

# 日本語母語英語学習者が使用する定型表現の分析

—「話すこと[やりとり]」と「話すこと[発表]」の技能育成をめざして—

研究者:小出 凱渡 茨城県／筑西市立明野中学校 教諭・茨城県／筑波大学大学院 在籍

《研究助言者:竹内 理》

## 概要

本調査では、定型表現の一種である高頻度語連鎖 (Lexical Bundles) を観点に、モノログとダイアログという話し言葉の2つの形態における日本語母語英語学習者の定型表現の使用実態を明らかにすることを目的とした。英語母語話者と学習者の話し言葉コーパスを比較し、学習者が過剰・過少使用する語連鎖を特定してリスト化するとともに、それらの語連鎖の特徴について分析・考察を行った。その結果、モノログでは、学習者は特定の句動詞を含む語連鎖や理由を述べる際に用いる表現を多用する傾向が見られた。一方、母語話者と比較して、学習者は多様な句動詞表現や助動詞を構成要素とする語連鎖をあまり使用していないことが明らかとなった。ダイアログでは、学習者は同じ発話の繰り返しで構成される語連鎖やフィラーを含む表現を過剰使用していた一方で、would like toや多様な時制表現を示す語連鎖を過少使用していることが明らかとなった。本調査の結果は、話し言葉における英語学習者の語連鎖の使用実態について有益な情報を提供し、定型表現を効果的に活用した「話すこと[やりとり]」及び「話すこと[発表]」の技能育成に示唆を与えるものである。

## 1 はじめに

近年、主にコーパスを活用した言語研究によって、話し言葉や書き言葉の多くが2語以上のまとまりからなる複数語で構成されていることが明らかにされてきた (Altenberg, 1998; Biber et al., 1999; Erman & Warren, 2000; Foster, 2001)。このように、自然談話内において一定の単語の組み合わせで繰り返し用いられる言語パターンは、定型表現 (formulaic language) と呼ばれており、外国語教育研究においても近年ますます注目を集めている (金澤, 2020)。定型表現は言語処理や言語習得と深く関わりがあり、正確で適切かつ流暢な言語使用に欠かせないものである (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Schmitt & Carter, 2004; Wray, 2002)。定型表現は母語話者だけではなく、学習者にとってもその有用性が明らかにされており (Conklin & Schmitt, 2008; Isobe, 2011; Jiang & Nekrasova, 2007)、第二言語習得における定型表現の重要性が示唆されている。中田 (2022) は、英語学習者も定型表現の知識を身につけることで多くのメリットを享受することができるとし、言語使用における正確性・流暢性が向上することに加え、状況に応じた適切な言語使用ができるようになること、すでに知っている単語の知識が深まること、文法知識の習得が高まることなどを挙げている。

このように、定型表現は学習者にとって学習・習得の必要性が高いことは明らかであるものの、学習者にとって定型表現の習得・使用は必ずしも容易なものではない。先行研究では、学習者の使用する言語は母語話者と比べ定型性が低く (Granger, 1998; Howarth, 1998; Li & Schmitt, 2009)、学習者は、言語産

出の場面に応じて定型表現を適切に使い分けることができないことが報告されている (Chen & Barker, 2010, 2016)。このような問題を解決し、学習者が定型表現を効率的に習得するためには、学習者言語における定型表現使用の実態を詳細に調査し、その特徴を踏まえた指導・学習を行っていくことが重要である。

定型表現には、イディオムや句動詞、コロケーションなど複数のタイプが存在するが、その中の1つに高頻度語連鎖 (Lexical Bundles) がある。これまでの研究では、書き言葉における日本語母語英語学習者の語連鎖の使用 (石川, 2019) や日本の英語検定教科書に見られる語連鎖の扱いについて調査したもの (Mikajiri, 2022; Northbrook & Conklin, 2018) はいくつか存在するが、話し言葉における日本語母語英語学習者の語連鎖の使用実態を調査した研究は非常に限られている。また、現在の学習指導要領では、「話すこと」が [発表] と [やりとり] の2つの領域に分けられ、別々に目標や指導内容が定められている。これは、一人の話者が連続して話す場合と、複数の話者が相互に話す場合で、「話すこと」の特性が異なるためである (文部科学省, 2017b, 2018)。これらを踏まえ、本調査は、話し言葉を聞き手に対して一方向で話すモノローグ (Monologue) と、聞き手と話し手の双方向のやりとりであるダイアログ (Dialogue) の2つに区別し、コーパスを用いた先行研究の方法論に基づきながら、日本語を母語とする英語学習者の語連鎖使用の実態とその特徴を明らかにすることを目指した。

## 2 先行研究

### 2.1 言語使用における定型表現の役割

定型表現 (formulaic language) とは、複数の単語が組み合わせられ、1つのまとまりとして用いられる結束性の高い語の連鎖と定義され (Wray, 2008)、イディオムや句動詞、コロケーションなどが存在する。定型表現は、言語使用や言語学習において重要な役割を果たすと考えられている (Conklin & Schmitt, 2012; 金澤, 2020; 中田, 2022; Siyanova-Chanturia & Pellicer-Sánchez, 2019)。その理由として、第一に、定型表現の普遍性がある。Howarth (1998) では、およそ24万語の学術テキストを収集したコーパスを分析したところ、約30%から40%がコロケーション (e.g., blow the fuse) やイディオム (e.g., blow the guff) といった単語連鎖から構成されていることが明らかとなった。また、Erman and Warren (2000) によると、談話中に定型表現が占める割合は書き言葉で52.3%、話し言葉においては58.6%に及ぶとされている。Conklin and Schmitt (2012) では、先行研究で得られた知見を踏まえ、定型表現は会話の3分の1から半分程度を構成していると報告している。このように、自然談話内において定型表現は多くの割合を占め、言語を使用する上で必要不可欠なものであると考えられている。

第二に、定型表現は言語処理を効率化し、適切で流暢な発話を可能にするとされる。Jiang and Nekrasova (2007) は、英語母語話者とトルコ語や中国語、アラビア語など多様な言語的背景を持つ英語学習者を対象に文法性判断課題を実施したところ、定型表現 (e.g., to tell the truth) の方が、非定型表現 (e.g., to tell the price) よりも反応時間が短く、正答率が高いことがわかった。この結果から、母語話者や学習者にかかわらず、定型表現は1つの語彙項目として心的辞書に保有され、ひと固まりとして認識されることによって各構成語に注意を払う必要がなくなり、より短い時間で正確に言語処理がなされることが示唆された。Conklin and Schmitt (2008) は、英語母語話者と第二言語 (L2) として英語を学ぶ協力者を対象に、自己ペース読み課題を実施し、定型表現を含む文とそうでない文ではどちらが速く読解されるかについて検証した。分析の結果、母語話者と学習者ともに、定型表現を用いた文の方がより速く読解されたことから、定型表現は言語処理上の利点 (processing advantage) を持つと結論づけた。また、中田 (2022) は、定型表現の知識は、言語使用における流暢性 (fluency) を高め、正しい文を淀みなく流暢に産出したり、素早く内容を理解したりする上では不可欠であるとしている。さらに、定型表現を使用することで、話している相手や状況に応じた適切な言語使用が可能となり、The point is that や Help yourself

