM’)は B2 であり、構成単語のうちで最も高いレベル（B2）よりもコロケーションのレベル（C1）の方が高い。（ここでは burning の「燃えている」という意味をこのコロケーションの構成単語の意味とすることは必ずしも適切とは言えないが、EVP は単語としての burning はこの本来の意味でしか収録していないため、ここではこれを構成単語の意味として扱う。本来の意味とコ

ロケーションになった場合のレベルの違いも重要な研究課題ではあるが、本研究では扱わない。）もうひとつの例を挙げると、a fair chance (‘an equal and honest opportunity’)のレベルは C2 であるが、fair (‘EQUAL’)と chance (‘OPPORTUNITY’)のレベルはともに B1 であり、コロケーションと構成単語のレベルに大きな乖離がある。

筆者の知る限りでは、コロケーションの難度と構成単語の難度の関係についての包括的な調査はこれまでに行われていないことから、本研究では、この両者の関係について調査し、考察を行う。本研究では、CEFR レベル情報が付与された語彙リストである EVP を利用し、この中でコロケーションと考えられる項目を抽出し、項目全体の CEFR レベルと、構成単語の CEFR レベルの関係を調査する。なお、EVP は CEFR レベルの情報付きでコロケーションを収録しているが、EVP が収録するコロケーションの項目数は筆者の調査では約 900 に過ぎず、EVP が網羅的なレベル情報付きのコロケーションリストであるとは言えない。しかしながら、EVP はレベル情報の付いたコロケーションリストとして代替のない有用なリソースであることから、本研究では EVP の情報を利用する。

本研究における研究設問は次の 3 つである。

RQ1：コロケーションのレベルと構成単語の中で最も高いレベルの間には概略どのような関係があるか。

RQ2：構成単語のレベルよりもコロケーションのレベルの方が高い項目はどの程度の割合を占めるか。また、そのような項目に共通して見られる特徴はあるか。

RQ3 : コロケーションのレベルよりも構成単語のレベルの方が高い項目はあるか。また、それはどのような項目か。

これらの研究設問への解を検討することを通して、コロケーションとその構成単語の難易度の関係について基礎的な知見を得ることが本研究の目的である。コロケーションの習得難度を規定する要因の一端を明らかにし、コロケーションの習得難度を判定する包括的な方法を開発するための端緒を探ることができれば、学術的な観点でも言語教育の観点でも意義があると考える。

2. データと分析手法

2.1 本研究で利用したデータ

本研究では、コロケーションの習得難度情報を取得するために EVP を利用した。EVP は主に Cambridge Learner Corpus と Cambridge English Corpus に基づいて作成された語彙項目のリストで、Cambridge University Press によって無償提供されている。収録されている単語・フレーズに対して、学習者がどの CEFR レベルであれば正しく使えるかという観点で CEFR のレベルが付与されている（Capel, 2010, 2012）。CEFR レベル以外に、定義や辞書での用例、そして学習者が産出したライティングの例なども収録されている。形態的に同じ語や句であっても、意味ごとに分けて扱っている点が、後述の CEFR-J Wordlist とは異なる特徴である。EVP にはイギリス英語版とアメリカ英語版があるが、本研究では

アメリカ英語版を利用した。EVP は、コロケーションも難度情報付きで収録しているが、収録するコロケーションの項目数は、2.2.1 で述べるように約 900 項目に過ぎない（筆者の調査による）。しかしながら、より大規模で、広く利用されている難度情報付きのコロケーションのリストは筆者の知る限りでは存在しないため、本研究では EVP が収録するコロケーションを対象として分析を行う。