といった特定の機能を持つ定型表現を用いることで、効果的にメッセージを伝えたり、理解したりすることができるとして、言語使用における定型表現の重要性を指摘している。

以上のように、これまでに様々な研究から、言語使用における定型表現の重要性が示されてきている。

## 2.2 高頻度語連鎖 (Lexical Bundles) に着目した定型表現研究

定型表現と聞くと、イディオムや句動詞などをイメージすることが多いが、これまでの研究では、コーパスを活用し、その中に見られる高頻度語連鎖 (Lexical Bundles) を対象に、話者の使用する定型の言語表現の特徴を明らかにしようとしたものが多く行われている (e.g., Biber et al., 1999; Biber et al., 2004; Biber & Barbieri, 2007)。Lexical Bundles とは、ある談話内において、繰り返し用いられる結束性の高い語の連鎖である (Biber et al., 1999)。

Biber and Barbieri (2007) は、スピーキングとライティングといった2つの産出モードにおいて、英語母語話者が使用する高頻度語連鎖について分析を行った。この研究では、大学の様々な場面における発話を収集したコーパスを用い、100万語の中で40回以上出現する4語の高頻度語連鎖を対象に調査がなされた。話し言葉コーパスには、大学での講義場面や教員と学生による面談の場面などで発話されたデータが含まれている。一方、書き言葉コーパスは、講義で使用する複数の教科書や参考書、大学のパンフレットなどから抽出されたデータで構成されている。分析の結果、話し言葉と書き言葉のどの場面においても高頻度で使用する語連鎖が観察され、特に話し言葉では、教師が課題や授業の活動について指示する表現 (e.g., you just have to, you need to do), 学生に与えられた課題を遂行するように指示する表現 (e.g., you should be able to), 講義でトピックや活動について言及する際に用いる表現 (e.g., I want to talk about, what we're going to), 教材に注意を向けさせる表現 (e.g., if you look at) などが抽出された。それに対し、書き言葉では、時間を示す表現 (e.g., the beginning of each, the end of each, at the time of) に加え、講義の内容や構成について導入する表現 (e.g., an introduction to the, the nature of the) が観察された。

Jalilifar et al. (2016) は、芸術や文学、応用言語学、科学といった異なる学問分野に見られる高頻度語連鎖について調査を行った。それぞれの分野の研究論文を収集したコーパスを用いて100万回に10回以上出現する3, 4, 5語連鎖を抽出したところ、3語連鎖では、as well as や the number of, based on the, in terms of などが全ての分野に共通して出現していた。また、4語連鎖では、at the same time や in the case of, on the other hand, 5語連鎖では at the end of the や in the case of the, as a result of the といった表現が抽出された。特に、この研究で観察された4語連鎖の頻度順上位50位に含まれる項目の多くが、同様に複数の学問分野における Lexical Bundles を調査した Hyland (2008) で見られた語連鎖と類似していることが明らかとなった。

## 2.3 学習者における語連鎖の使用実態

2.2に挙げた研究はいずれも母語話者を対象としたものであるが、L2学習者を対象とした同様の調査も行われている。Chen and Baker (2010) では、中国語を母語とする英語学習者を対象に、学習者が使用する語連鎖 (4語連鎖) の言語的機能や言語的構造について調査を行った。学習者コーパスと母語話者コーパスを用いて分析を行ったところ、母語話者と比較して学習者は、the extent to which といった関係詞節の一部を構成する名詞句、the + 名詞 + of + the や in + the + 名詞 + of などの of を含む語連鎖をあまり使用しないことが明らかとなった。また、受動動詞 + 前置詞句を構成する語連鎖である be taken into account や be used in the に加え、自分の意見や主張を和らげる表現である is likely to be や would have to be などの表現も過少使用していた。一方で、学習者は母語話者と比べ、all over the world といった特定の表現や、不定詞節を導く語連鎖である in order to achieve や to be able to を多用していること

がわかった。さらに、学習者の語連鎖使用の特徴として、in the long runのように英語母語話者が主に話し言葉で使用する表現を書き言葉で用いる傾向が観察された。

日本語を母語とする英語学習者を対象とした研究としては、石川(2019)がある。石川は、複数のコーパスを用いて日本の英語教育で使用されている教材や試験問題における語連鎖、学習者の語連鎖使用の特徴について分析を行った。その結果、学習者が使用する語連鎖は、教材や試験問題に現れる語連鎖と比べて、母語話者が使用する語連鎖と大きく乖離しており、使用傾向が異なっていることが明らかとなった。また、学習者の過剰使用語連鎖の大半は、英作文の際に与えられたトピック文の一部となっており、トピック文に含まれる表現をそのまま流用する傾向が観察された。トピック文にない語連鎖としては、意見や主張を導くI thinkの過剰使用が特徴的であった。一方で、学習者が過少使用していた語連鎖を見ると、one of the やsome of the, as well asといった物事の量や比率、要素の添加などを表すものなどが抽出されており、母語話者と比較して学習者はこれらの表現をあまり使用しない傾向が見られた。

上記2つの研究のように、学習者の書き言葉における語連鎖使用の実態を調査した研究は他にもいくつか存在するが(e.g., Ådel & Erman, 2012; Güngör & Uysal, 2016; 小出, 2024; Pan et al., 2013; Staples et al., 2013), 話し言葉について分析した研究は限られている。Yan(2019)は、中国語を母語とする英語学習者を3つの習熟度に分け、話し言葉で使用する語連鎖の特徴について習熟度別に調査を行った。この研究では、Test for English Majors(TEM)と呼ばれる中国で実施されている英語力を測定する試験において、受験者が発話したデータを対象として分析を行った。その結果、習熟度が低い学習者は、量や数、程度の大きさを表す語連鎖(quantifying bundles)である more and more people やthere are a lotなどを多く使用する一方で、習熟度の高い学習者は、pay more attention toやintegral part of theといったトピックに焦点を当てる表現に加え、we should have aやwe should learn fromなどのshouldを含む語連鎖を多用することが示された。さらに、I would like toやis the most importantなどの自分の考えを表す語連鎖、話者の意見の確信度を示すI think it isやI don't think, トピックを導入する際に用いる表現であるwe all know thatやfirst of all Iなどは、3つの習熟度において同程度の使用頻度であり、話し言葉において学習者が共通して使用する語連鎖であることがわかった。特に、この研究で抽出されたon the other handやat the same timeなどのいくつかの表現は、学習者の書き言葉を調査した他の研究においても同様に観察されており(Chen & Baker, 2016; Staples et al., 2013), 産出モードの違いにかかわらず、学習者が好んで使用する語連鎖であることが明らかとなった。

### 3 研究の目的

先行研究では、日本語母語英語学習者を対象に、書き言葉における学習者と母語話者が使用する定型表現(語連鎖)の比較・分析がコーパスに基づいて行われ、学習者の定型表現の使用実態が調査されている(石川, 2019)。しかしながら、話し言葉について同様の調査はほとんど行われておらず、日本語母語英語学習者の定型表現の実態を明らかにする上で検証の余地が残る点である。そこで、本調査では、日本語母語英語学習者が話し言葉で使用する語連鎖に焦点を当て、先行研究と同様にコーパスを用いたアプローチによって、その実態と特徴を明らかにすることを目的とした。

現在の日本の学校英語教育において、「話すこと」が[発表]と[やりとり]の2つの領域に分けられて指導が行われていることを踏まえ、本調査では、話し言葉を聞き手に対して一方向で話すモノローグ(Monologue)と、聞き手と話し手の双方向のやりとりであるダイアログ(Dialogue)の2つに区別し、それぞれの産出モードにおける学習者の定型的言語表現の分析を行うこととした。研究課題は以下の2つである。

RQ1

モノローグ (Monologue) において、日本語母語英語学習者が過剰・過少使用する語連鎖は、どのようなものがあるか。また、その特徴 (品詞、文法、語用論的機能など) は何か。