コロケーションを構成する単語の習得難度情報を取得するために利用したのは、CEFR-J Wordlist Version 1.6（東京外国語大学投野由紀夫研究室, 2020）（以下「CEFR-J Wordlist」）である。これは、CEFR を日本の英語教育環境に適用した CEFR-J（投野, 2013；投野・根岸, 2020）の一部を構成する単語リストであり、主に、見出し語・品詞・CEFR レベル（A1 から B2 までの 4 段階）の情報から成る。CEFR-J Wordlist は、東アジア圏（中国・台湾・韓国）の英語教科書を CEFR レベル別に分類し、各レベルで地域サブコーパスに共通して出現する単語を抽出し、さらに EVP も参照して補充するという方法で構築された（投野, 2013, pp. 131–132）。EVP でも単語単位の語彙項目が収録されているにも関わらず、本研究で CEFR-J Wordlist を使用した第一の理由は、EVP ではコロケーション中の構成単語が見出し語としては収録されていても、そのコロケーション中での語義が収録されていないことが少なくないため、コロケーション中での語義に対応する CEFR レベルを特定できない場合が多いためである。例えば、tell the difference (‘to notice a difference between two things’)（EVP の CEFR レベルは C1）の構成単語である tell は EVP では単語としては、‘to say something to someone, usually giving them information’の語義（A1）と‘If something tells you something, it gives you information.’の語義（B2）でしか収録されていない。これに対し、CEFR-J Wordlist では語彙項目・品詞ごとに CEFR レベルが付与されているため、単語が収録されていれば、その CEFR レベルを特定することができる。（実際には、EVP の実践に見られるように、同じ単語でも語義によって習得難度は異なり、本来はコロケーション中で使用されている語義に基づいて構成単語の習得難度を特定することが理想的であるが、現時点ではそれを可能にするリソースがない。）構成単語のレベル特定に CEFR-J Wordlist を利用した第二の理由は、CEFR-J Wordlist が日本の英語教育環境で広く認知され、使用が広がっていることから、他の語彙研究との比較調査の可能性を考慮したためである。

EVP と CEFR-J Wordlist は異なる方法で CEFR レベルを規定しているということは事実であるが、一定数のコロケーションをレベル情報付きで収録しているリストは EVP しかなく、しかしその EVP はコロケーション中での構成単語の語義のレベル情報を提供していない場合が多いという理由により、構成単語のレベルの判定には、EVP 以外の CEFR レベル付きの単語リストを利用せざるを得ない。そのような単語リストには、例えば Oxford 3000 などもあるが、本研究では日本の英語教育研究において広く利用されている CEFR-J Wordlist を利用した。

2.2 手法

2.2.1 EVP データ中のコロケーション項目の特定

EVP が収録する約 15,300 項目のうち、2 語以上からなるフレーズは約 4,100 項目ある。（項目数は筆者調査による。）このフレーズ項目を、下記の基準によりコロケーションとそれ以外に分類したところ、コロケーションは約 1,200 項目あった。なお、上記の項目数は、実質的に同じ項目が複数の異なる構成単語のエントリーに含まれているもの（例えば、a broken home は broken のエントリーと home のエントリーの両方に含まれている）を別項目として数え上げた項目数であり、これらの重複を除くと、コロケーションは約 900 項目となる。

フレーズをコロケーションと考えられる項目とそれ以外の種別に分類する際には、判断が難しい場合も少なくない。例えば、Ackermann and Chen (2013)は、Academic Collocation List を作成する際に、ある単語同士の結びつきをコロケーションとして収録するかどうか判断するために、頻度・MI スコア・t スコアという客観的な基準に加えて、それだけで意味的に完結しているか、固定的な複合語でないか、good のような一般的で意味的に透明な形容詞（common transparent adjectives）を含まないかなどの条件を設けて選別し、さらに専門家の検討（expert review）を行っているが、上記の条件を満たすかどうかの判断でも、専門家の検討結果でも揺れがあったことが報告されている。このことは、コロケーションとそれ以外を明確に区別することは極めて困難であることを示していると考えられるが、本研究ではコロケーションとそれ以外を区別することが主眼ではなく、コロケーションと考えられる項目とその構成要素の習得難度を分析することが目的であるため、以下の基準により、筆者と研究補助者 1 名の判断によりコロケーションか否かの分類を行った。