RQ2

ダイアログ (Dialogue) において、日本語母語英語学習者が過剰・過少使用する語連鎖は、どのようなものがあるか。また、その特徴 (品詞、文法、語用論的機能など) は何か。

## 4 研究の方法

本節では、調査に用いる各コーパスの基本情報、分析対象とする語連鎖の基準、そしてコーパス分析から得られたデータの分析手法について詳述する。

### 4.1 対象とするコーパスデータ

#### 4.1.1 ICNALE の概要

本調査では、ICNALE (International Corpus Network of Asian Learners of English) コーパスにおける発話データを分析対象とした。ICNALE は、アジア圏における10の国や地域の英語学習者 (大学 (院) 生) と英語母語話者およそ4,000人による発話や英作文データを収集した大規模コーパスであり、14,000以上のサンプル数で構成されている (Ishikawa, 2023)。参加者の習熟度を判定するため、全ての学習者には、Common European Framework of Reference (CEFR) の基準に基づき、A2, B1\_1, B1\_2, B2の4段階で習熟度情報が付与されている。ICNALE は5つのモジュールから構成されており、1分間の独話データを集めた Spoken Monologue (SM)、インタビュー形式の対話を集めた Spoken Dialogue (SD)、英作文データを収集した Written Essays (WE)、学生の作文を専門家が校閲した Edited Essays (EE)、学生の作文・発話の評価データを収集した Global Rating Archive が含まれている (Ishikawa, 2023)。

本調査で ICNALE を分析対象とした理由は、以下の2つである。1つ目は、言語の産出条件が統制されている点である。ICNALE の全モジュールにおいて、学習者と母語話者は、「大学生のアルバイトの是非 (It is important for college students to have a part-time job.)」及び「レストラン全面禁煙の是非 (Smoking should be completely banned at all the restaurants in the country.)」という2つのトピックについて作文もしくは発話をし、データが収集されている。また、詳細は後述するが、SM や SD では、全ての参加者が同一条件の下で言語産出ができるように、事前に定められた順序に従いながらデータ収集がなされている。このような ICNALE の特性は、母語話者と学習者での比較や異なる母語を持つ学習者間での比較、習熟度別による比較といった対照分析の妥当性を高め、本調査のように、母語話者と比較して学習者言語の特徴を解明しようとする際に適したコーパスであるといえる。

2つ目は、2種の話し言葉モジュールが含まれている点である。ICNALE の話し言葉データには、話者が単方向で話した独話データである SM と、双方向のやりとりを集めた対話データである SD の2つがある。これらは、本調査で示唆を与えたいと考える「話すこと [発表]」と「話すこと [やりとり]」の特性を反映したものであると考える。以下、これら2つのモジュールにおけるデータを Ishikawa (2023) と石川ほか (2020) に基づいてさらに詳しく述べる。

#### 4.1.2 ICNALE Spoken Monologue

ICNALE Spoken Monologue は、1,100人の参加者に、上述の「大学生のアルバイトの是非」と「レストラン全面禁煙の是非」という2つのテーマについて、それぞれ2回、1分間のスピーチを行わせた独話データを含む (Ishikawa, 2023)。まず、参加者は指定された電話番号に電話をかけ、システムから自動で流れる指示に従い、自己紹介発話 (60秒) を実施する。次に、1つ目のテーマである「大学生のアルバイトの是非」

について、第1発話(20秒準備後に60秒で実施)を行い、その後、第2発話(10秒準備後に60秒で実施)をする。ここで2度発話を行う理由は、緊張により第1発話で本来のパフォーマンスが発揮できない参加者がいる可能性を考慮したためである。自己紹介の発話はコーパスに含まれず、第1発話と第2発話がコーパスに含まれている。最後に、2つ目のトピックである「レストラン全面禁煙の是非」について、「大学生のアルバイトの是非」の発話手順と同様に、第1発話、第2発話を連続して行う。全ての参加者は同一条件の下、発話を行っており、発話データを収集後、それらの音声の書き起こしテキストを構築している。本調査では、その中の日本語母語英語学習者( $n = 150$ )と英語母語話者( $n = 150$ )の全データ(第1発話と第2発話)を調査対象とする。

#### 4.1.3 ICNALE Spoken Dialogue

ICNALE Spoken Dialogue は、参加者425名に対してインタビューを行い、その発話を書き起こしたデータである(石川ほか, 2020)。インタビューの進め方は詳細なプロトコルブックによって定められており、①導入対話→②「大学生アルバイト」関連タスク→③「レストラン禁煙」関連タスク→④振り返りの手順で進み、計4部で構成されている。①は、インタビューの導入部分であり、英語学習などをテーマとした質疑が行われる。

②と③では、それぞれ、ICNALEの共通テーマである「大学生のアルバイト」と「レストランでの禁煙」に関連した課題が設定されており、(a)イラスト描写、(b)イラスト関連質疑、(c)ロールプレイ、(d)ロールプレイ関連質疑の4つのタスクで構成されている。③を例にすると、(a)のイラスト描写では、公園でタバコを吸っている男性に対して、子連れ母親が注意して公園内での喫煙を止めさせるという内容の6コマ漫画を時系列で描写させるというタスクが与えられる。(b)では、漫画の内容に関連した質問(公園に行くのは好きか、好きな公園はどこかなど)を提示して、参加者に回答させる。(c)では、インタビュワーと参加者がそれぞれ、レストランの店主役とその客役を演じてロールプレイを行う。参加者は、レストランでのタバコの臭いに耐えられず店を出ざるを得なかったことを抗議し、店主に返金を求める客として、インタビュワーを説得するという課題が与えられる。(d)では、ロールプレイの内容に関連した質問(どんなレストランが好きか、レストランの全面禁煙についてどう思うかなど)について参加者に答えさせる。最後に、④では、(a)英語による振り返り質疑と、(b)母語による振り返り質疑を続けて行わせる。なお、SDの日本モジュールに関しては、1名の非英語母語話者の教員が100名全てのインタビューを行っており、これにより、調査者の性別や外国語知識などといった違いによるインタビューへの影響が抑制され、データの統制性の向上にも寄与すると考えられる(石川, 2018)。

SDデータのうち、本調査では、日本語母語英語学習者( $n = 100$ )と英語母語話者( $n = 20$ )の発話データを調査対象とする。また、本調査では、関連タスクのうち、(b)イラスト関連質疑、(c)ロールプレイ、(d)ロールプレイ関連質疑の計6部のデータのみを分析対象とし、②と③の(a)イラスト描写と、①導入対話と④振り返りの4つのパートを分析対象から除外した。これらのデータを分析対象外とした理由として、①と④は、インタビューの導入や振り返りのパートであり、実際に2つのテーマについてのタスクを行っている発話ではないこと。また、②と③の(a)は、与えられた課題についての発話ではあるが、参加者が一方的にイラストについて描写しており、モノログの特性に近いデータであると判断したためである。以上2種のコーパスの基本情報に関しては、表1を参照されたい。

■表1: 本調査で分析対象とした各コーパスの基本情報

|        | Spoken Monologue       | Spoken Dialogue                  |
|--------|------------------------|----------------------------------|
| 形式     | 独話型                    | 対話型                              |
| 参加者    | 日本語母語英語学習者<br>150名     | 日本語母語英語学習者<br>100名               |
|        | 英語母語話者<br>150名         | 英語母語話者<br>20名                    |
| トピック   | 大学生のアルバイト<br>レストランでの禁煙 | 大学生のアルバイト<br>レストランでの禁煙           |
| タスク    | 意見主張                   | イラスト関連質疑<br>ロールプレイ<br>ロールプレイ関連質疑 |
| 総語数    | 英語母語話者<br>94,168       | 英語母語話者<br>30,753                 |
|        | 日本語母語英語学習者<br>41,737   | 日本語母語英語学習者<br>103,073            |
| 総ファイル数 | 英語母語話者<br>600          | 英語母語話者<br>120                    |
|        | 日本語母語英語学習者<br>600      | 日本語母語英語学習者<br>600                |