- コロケーションは、決定詞（sb's を含む）を除いた上で、典型的な形容詞＋名詞や、動詞＋副詞、名詞＋前置詞（of・out of を含む）などの 2 語から成る組み合わせ、あるいはこの組み合わせが複数結合したもの（例えば in the vicinity (of sth)）に限定する。Ackermann and Chen (2013, p. 238)は、（彼らは内容語のコロケーションに限定しているものの、）Nesselhauf (2005)などの先行研究に言及しながら、このような 2 語の組み合わせをコロケーションとする定義が先行研究の方法と一致するということを述べている。ただし、arrive at an [*sic*] solution/conclusion/answer, etc.の中の arrive at のような句動詞は 1 語の動詞に相当するものとして扱い、be in the public eye の public eye のような複合名詞は 1 語の名詞に相当するものとして扱う。したがって、上記 2 例はいずれもコロケーションとみなす。（後者については、後述のように be を除いた上で判断する。）
- add to や back up (sth) or back (sth) up のような句動詞はコロケーションとはみなさない。
- あくまで EVP が提示する各項目の定義に基づいて判断する。
- be で始まる項目は原則として be を除いて判断する。（be out of breath や be poor in

sth のように通常 be 動詞の後の補部の位置で使用されるコロケーションは、EVP では be で始まる形で収録されていることが少なくないものの、これらのコロケーションでは be は重要な要素ではないことから、be を除いて前置詞＋名詞などのコロケーションの形になるかどうかを判断する。）

- sth や sb は様々な語句が入り得る可変部分を示すものであるため、除いて判断する。you も sb の代わりに用いられていると考えられる場合は除いて判断する。
- not believe your ears (‘to be very surprised when you hear what someone says’)のように not＋動詞で始まる項目は、原則として not を除いて判断する。
- Honduran/Canadian/Indian, etc. by birth のように比較的自由度の高い可変部分を含む場合は、可変部分はあくまで具体的な表現の例として挙げられていると考えられるため、可変部分を除いて判断する。したがって、この例の場合は by birth としてコロケーションとみなす。
- 単に動（名）詞または現在分詞が後続することを表す do(ing)は除いて判断する。
- 不定詞標識の to はコロケーションを構成する一語として扱う。従って、aim to do や be authorized to do などはコロケーションとみなす。ただし、be about to do や be going to do や have (got) to do sth は前置詞等＋to 不定詞のコロケーションとは言いがたいためコロケーションとはみなさない。
- be at liberty to do のような表現では、末尾の to do はコロケーションの重要な要素ではないことから、to do を除いて（さらに上記の通り最初の be も除いて）at liberty として判断する。したがって、be at liberty to do はコロケーションとみなす。
- 形式上は前置詞＋副詞などのコロケーションである at all・at first・before long などと、形式上は名詞＋前置詞のコロケーションである a number of sth (‘several’)などは、他の語の選択の余地がほぼない定型表現であるため、コロケーションとはみなさない。（固定的な表現をコロケーションとして扱わないという方針は、Ackermann and Chen (2013, p. 239)で採用された方法と共通である。（ただし、彼らは内容語のコロケーションに限定している。））
- (by) herself・beside yourself・help yourself のような再帰代名詞を必須要素として伴うフレーズはコロケーションとはみなさない。
- kind of (‘used to describe something approximately’)のように語用論的な機能を持ったイディオム、be on cloud nine (‘to be extremely happy and excited’)・be on the ball (‘to be quick to understand and react to things’)・a nest egg (‘a sum of money that has been saved or kept for a specific purpose’)・get the picture (‘used to say that someone understands a situation’)などの形式上はコロケーションであってもイディオム（的性質が強い）と考えられる表現はコロケーションとはみなさない。
- 判断に迷うものは Oxford Collocations Dictionary（McIntosh, 2009）と Longman Collocations Dictionary（Mayor, 2013）のうちのいずれかのコロケーション辞典に収