## 4.2 語連鎖の抽出基準

高頻度語連鎖を対象とした多くの研究においては、任意の語連鎖が定型表現であるかどうか、すなわち、話者が高頻度で繰り返し用いる表現であるかどうかについて判定する方法として、頻度 (frequency) と分布度 (dispersion) の2つの指標が用いられている (Ädel & Erman, 2012; Biber et al., 2004; Biber & Barbieri, 2007; Chen & Baker, 2016; Hyland, 2008; Yan, 2019; 石川, 2019)。頻度については、研究者によって様々な閾値が用いられている (Biber et al., 2004; Hyland, 2008)。英語学習者の話し言葉における語連鎖を調査した Yan (2019) では、同様に英語母語話者の話し言葉について検証した Biber et al. を参考に、100万語中40回以上の頻度で現れる語連鎖を抽出対象としている。本調査では、話し言葉を分析対象とすることを鑑み、話し言葉における語連鎖を調査した Yan と Biber et al. を参考に、100万語中に40回以上出現する語連鎖を対象にした。なお、Yan と同様、本研究で使用した4つのコーパスはどれも100万語を満たしていないため、100万語に換算したときに40回以上となるものを抽出したことに注意されたい。

上述の通り、頻度の抽出基準については研究によって様々であり、どの程度の頻度であれば高頻度としてみなされるかについては明らかではないため、頻度のみを用いて定型表現を判定するのは注意を要する。そこで重要となる2つ目の指標が、分布度である。Biber et al. (2004) では、高頻度語連鎖は、コーパス内のテキストに幅広く分布していると主張し、分布度を用いることで個々人の偏った語連鎖使用で見られる表現の抽出を防ぐことができるとして、高頻度語連鎖を異なる5つ以上のファイルに用いられるものと

定めている。これを踏まえ、Yan(2019)では、分析対象とした3つのコーパスの総ファイル数がそれぞれ276, 314, 217と異なっていることを考慮し、また、Biber et al. が用いた5と近接するように、各コーパスの総ファイル数のうち2%以上のファイルに出現する語連鎖を抽出している。上記2つの先行研究を参考に、本調査においても対象とするコーパスサイズが異なること、分布度が先行研究に近似した値(5)になることに留意し、各コーパスの総ファイル数のうち1%以上のテキストに出現する語連鎖を対象とすることとした。

最後に、どの長さの語連鎖を抽出するかについてであるが、これまでに最も多く研究されてきた長さは4語であるとされている(Chen & Baker, 2010)。その理由として、4語連鎖は、手動で語連鎖を分類したり、コンコーダンスを確認したりする際に最も扱いやすく(Chen & Baker, 2010)、5語連鎖よりも頻度が高く現れ、3語連鎖よりもはっきりとした様々な構造や機能を示すとされるためである(Hyland, 2008)。また、話し言葉を調査したBiber et al.(2004)やYan(2019)でも4語連鎖を扱っており、4語を対象とすることで、多くの先行研究との整合性を図ることができる。これらを踏まえ、本調査においても4語連鎖について分析することとした。

表2は、これまでに述べたコーパスごとの抽出基準を示したものである。ただし、ICNALE Spoken Monologueの英語母語話者と日本語母語英語学習者、ICNALE Spoken Dialogueの日本語母語英語学習者については、分布度との関連で、頻度の抽出基準は実質的には6以上となっていることに留意されたい。

■表2: 各コーパスにおける語連鎖の抽出基準

| コーパス                                    | 頻度 (frequency) |      | 分布度 (dispersion) |      |
|-----------------------------------------|----------------|------|------------------|------|
|                                         | 総語数            | 抽出基準 | 総ファイル数           | 抽出基準 |
| ICNALE Spoken Monologue<br>(英語母語話者)     | 94,168         | 4    | 600              | 6    |
| ICNALE Spoken Monologue<br>(日本語母語英語学習者) | 41,737         | 2    | 600              | 6    |
| ICNALE Spoken Dialogue<br>(英語母語話者)      | 30,753         | 2    | 120              | 2    |
| ICNALE Spoken Dialogue<br>(日本語母語英語学習者)  | 103,073        | 5    | 600              | 6    |

### 4.3 語連鎖の抽出方法

語連鎖の抽出は、フリーコーパスソフトウェアであるAntConc(Anthony, 2023)のN-gram機能を用いて実施した。N-gramは、コーパス内に共通して現れる任意の長さNの語連鎖を抽出する分析機能であり、Nの語連鎖とは、it isのような2語連鎖やas well asのような3語連鎖を示す。なお、本調査では書き言葉における日本語母語英語学習者の語連鎖使用について調査した石川(2019)に倣い、抽出した各項目に動詞や形容詞の活用形が含まれる場合も、原形に戻さず、テキストの中で用いられているままの表記で集計を行った(例: is going toとam going to は別個に処理)。また、AntConcでの処理の際、抽出された語連鎖は、全て小文字扱いとなり(例: I want toは、i want toとして処理)、ハイフン、アポストロフィ、コンマ、ピリオド、クエスチョンマークといった記号は1語と見なさず、スペース扱いとなり省略され処理されている(例: part-timeはpartとtimeの2語、can'tは、canとtの2語として処理)。

#### 4.4 トピックに関連した語連鎖の扱い

石川(2019)によると、学習者は言語産出の際に与えられた指示文に含まれる表現をそのまま流用する傾向が見られ、学習者が過剰使用する語連鎖の大半はトピックに関連した表現であるとされる。この点を踏まえ、本調査では、トピック文に含まれる表現や直接関連した表現以外の高頻度語連鎖に焦点を当てるため、語連鎖の抽出後、トピックに関連した語連鎖(topic bundles)を分析対象から除外した。具体的には、a part time jobやsmoking is bad forといったICNALEの共通トピックに関連する語連鎖、SDの絵描写やロールプレイの内容に関する表現であるswim in the sea やsmoking in the parkなどは分析対象から除外した(絵描写の発話自体は分析に含めていないが、このような表現は主に関連質疑において見られた)。このような表現からは、定型表現に関する教育的示唆を得られる可能性が低く、話し言葉を調査したYan(2019)においても分析対象外とされていることから、本調査でも対象外とした。

#### 4.5 語連鎖の分析方法

本調査では、学習者が使用する語連鎖の使用実態を調査するため、日本語母語英語学習者の書き言葉を対象とした石川(2019)を参考に分析を行うこととした。分析の手順は以下の通りである。

初めに、各語連鎖の1万語あたりの調整頻度と分布度比率を算出した。ここで、調整頻度を1万語あたりとした理由は、元のコーパスサイズを超えて基準値を設定することは、データが持つ情報量を過大解釈することにつながるためである(石川, 2021)。したがって、基準値の決定については、調整しようとする全てのコーパスサイズの下限値を超えない範囲で最も大きな10の乗数を選択するのが原則であるとされる(石川, 2021)。そのため、本調査で分析対象とした4つのコーパスサイズを考慮し、1万語あたりの調整頻度を用いることとした。