録されていればコロケーションとみなす。

EVP が収録するフレーズ項目がコロケーションであるかを判断する際に、上記のように一部の語を削除しているが、削除対象の単語は、様々な要素が入り得る可変部分を除くと、冠詞・my などの人称代名詞所有格・be・not・do(ing)・不定詞標識 to などであり、これらはいずれも CEFR-J Wordlist では A1 レベルであるため、3.1 で述べる、構成単語のうち最も高いレベルの調査には影響しない。

2.2.2 コロケーションの構成単語レベルの特定

コロケーションの構成単語の CEFR レベルの調査に当たっては、下記の基準（使用した基準の中で主なものを抜粋）で調査対象とする単語を特定し、該当する語のコロケーション中での品詞に基づいて CEFR-J Wordlist で CEFR レベルを調査した。

- 括弧内の語句も構成単語とみなす。例えば、in addition (to)では in と addition と to を構成単語とみなす。
- 「/」と「;」による並列がある場合は全項目を構成単語とみなす。
- 冠詞（a・an・the）・不定詞標識 to・be・sb('s)・sth・人称代名詞・数字（80s 等を含む）・単に動（名）詞または現在分詞が後続することを表す do(ing)・, etc.等はレベル調査の対象となる構成単語とみなさない（2.2.1 参照）。
- 名詞複数形は単数形に修正する。
- CEFR-J Wordlist にない動詞屈折形は原形に修正する。
- out of は CEFR-J Wordlist に従って 1 語扱いとして処理する。
- 名詞の後に(s)が付いている場合は単数形として処理する。

以上の基準に基づいて作成したデータの一部を表 1 に示す。

表 1

コロケーションと構成単語のレベル調査の結果（一部）

コロケーション	CEFR レベル	構成単語と CEFR レベル
a bad/rough, etc. patch	C2	bad_A1, rough_B1, patch_B2
adverse effect(s)/ reaction(s)/conditions	C2	adverse_B2, effect_A2, reaction_B2, condition_A2
at sb's disposal	B2	at_A1, disposal_B2
be at liberty to do sth	C2	at_A1, liberty_A2
be friends (with sb)	B1	friend_A1, with_A1
go/run deep	C2	go_A1, run_A1, deep_B2
keep sb posted	C1	keep_A1, post_A2

注：品詞の表示は省略している。

3. 結果と考察

本節では、2 節で述べた手法により得られた結果を示しながら、1 節で挙げた研究設問への解答を試みる。

3.1 コロケーションのレベルと構成単語のレベルの関係の概要

コロケーションを含むフレーズのレベルは、単純に考えれば、その構成単語のレベルの中で最も高いものと一致すると言える。例えば、at ease というコロケーションを使えるためには、その構成単語である at と ease の両方を使える必要があるという考え方である。at ease は EVP では B2、構成単語である at と ease は CEFR-J Wordlist ではそれぞれ A1 と B2 となっており、上記の考え方で説明が可能である。しかし、実際には、1 節で例示したように、コロケーションのレベルと構成単語のレベルの間に乖離がある例は少なくない。この両者の関係について検討するために、両者のレベルを集計した結果が表 2 である。

コロケーションを構成する単語の中には、a handful of の handful や、cramp sb's style の cramp のように、CEFR-J Wordlist に収録されていないものもあった。また、endeavor to do sth の endeavor や shed tears の shed のように、当該の品詞が CEFR-J Wordlist に収録されていないものもあった（名詞の endeavor と名詞の shed は収録されている）。これらは、表 2 の集計結果では除外されている。その結果、表 2 に含まれている項目の総数は 797 である。

表 2

コロケーションのレベルと構成単語の中で最も高いレベルの集計結果

コロケーションのレベル	構成単語の最高レベル			
	A1	A2	B1	B2
A1	<u>7</u>	5	0	0
A2	23	<u>10</u>	8	1
B1	48	43	<u>26</u>	6
B2	53	67	71	<u>24</u>
C1	41	40	40	27
C2	45	73	80	59