次に、分布度比率の算出を行った。分布度比率とは、各コーパスに所収の総ファイル数のうち、当該語連鎖が1回以上出現しているファイル数の比率である(例: 総ファイル数が300であり、そのうち任意の語連鎖が1回以上出現しているファイル数が210の場合、分布度比率は70%となる)。

最後に、1万語あたりの調整頻度に分布度比率を掛けた値である重要度を求めた。重要度は、頻度と分布度を同時に評価する指標であり、例えば、調整頻度が200で分布度比率が50%の語連鎖と、調整頻度が150で分布度比率が80%の語連鎖の場合、前者の重要度指標は100、後者の重要度指標は120となる(石川, 2019)。

その後、各語連鎖における母語話者の重要度から学習者の重要度を引いた差分値を算出した。重要度の値が正で高いほど学習者が過少使用している表現であり、負で低いほど学習者が過剰に使用している語連鎖であると解釈される。ただし、学習者データもしくは母語話者データ的一方のみから抽出された項目については、抽出されていない方の当該語連鎖の重要度を0として差分値を求めた。

差分値を算出した後、2つのRQについて順に調査を行った。まず、RQ1について、SMにおいて学習者が過剰・過少使用している差分値上位100位までの項目を取り上げ、過剰・過少使用のそれぞれについて差分値の絶対値順にリスト化した。次に、作成したリストを元に、学習者が使用する高頻度語連鎖の言語的特徴について、各語連鎖を構成する語の品詞やその表現の文法について検討した。また、抽出した語連鎖の前後の文脈についても観察し、その語用論的機能について検討した。RQ2については、SDを対象に同様の手順で分析を行った。

## 5 結果と考察

表3は、SMとSDそれぞれで抽出された語連鎖数である。これらの語連鎖のうち、本調査では4.4で述べ

た通り, non-topic bundlesを対象に分析を行った。また, 表4と表5は抽出された語連鎖について, 語の品詞や文法事項に着目し, いくつかのカテゴリーに分類したものである。表の左側には各カテゴリーの名称, 中央には本研究で抽出された語連鎖のうち, そのカテゴリーに分類された特徴的な語連鎖, 右側には各語連鎖の差分値が記載されている。これらのカテゴリーごとに, 本節では, 学習者の語連鎖使用の実態をその考察と併せて報告する。

■表3: 各コーパスから抽出した語連鎖数

| カテゴリー                                   | 抽出した語連鎖     |               |                   |
|-----------------------------------------|-------------|---------------|-------------------|
| コーパス                                    | all bundles | topic bundles | non-topic bundles |
| ICNALE Spoken Monologue<br>(英語母語話者)     | 455         | 203           | 252               |
| ICNALE Spoken Monologue<br>(日本語母語英語学習者) | 263         | 126           | 137               |
| ICNALE Spoken Dialogue<br>(英語母語話者)      | 730         | 104           | 626               |
| ICNALE Spoken Dialogue<br>(日本語母語英語学習者)  | 541         | 115           | 426               |

■表4: SMにおいて抽出した語連鎖のカテゴリー化

| カテゴリー      | 項目                 | 差分値     |
|------------|--------------------|---------|
| 句動詞を含む語連鎖  | i agree with this  | -4.1131 |
|            | it depends on the  | 0.0234  |
| 意見主張を示す語連鎖 | i think it is      | -0.6074 |
|            | believe that it is | 0.0299  |
| 助動詞を含む語連鎖  | i think we should  | -0.1138 |
|            | should be able to  | 0.0673  |
| その他の語連鎖    | i have two reasons | -0.3358 |
|            | at the same time   | 0.1168  |

■表5: SDにおいて抽出した語連鎖のカテゴリー化

| カテゴリー                       | 項目              | 差分値     |
|-----------------------------|-----------------|---------|
| 願望を表す語連鎖                    | so i want to    | -0.3936 |
|                             | i would like to | 0.2212  |
| 発話の繰り返しやフィラーを<br>構成語彙とする語連鎖 | i can i can     | -0.2411 |
|                             | uh let me see   | 0.0244  |
| その他の語連鎖                     | i am going to   | 0.2385  |
|                             | as long as they | 0.0542  |

## 5.1 モノローグ (Monologue) における学習者の語連鎖使用の実態 (RQ1)

### 5.1.1 句動詞を含む語連鎖

過剰使用語連鎖の差分値順上位100位のうち、agreeを含む語連鎖が23項目、disagreeを含む語連鎖は5項目観察された。このうち、句動詞であるagree withやdisagree withが含まれていた項目は計19項目であった。特に、i agree with thisは、過剰使用語連鎖の中で最も差分値の高い項目であり、母語話者に比べ学習者がとりわけ多用する表現であるといえる。また、i disagree with thisにおいても差分値順位が5位であったことから、母語話者と比較し学習者が好んで使用する表現であると考えられる。書き言葉を調査した石川(2019)においても、i agree with thisとi disagree with thisは共に学習者が過剰使用する表現として抽出されており、これらは話し言葉や書き言葉にかかわらず、母語話者と比べ学習者が多用する語連鎖であるといえるだろう。さらに、agreeを構成語彙とする語連鎖の中には、i agree this statementやi agree the opinionなどのagree＋目的語の構成を持つ語連鎖が9項目観察された。これらはwithが不足した文法的に誤った表現であり、学習者コーパスからのみ抽出され、母語話者は使用しない表現であった。

一方、学習者が過少使用する語連鎖に注目すると、have to worry aboutやit depends on the, to take care ofなどの句動詞を含む語連鎖が抽出され、これらの学習者の使用頻度は0であり、母語話者のみが使用した表現であることがわかった。話し言葉における句動詞使用を調査した石井(2018)では、句動詞使用について、学習者の習熟度が上がるにつれて使用頻度は上昇していくものの、上級の学習者であっても使用する句動詞は母語話者とは大きく異なり、母語話者の使用頻度が高いにもかかわらず、学習者がうまく使えないものがあることを報告している。本調査においてもその傾向が現れ、上記のような句動詞を含む語連鎖の使用傾向に違いが見られたと考えられる。

### 5.1.2 意見主張を示す語連鎖

これまでの研究では、書き言葉において、日本語母語英語学習者のi thinkのようなthinkを構成語彙とする語連鎖の過剰使用が報告されていた(石川, 2019; 阪上, 2013)。話し言葉を対象とした本調査においても、過剰使用語連鎖上位100位を見ると、10項目がthinkを含む語連鎖であった(e.g., i think it is)。一方、過少使用語連鎖では24項目が抽出されており(e.g., think it s important)、話し言葉(SM)においては、学習者だけではなく、母語話者についてもthinkを構成語彙とする語連鎖を多用する傾向があるといえる。

さらに、学習者の過少使用では、believe that it isやi believe it sのようなbelieveを含む表現が抽出されており、母語話者は自分の考えや意見を述べる際に、thinkに加え、believeを用いていると考えられる。加えて、学習者は、i do think thatやi do believe thatといった強調の助動詞doを構成語彙とする語連鎖を過少使用していた。

### 5.1.3 助動詞を含む語連鎖

次に、助動詞を含む語連鎖に注目すると、過剰使用語連鎖では、i think we shouldのみが抽出された一方で、過少使用では、計13項目が見られた。これらの表現の構造を確認すると、i think they shouldやi think it wouldなどのthink＋主語＋助動詞の構造を持つ語連鎖に加え、should not be allowedのように受動動詞を持つ表現やshould be able toといった3語連鎖であるbe able toを含む項目が見られ、学習者に比べ母語話者は、助動詞を含む多様な構造の語連鎖を使用していることがわかった。助動詞の使用について、これまでの研究で学習者と母語話者では助動詞の使用傾向が異なることが報告されている(石田, 2011; 鈴木, 2022)。特に、wouldについては、そもそも学習者は母語話者に比べ、wouldをあまり使用しないことが知られている(Ishikawa, 2019; 鈴木, 2022)。同様に、本調査でも学習者においてwouldを含む語連鎖の過少使用が見られたことになる。