注 : 下線は、コロケーションのレベルと構成単語の最高レベルが一致している場合を表す。

表 2 で下線を引いた対角線上のセルが、コロケーションのレベルと構成単語の最高レベルが一致している項目の数を表している。例えば、コロケーションのレベルが B1 で、構成単語の最高レベルも B1 であるものは 26 項目あり、その中には a wide range/selection/variety, etc.（wide と range が A2 で selection と variety が B1）や shake hands（shake が B1 で hand が A1）などがある。これらのレベルが一致する項目は全部で 67 項目であ

り、全体のうちの 8.4%に過ぎない。

表 2 の下線部より右上のセルは、コロケーションのレベルよりも構成単語の最高レベルの方が高い項目の数であるが、これらの項目数は合計で 20 項目であり、全体のうちの 2.5%に過ぎない。この群については 3.3 で論じる。

コロケーションのレベルと構成単語の最高レベルの関係において圧倒的多数を占めるのは、前者が後者よりも高い項目であり、表の下線部より左下に位置するこれらの項目は合計で 710 項目あり、全体のうちの 89.1%を占める。この群については 3.2 でもう少し詳しく論じる。

これらの結果から、RQ1 への解は、次のようになる。

RQ1：コロケーションのレベルと構成単語の中で最も高いレベルの間には概略どのような関係があるか。

RQ1 への解：ほとんどの場合、コロケーションのレベルは、その構成単語の中で最も高いレベルよりも高い。

コロケーションのレベルがその構成単語のレベルよりも高くなることが大半である理由としては、いくつか考えることができる。第一の理由は、理論上は可能であっても慣習上用いられない語を組み合わせるのではなく、言語共同体の慣習上自然と感じられるコロケーションを使えるということは、単にその構成単語を使えるというレベルを超えた言語知識であると考えられるということである。つまり、単語を使えたとしても、その単語を適切なコロケーションで使えるとは限らないということである。例えば、EVP では B1 レベルとされている by hand (‘done or made by a person instead of a machine’)を A1 レベルの学習者が正しく使えるとは考えにくいが、構成単語のレベルとしては、前置詞の by も名詞の hand も、CEFR-J Wordlist では A1 である。by hand というコロケーションを使えるということと、単に by も hand も使えるということには隔たりがあり、その隔たりこそがコロケーションとしての難しさであると考えられる。第二の理由は、第一の理由とも関係するが、一般にフレーズの頻度は単語の頻度よりも大幅に低いため、現実の言語使用場面においてコロケーションに遭遇する機会が、構成単語に遭遇する機会よりも大幅に少ないために習得がより困難であるということである。非母語話者にとって、コロケーションの習得は難しいということが多くの先行研究で述べられてきている（堀, 2009 ; Henriksen, 2013 ; Wolter & Gyllstad, 2013）が、これは上記の理由によるところが大きいと考えられる。

3.2 構成単語のレベルよりもコロケーションのレベルの方が高い項目

3.1 で見たように、表 2 で下線部の左下に位置する数値が、構成単語の最高レベルよりもコロケーションのレベルの方が高い項目の数であり、全体のうちの 89.1%という大多数がこの群に属している。レベル差が 2 段階以上あり、顕著な差があると言えるものに限っても、構成単語のレベルが A1 である一方でコロケーションのレベルが B1 であるものは 48 項目、構成単語のレベルが A2 以下である一方でコロケーションのレベルが B2 であるもの

は 120 項目、構成単語のレベルが B1 以下である一方でコロケーションのレベルが C1 であるものは 121 項目、構成単語のレベルが B2 以下である一方でコロケーションのレベルが C2 であるものは 257 項目もある。これらを合計すると 546 項目となり、全 797 項目中の 68.5%を占め、コロケーションのおよそ 3 分の 2 が構成単語のレベルよりもコロケーションのレベルの方が 2 段階以上高いものであるということになる。