#### 5.1.4 その他の語連鎖

ここでは、差分値上位に占める項目数は少ないが、特徴的な構造や機能を持つ語連鎖を記述する。学習者の過少使用語連鎖では、it s going toやthey re going toといったbe going toを含む表現に加え、be able to haveやnot be able toなどのbe able toを構成語彙とする表現、when i was a やas i was saying, while they are in, if you don tといった多様な副詞節を導く語連鎖が抽出された。また、母語話者と比較して学習者はat the same timeやon the other handを過少使用していた。中国語母語英語学習者の書き言葉を調査したChen and Baker (2016)や話し言葉を扱ったYan (2019)では、at the same timeやon the other handは学習者コーパスから抽出されたが、本調査においては確認されなかった。このことから、これらは日本語母語英語学習者が特徴的にあまり使用しない語連鎖であると推察される。上記に挙げた表現については、学習者にとって習得が困難であり、定着に時間のかかる項目である可能性がある。

次に、過剰使用語連鎖を見ると、i have two reasons, first reason is thatなどのreasonを含む表現が4項目観察された。これらは、母語話者は一切使用せず、学習者のみが使用した語連鎖であった。これらはスピーチなどを行う際に典型的に指導される表現であると推察されるが、そのような表現が学習者にある程度定着しているといえる一方、母語話者がほとんど用いることのない表現を繰り返し指導していることにもなり得る。

### 5.2 ダイアログ (Dialogue) における学習者の語連鎖使用の実態 (RQ2)

#### 5.2.1 願望を表す語連鎖

過剰使用語連鎖の差分値上位100位のうち、wantを含む語連鎖は28項目であり、およそ3割を占めていた。これらの語連鎖の構造を確認すると、so i want toやi want to goなどの不定詞節を導く表現や、i want you toのようにwant＋目的語＋toの構造を持つ項目が散見された。一方、過少使用語連鎖を見ると、want toと類似した意味を持つwould like toを含む項目が抽出され、願望を表す語連鎖として学習者がwant toを多用する一方で、母語話者はwould like toを用いて表現していると考えられる。

#### 5.2.2 発話の繰り返しやフィラーを構成語彙とする語連鎖

過剰使用語連鎖の中には、when i when iやwe can we canのように同じ発話の繰り返しで構成されている語連鎖が9項目見られており、SMでは5項目抽出されていた。過少使用語連鎖においてもSDでこのような項目は5項目観察されたが、SMでは1項目のみであった。この結果から、学習者と母語話者共に繰り返しの発話を含む語連鎖をある程度使用すると考えられるが、比較的学习者の方がその傾向が強く、SMよりもSDにおいてその使用頻度は上昇するといえる。

また、SMの差分値上位の語連鎖では確認されなかったが、SDの過剰使用語連鎖では、umm i want toやuh i don tといったフィラーを含む表現が22項目存在していた。上記のような語連鎖が見られた理由については、学習対象の言語産出に伴う認知的負荷が高くなることで同じ語連鎖を繰り返し使用したり、言語産出に時間を確保するために、フィラーを含む表現を多用したりするためであると考えられる。特に、自分のペースで単方向的に話すSMと違って、SDは他者との対話であり、相手を意識した状況となるため、このような表現の使用頻度が増加する傾向があると考えられる。

加えて、SDでは、yes yes yes yesのような構成語彙が全て同じ項目が2つ観察された。これらの表現はSMでは抽出されなかったが、SDは対話データでありインタビュー어가質問し、参加者が回答する場面が多く存在しているため、上記のような語連鎖が見られたといえるだろう。ここで過少使用語連鎖を見ると、yes i do iやno i don tが抽出されていることから、回答する際に用いる表現についても学習者と母語話者では違いがあると考えられる。さらに、学習者はuh let me seeやwell i think itなどの語連鎖をあまり使用しておらず、会話をつなぐ表現の1つであるlet me seeやwellを含む表現を過少使用していた。

### 5.2.3 その他の語連鎖

過少使用語連鎖では、i was working inやi am trying toのような過去進行形や現在進行形、現在完了形を示すi ve been toや過去の習慣を表現するi used to go、予定を表すi am not goingなど多様な時制表現を含む語連鎖が見られた。また、学習者はas long as theyやas you can seeといったasを構成語彙とする表現に加え、SMでも過少使用として抽出されたat the same timeをあまり使用していなかった。上記に挙げた語連鎖については、学習者が時制表現の用法や意味をあまり理解できていなかったことや、理解できていたとしてもアウトプットする練習があまり与えられていなかったことから、実際に使用できなかった可能性が考えられる。

## 6 結論と今後の課題

本調査では2つの研究課題を通して、話し言葉における日本語母語英語学習者の語連鎖使用の実態を調査した。RQ1では、モノログを対象に分析を行い、母語話者と比較して学習者が過剰使用する語連鎖やあまり使用していない表現などについて検証した。その結果、学習者は特定の句動詞を含む語連鎖 (e.g., agree with) や理由を述べる際に用いる語連鎖 (e.g., i have two reasons) を多用する傾向が見られた。一方で、学習者の過少使用を見ると、他の多様な句動詞表現 (e.g., worry about, depends on) や助動詞を構成語彙とする表現に加え (e.g., i think we should, should not be allowed), at the same timeやon the other handなどの定型表現が抽出され、学習者は母語話者と比較してこれらの表現をあまり使用していないことが明らかとなった。また、自身の意見や考えを示す定型表現に着目すると、学習者と母語話者共にthinkを含む語連鎖を使用しており、さらに学習者はbelieve that it isのようにbelieveを構成語彙とする語連鎖を過少使用していることが明らかとなった。

次に、RQ2では、同様の手順でダイアログについて分析を行った。対話データにおける学習者言語の特徴として、学習者は同じ発話の繰り返しで構成される語連鎖やフィラーを含む項目 (e.g., when i when i, uh i don t), wantを構成語彙とする語連鎖 (e.g., so i want to, i want to go) を過剰使用していた一方で、would like toや多様な時制表現を示す表現 (e.g., i was working in, i ve been to), asを含む語連鎖 (e.g., as long as they) などを過少使用することが明らかとなった。フィラーや繰り返しを含む語連鎖を多く使用するという傾向は、モノログとは異なるダイアログの特性を反映しているといえる。

本調査で示した語連鎖のデータは、同一のテーマ・条件で収集された母語話者の発話と比較して、学習者が過剰使用・過少使用する表現を抽出したものであり、妥当性が高く、一定の資料的価値があると考えられる。本調査で得られた結果を基に、発話における学習者言語の特徴を把握することで、学習者がよく使う表現と類似した別の言い方や、会話場面で方略的に使われる表現などを指導したり、特定の動詞とよく使われる前置詞や副詞を組み合わせる指導したりするなど、指導上の工夫を考えることができるだろう。本調査の結果が、定型表現を効果的に活用した「話すこと[やりとり]」及び「話すこと[発表]」の技能育成のための一助となることを期待する。

一方で、本調査にはいくつかの限界点が存在することを述べておく必要がある。まず、本調査で利用したデータについて、SMは2014年頃、SDは2018年頃の時点で収集されたものであり、現在の日本国内における英語学習者の現状を反映したものではない点に留意が必要である。同様に、本調査で分析対象としたSMとSDは、現行の学習指導要領で示されている「話すこと[やりとり]」と「話すこと[発表]」の特性を厳密に反映したものではない。例えば、やりとりでは、即興で互いの考えや気持ちなどを伝え合ったりすることに加え、相手の発話に応じたり、それに関連した質問や意見を述べたりして、互いに協力して対話を継続・発展させることが必要とされているが(学習指導要領, 2017b)、SDではインタビューする人とされる人という役割が与えられ、主に事前に決められた順序で会話が進むという構成になっていた。また、発表では、スポーツや音楽など身の回りのことで生徒が共通して関心を持っていることについての話す

力の育成が示されているが、SMでは指定された2つのトピックのみについての発話であった。今後は、現在あるいは近年の学校英語教育を経験した学習者からデータを収集したり、「話すこと[やりとり]」と「話すこと[発表]」の特性をより厳密に反映したデータを用いて検証を重ねていくことが求められるだろう。

また、本研究で分析対象としたコーパスは、大学(院)生が産出した言語データであったため、中学生や高校生といった発達段階における学習者の語連鎖の使用については調査できていない。今後、他の発達段階における学習者の言語データも分析することで、各発達段階の語連鎖使用の実態についても明らかにすることができるだろう。

さらに、本研究では英語母語話者の定型表現の使用傾向を基準に、学習者の過剰・過少使用する表現を求めているが、母語話者の言語使用を厳密な基準とすることには議論の余地があるだろう。近年、批判的応用言語学の観点から、英語教育を取り巻く規範とその問題点の検討がなされている(久保田, 2018)。この批判的応用言語学の視点から見ると、母語話者を基準にした教育を行うことは現実と乖離しているとも考えられる。また、このような教育は、ネイティブスピーカリズム(native-speakerism)を助長する可能性もある。これらの点を考慮した上で、本研究の結果をどのように言語指導や学習に取り入れていくかを慎重に検討していく必要があるだろう。

最後に、本研究で分析した学習者と母語話者コーパスは総語数やサンプル数に大きな違いがあり、加えて、学習者コーパスでは習熟度に偏りがあることに留意が必要である。これらの要素は本研究で得られた結果に影響を及ぼしている可能性があり、その点を踏まえた上で本研究の結果は解釈される必要がある。今後はデータの均衡を図った上で、学習者の習熟度別に語連鎖使用の特徴などを検討することができれば、定型表現の学習・指導に対してさらに有益な示唆が得られるであろう。

## 謝辞

本研究を実行する機会を与えてくださった公益財団法人 日本英語検定協会、関係者の皆様、ならびに、選考委員の先生方に、心より御礼申し上げます。とりわけ、私の調査を担当してくださり、激励のお言葉や丁寧なご指導をいただいた研究助言者の竹内理先生に厚く感謝申し上げます。そして、筑波大学大学院の名畑目真吾先生には、研究の着想から、調査、論文執筆まで親身なご指導をいただきました。心より感謝申し上げます。最後に、多くのご助言と励ましをいただきました研究室の皆様に深く御礼申し上げます。今後もこの貴重な経験を生かして、教育現場に少しでも還元できるよう、研究に精進いたします。なお、本報告書に関して、開示すべき利益相反関連事項はありません。

## 引用文献

- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.101-122). Oxford University Press.
- Anthony, L. (2023). AntConc (Version4.2.0) [Computer software]. Tokyo: Waseda University. Retrieved from <https://www.laurenceanthony.net/software/antconc/>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of Spoken and Written English*. Longman.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26(3), 263-86.
- Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, 14, 30-49.
- Chen, Y., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849-880.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? . *Applied linguistics*, 29(1), 72-89.
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61.

## 引用文献

- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning. Teaching and testing* (pp. 75-93). Longman.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.79-100). Oxford University Press.
- Güngör, F. & Uysal, H. H. (2016). A comparative analysis of lexical bundles used by native and non-native scholars. *English Language Teaching*, 9(6), 176-188.
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp.155-175). Oxford University Press.
- Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21.
- 石井康毅 (2018).「話し言葉コーパスと検定教科書に基づく日本人英語学習者の句動詞使用実態の分析」*Learner Corpus Studies in Asia and the World*, 3, 101-119.
- 石川慎一郎 (2018).「The ICNALE Spoken Dialogue の設計:対話における L2口頭産出研究のために」*Learner corpus studies in Asia and the world*, 3, 9-26.
- 石川慎一郎 (2019).「英語教育における連語:ターゲット・インプット・アウトプットの三元コーパス分析をふまえた English N-gram List for Japanese Learners of English (ENL-J) の開発と利用」仁科恭徳・吉村由佳・吉川祐介 (編).『言語分析のフロンティア』(pp.32-47) 金星堂.
- 石川慎一郎 (2020).「第4章 コーパスと英語教育研究」石川慎一郎・長谷部陽一郎・住吉誠『コーパス研究の展望』(pp. 193-244). 開拓社.
- 石川慎一郎 (2021).『ベーシックコーパス言語学 第2版』ひつじ書房.
- Ishikawa, S. (2019). The ICNALE Spoken Dialogue: A New Dataset for the Study of Asian Learners' Performance in L2 English Interviews. *English Teaching*, 74(4), 153-177.
- Ishikawa, S. (2023). *The ICNALE Guide: An Introduction to a Learner Corpus Study on Asian Learners' L2 English*. Routledge.
- 石田知美 (2011).「日本人英語学習者コーパスを用いた法助動詞の使用に関する研究」『JACET 中部支部紀要』第9号, 57-73.
- Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: How do Japanese EFL learners process multi-word expressions. *JACET Kansai Journal*, 13, 38-49
- Jalilifar, A., & Ghoreishi, S. M., & Roodband, S (2016). Developing an inventory of core lexical bundles in English research articles: a cross-disciplinary corpus-based study. *Journal of World Languages*, 3(3), 184-203.
- Jiang, N. A., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433-445.
- 金澤佑 (編) (2020).『フォーミュラと外国語学習・教育:定型表現研究入門』くろしお出版.
- 小出凱渡 (2024).「英作文における学習者の定型表現使用の実態 —母語話者・学習者コーパスに基づく比較分析—」筑波大学大学院教育学学位プログラム修士論文 (未公開)
- 久保田竜子 (2018).『英語教育幻想』筑摩書房.
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.
- Mikajiri, N. (2022). Quantitative Analysis of Lexical Bundles Used in Spoken English and Japanese Elementary and Junior High School English Textbooks. *Journal of Corpus-based Lexicology Studies*, 5, 1-23.
- 文部科学省 (2017a).『小学校学習指導要領 解説 外国語活動・外国語編』Retrieved from [https://www.mext.go.jp/content/20201029\\_mxt\\_kyoiku01-100002607\\_11.pdf](https://www.mext.go.jp/content/20201029_mxt_kyoiku01-100002607_11.pdf)
- 文部科学省 (2017b).『中学校学習指導要領解説 外国語編 英語編』Retrieved from [https://www.mext.go.jp/content/20210531-mxt\\_kyoiku01-100002608\\_010.pdf](https://www.mext.go.jp/content/20210531-mxt_kyoiku01-100002608_010.pdf)
- 文部科学省 (2018).『高等学校学習指導要領解説 外国語編 英語編』Retrieved from [https://www.mext.go.jp/content/1407073\\_09\\_1\\_2.pdf](https://www.mext.go.jp/content/1407073_09_1_2.pdf)
- 中田達也 (2022).『英語は決まり文句が8割:今日から役立つ「定型表現」学習法』講談社.
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford University Press.
- Northbrook, J., & Conklin, K. (2018). "What are you talking about?" An analysis of lexical bundles in Japanese junior high school textbooks. *International Journal of Corpus Linguistics*, 23(3), 311-334.
- Northbrook, J., & Conklin, K. (2019). Is what you put in what get out? Textbook — derived lexical bundle processing in beginner English learners. *Applied Linguistics*, 40(5), 816-833.
- Pan, F., Reppen, R., & Biber, D. (2016). Comparing patterns of L1 versus L2 English academic professionals: Lexical bundles in telecommunications research journals. *Journal of English for Academic Purposes*, 21, 60-71.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp.191-226). Longman.
- 阪上辰也 (2013).「日本語母語英語学習者のエッセイに見られる共起表現の分析」『広島外国語教育研究』第16号, 159-169.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt. (Ed.), *Formulaic Sequences:*