ここでは、特にレベル差が顕著である、構成単語のレベルが A1 でコロケーションのレベルが C2 である 45 項目のうち、20 項目に注目し、いくつかのタイプに分類を試みる。複数のタイプにまたがると思われるものもあるが、最もよく当てはまると考えられるタイプに分類した。

- 構成単語中に多義語があり、その意味がやや抽象的であるもの
  a broken home (‘a family in which the parents do not now live together’)、a run of sth (‘when something happens several times without something different happening during that period’)、a suggestion of/that sth [*sic*] (‘something that makes something seem likely to be true’)、be foreign to sb (‘to be something you know nothing about or do not understand’)、be hard on sb (‘to make someone unhappy by causing them problems’)、beside the point (‘not important or not connected with what you are talking about’)、break free (‘to suddenly escape or become separate from something or someone’)
- イディオムあるいは固定表現的なもの
  a tight corner (‘a difficult situation’)、be looking to do sth (‘to plan to do something’)、be running at sth (‘to be at a particular level’)、be under way (‘to be already happening’)、be/feel under the weather (‘to feel ill’)
- 比喩的な意味を表すもの
  at heart (‘used to say what someone is really like’)、at/by your side (‘If you have someone at/by your side, they support you and help you.’)、be hungry for sth (‘to have a strong wish for something’)、be news to sb (‘to be a surprise to someone’)、behind sb’s back (‘If you do something behind someone’s back, you do it without them knowing, often in an unfair way.’)、blow your mind (‘If something blows your mind, you are very excited or surprised by it.’)、catch sb’s eye (‘to get someone’s attention by looking at them’)
- 理解をする上ではそれほど難しくないと思われるが、学習者が使用する頻度が低いためにコロケーションとしては C2 レベルと判断されたのではないかと推察されるもの
  be slow to do sth; be slow in doing sth (‘to take a long time to do something’)

以上の分類は筆者の主観的な分類であり、分類の方法やそもそものタイプの設定について、多様な考え方が可能であると思われる。しかしながら、構成単語の基本的な意味を足し合わせるだけでは、コロケーション全体の意味を正しく推測することが難しいものが多い

ということは言えそうである。

以上の考察に基づき、RQ2 への解は、次のようになる。

RQ2：構成単語のレベルよりもコロケーションのレベルの方が高い項目はどの程度の割合を占めるか。また、そのような項目に共通して見られる特徴はあるか。

RQ2 への解：構成単語のレベルよりもコロケーションのレベルの方が高い項目は約 90%を占める。そのようなコロケーションは、たいていの場合、構成単語の意味を足し合わせて全体の意味を推測するのが難しいものである。

3.3 コロケーションのレベルよりも構成単語のレベルの方が高い項目

3.1 で見たように、表 2 で下線部の右上に位置する数値が、コロケーションのレベルよりも構成単語の最高レベルの方が高い項目の数である。特徴を明らかにするために、いくつかの項目を具体的に見ていく。

コロケーションのレベルが A1 で、構成単語の最高レベルが A2 である項目は以下の 5 つである。

get here/there/home/to work, etc.は、副詞の home が A2（他の構成単語は全て A1）であるためこの結果になったが、直感的には易しい語であり、here・there・home・to work が並列でまとめて提示されていることもあり、コロケーションのレベルが A1 となっていることに違和感はないと感じられる。なお、CEFR-J Wordlist と同様に、EVP でも単語としての副詞の home (‘to the place where you live’)は A2 となっていて、同様のレベルのずれが見られる。

last night と last week/year/Monday, etc.、そして next week/year/Monday, etc.については、いずれも決定詞の last と next が A2（他の構成単語は全て A1）であるためこの結果になったが、一方で、時間表現としての last night などは、A1 レベルでも使用できる学習者は少なくないと思われ、これらのコロケーションのレベルが A1 となっていることにさほどの違和感はないと思われる。なお、EVP でも上記の last・next とかなり近いと思われる last (‘the most recent’)（用例としては What was the last movie you saw?と It rained for the last three days.が挙げられている）と next (‘The next time, event, person, or thing is the one nearest to now or the one that follows the present one.’)のレベルは A2 となっているが、last night などの表現の難度を考えると、これらの単語が A1 から A2 にまたがるものであるということが示唆される。

take a picture/photo(graph)は、photograph が A2（他の構成単語は全て A1）であるためこの結果になったが、photo は A1 である。EVP でも単語としての photo は A1 であるが、photograph は A2 となっていて、上記のコロケーションと同様に、コロケーションのレベルが A1 となっていることに違和感はないと感じられる。

コロケーションのレベルよりも、構成単語の最高レベルが 2 段階高いものとしては、spare time (‘time when you are not working’)がある。EVP での spare time のコロケーションと

しての CEFR レベルは A2 であるが、CEFR-J Wordlist での形容詞 spare の CEFR レベルは B2 となっている。ただし、EVP では形容詞 spare (‘If something is spare, it is available to use, because it is extra and not being used.’)（用例としては a spare key/tire などが挙げられている）の CEFR レベルは B1 となっていて、この意味であるとすれば、レベルのずれは依然として存在するものの、その差は 1 段階となる。EVP 作成の基礎となった Cambridge Learner Corpus では spare time は A2 レベルの学習者が使っていることから、このコロケーションのレベルは A2 と判断されたと推測されるが、spare という単語を様々な他の語と組み合わせて使えるようになるよりも、先に spare time というコロケーションのフレーズでだけ使えるようになるということも十分あり得ると考えられ、そう考えれば、このレベルのずれも、大きな違和感があるようなものではないと思われる。

コロケーションのレベルが A2 で構成単語の最高レベルが B1 である項目と、コロケーションのレベルが B1 で構成単語の最高レベルが B2 である項目についての詳細は割愛するが、概して上記と同様の傾向が認められる。

以上の結果から、RQ3 への解は、次のようになる。

RQ3 : コロケーションのレベルよりも構成単語のレベルの方が高い項目はあるか。また、それはどのような項目か。

RQ3 への解 : わずかながら存在するが、構成単語単体で見ればややレベルが高くても、コロケーションとしてはそれほど難度が高くないものが多い。

コロケーションのレベルよりも構成単語の最高レベルの方が高い項目の割合は 2.5%に過ぎないということも考え合わせると、コロケーションのレベルはその構成単語のレベルを上回ることは基本的にはないと考えて差し支えないであろう。

4. まとめと今後の課題

本研究では、コロケーションとその構成要素である単語の難易度の関係について基礎的な知見を得ることを目的として、コロケーションの CEFR レベルと、その構成単語の CEFR レベルの関係を調査し、これまで明らかになっていなかった両者の関係の一端を明らかにした。ほとんどの場合、構成単語のレベルよりもコロケーションのレベルの方が高いことが明らかになったが、その理由としては、単語の知識に加えて、コロケーションとしての語彙選択の知識が必要になるということと、コロケーションの頻度が単語の頻度よりも低いことが挙げられる。また、構成単語のレベルとコロケーションのレベルの差が大きいコロケーションでは、構成単語の意味を足し合わせて全体の意味を推測するのが難しい場合が多いと考察した。

本研究ではコロケーションの構成単語の意味を特定する際に CEFR-J Wordlist を利用したが、コロケーションのレベルを決定する際に参照した EVP を利用し、コロケーション中の単語の語義に近い場合の CEFR レベルを確認することで、より精緻な分析が可能になると考えられ、今後この課題に取り組む予定である。ただし、2.1 で述べたように、コロケー

ション中での構成単語の語義が EVP には収録されていないことも少なくないことから、全てのコロケーション項目について、構成単語の EVP における CEFR レベルの情報を取得することはできないため、分析可能な項目数は減少すると予想される。

また、構成単語のレベルよりもコロケーションのレベルの方が高いコロケーションについては、コロケーションのフレーズ全体としての意味的透明性、つまり構成単語の意味からコロケーションの意味をどの程度予測できるかという観点で個々のコロケーションを評価し、コロケーションと構成単語のレベルの差と相関が見られるのかどうかも検討したい。

既述のように、コロケーションを含むフレーズの頻度は、単語と比べると大幅に低い。生起頻度が低いフレーズの頻度を調査するためには相当に大規模な CEFR レベル別のコーパスが必要であるが、近い将来にそのような CEFR レベル別の大規模なコーパスが利用可能になるという可能性は高くない。したがって、現実的にコロケーションの習得難度を判断するためには、コーパス頻度とは異なる観点も用いる必要がある。本研究が取り組んだコロケーションと構成単語の習得難度の関係は、コロケーションの習得難度を判断する手法というより大きな研究の枠組みの中で一定の意味を持つと考えられる。

謝辞

本研究は、2020～2021 年度成城大学特別研究助成「コロケーションの CEFR レベル判定のための基礎研究」、2022～2023 年度成城大学特別研究助成「コロケーションと構成単語の CEFR レベルの関係についての研究」、JSPS 科研費「コロケーション・チャンクを含む高頻度フレーズの CEFR レベル別リストの作成と公開」（JP22H00677；研究代表者　内田諭）の助成を受けたものである。ここに記して謝意を表する。また、本稿の初稿を丁寧に読み、適切な助言をくださった 3 名の匿名の査読者に心からの謝意を表する。

引用文献

Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, *12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Browne, C., Culligan, B., & Phillips, J. (2013). The New General Service List. https://www.newgeneralservicelist.org

Cambridge University Press. (2015). English Vocabulary Profile. https://www.englishprofile.org/wordlists

Capel, A. (2010). A1–B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. *English Profile Journal, 1*(1). Cambridge University Press. https://doi.org/10.1017/S2041536210000048

Capel, A. (2012). Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal, 3*(1). Cambridge University Press.

https://doi.org/10.1017/S2041536212000013

Chen, Y.-H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14*(2), 30–49. https://doi.org/10125/44213

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://rm.coe.int/16802fc1bf

大学英語教育学会基本語改訂特別委員会（2016）『大学英語教育学会基本語リスト 新JACET8000』 桐原書店.

Davies, M., & Gardner, D. (2010). *A frequency dictionary of contemporary American English*. Routledge. https://doi.org/10.4324/9780203880883

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Blackwell. https://doi.org/10.1002/9780470756492.ch4

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–160). Clarendon Press.

Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 29–56). European Second Language Association. https://www.eurosla.org/monographs/EM02/Henriksen.pdf

堀正広（2009）『英語コロケーション研究入門』 研究社.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Clarendon Press.

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Language Teaching Publications.

Mayor, M. (Ed.). (2013). *Longman collocations dictionary and thesaurus*. Pearson Education.

McIntosh, C. (Ed.). (2009). *Oxford collocations dictionary for students of English* (2nd ed.). Oxford University Press.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins. https://doi.org/10.1075/scl.14

Rundell, M. (Ed.). (2010). *Macmillan collocations dictionary*. Macmillan Education.

東京外国語大学投野由紀夫研究室（2020）『CEFR-J Wordlist Version 1.6』 https://www.cefr-j.org/download.html#cefrj_wordlist

投野由紀夫（編）（2013）『CAN-DO リスト作成・活用 新しい英語到達度指標 CEFR-J ガイドブック』 大修館書店.
投野由紀夫・根岸雅史（編）（2020）『教材・テスト作成のための CEFR-J リソースブック』 大修館書店.
塚本倫久（2013）『プログレッシブ 英語コロケーション辞典』 小学館.
Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451–482. https://doi.org/10.1017/S0272263113000107
八木克正（監修）・樋口昌幸・住吉誠（編）（2015）『小学館オックスフォード 英語コロケーション辞典』 小学館.