## 引用文献

- Acquisition, processing and use* (pp. 1–22). John Benjamins.
- Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). *Understanding formulaic language: A second language acquisition perspective*. New York: Routledge.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for Academic Purposes*, 12(3), 214–225.
- 鈴木陽子 (2022). 「モノローグ発話における日本人英語学習者の 法助動詞の使用」『明治学院大学教養教育センター紀要：カルチュラル』第16巻, 1-10.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford University Press.
- Yan, H. (2019). I think we should…: Investigating lexical bundle use in the speech of English learners across proficiency levels. *International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL)*, 1(2), 1–16.

資料1：SMにおける学習者の過剰使用語連鎖リスト(差分値の絶対値が大きい順).....

| 順位 | 項目                           | 差分値     | 順位  | 項目                         | 差分値     |
|----|------------------------------|---------|-----|----------------------------|---------|
| 1  | i agree with this            | -4.1131 | 51  | it is bad for              | -0.0467 |
| 2  | agree with this statement    | -1.7835 | 52  | i agree this idea          | -0.0439 |
| 3  | i agree with the             | -1.3588 | 52  | i i don t                  | -0.0439 |
| 4  | with this statement because  | -0.8821 | 52  | so they have to            | -0.0439 |
| 5  | i disagree with this         | -0.7835 | 55  | don t don t                | -0.0431 |
| 6  | i think it is                | -0.6074 | 56  | so i agree this            | -0.0399 |
| 7  | agree with the statement     | -0.5852 | 56  | statement because i think  | -0.0399 |
| 8  | it is important for          | -0.3964 | 56  | t agree with the           | -0.0399 |
| 9  | i don t agree                | -0.3953 | 56  | this statement this is     | -0.0399 |
| 10 | is not good for              | -0.3666 | 60  | can learn a lot            | -0.0395 |
| 11 | i have two reasons           | -0.3358 | 60  | learn a lot of             | -0.0395 |
| 12 | disagree with this statement | -0.2907 | 60  | we can we can              | -0.0395 |
| 13 | i don t like                 | -0.2529 | 63  | have a right to            | -0.0383 |
| 14 | agree with this opinion      | -0.2492 | 64  | agree this opinion because | -0.0323 |
| 14 | i agree this statement       | -0.2492 | 64  | agree with the opinion     | -0.0323 |
| 16 | don t agree with             | -0.2108 | 64  | because i don t            | -0.0323 |
| 17 | with the statement because   | -0.1933 | 64  | have two reasons first     | -0.0323 |
| 18 | it is important to           | -0.1757 | 64  | i think i think            | -0.0323 |
| 19 | with this statement i        | -0.1442 | 64  | some people don t          | -0.0323 |
| 20 | think it is important        | -0.1340 | 64  | with the idea that         | -0.0323 |
| 21 | this statement because i     | -0.1294 | 71  | so that s why              | -0.0319 |
| 22 | a lot of money               | -0.1222 | 72  | can learn how to           | -0.0288 |
| 23 | i don t i                    | -0.1214 | 73  | agree this statement i     | -0.0256 |
| 24 | with this opinion because    | -0.1154 | 73  | i agree with that          | -0.0256 |
| 25 | people who don t             | -0.1143 | 73  | i i agree with             | -0.0256 |
| 26 | i think we should            | -0.1138 | 73  | need a lot of              | -0.0256 |
| 27 | agree with the idea          | -0.1086 | 73  | statement this is because  | -0.0256 |
| 27 | i agree this opinion         | -0.1086 | 73  | this statement i think     | -0.0256 |
| 29 | this statement i have        | -0.1022 | 73  | with this statement this   | -0.0256 |
| 30 | it is not good               | -0.0839 | 80  | it is good for             | -0.0252 |
| 30 | that s why i                 | -0.0839 | 81  | i can t i                  | -0.0224 |
| 32 | i don t want                 | -0.0827 | 82  | have a lot of              | -0.0210 |
| 33 | agree with this idea         | -0.0783 | 83  | and second reason is       | -0.0196 |
| 33 | statement i have two         | -0.0783 | 83  | because if we have         | -0.0196 |
| 35 | t i don t                    | -0.0779 | 83  | disagree with the idea     | -0.0196 |
| 36 | i disagree with the          | -0.0752 | 83  | first reason is that       | -0.0196 |
| 37 | agree this statement because | -0.0675 | 83  | how to how to              | -0.0196 |
| 37 | with this idea because       | -0.0675 | 83  | i agree the opinion        | -0.0196 |
| 39 | it is it is                  | -0.0623 | 83  | i agree with it            | -0.0196 |
| 40 | so i don t                   | -0.0576 | 83  | i think we can             | -0.0196 |
| 41 | so i think we                | -0.0575 | 83  | if we have a               | -0.0196 |
| 42 | disagree with this opinion   | -0.0571 | 83  | opinion it is important    | -0.0196 |
| 42 | don t i don                  | -0.0571 | 83  | people don t like          | -0.0196 |
| 44 | is very bad for              | -0.0547 | 83  | t agree with this          | -0.0196 |
| 45 | so we have to                | -0.0527 | 83  | think it is not            | -0.0196 |
| 45 | there are many people        | -0.0527 | 83  | this opinion because i     | -0.0196 |
| 47 | so i think it                | -0.0514 | 83  | we don t have              | -0.0196 |
| 48 | a lot of things              | -0.0486 | 98  | we can know how            | -0.0192 |
| 49 | is bad for our               | -0.0483 | 99  | that it is important       | -0.0184 |
| 50 | it is very important         | -0.0480 | 100 | is very important for      | -0.0174 |

注．分析では少数点第9位までの差分値を求めているが，本資料では少数点第4位までの値を掲載。

資料2: SMにおける学習者の過少使用語連鎖リスト(差分値の絶対値が大きい順).....

| 順位 | 項目                      | 差分値    | 順位 | 項目                          | 差分値    |
|----|-------------------------|--------|----|-----------------------------|--------|
| 1  | i think it s            | 2.5798 | 53 | i believe it s              | 0.0255 |
| 2  | i don t think           | 1.3586 | 53 | i don t believe             | 0.0255 |
| 3  | to be able to           | 0.5752 | 53 | think it would be           | 0.0255 |
| 4  | don t think it          | 0.2825 | 56 | they want to do             | 0.0254 |
| 5  | think it s important    | 0.2676 | 57 | believe that it s           | 0.0253 |
| 6  | i don t know            | 0.1848 | 58 | it depends on the           | 0.0234 |
| 7  | t think it s            | 0.1434 | 58 | um i think it               | 0.0234 |
| 8  | i think that it         | 0.1283 | 60 | i think they should         | 0.0214 |
| 9  | and i think that        | 0.1189 | 60 | it s important that         | 0.0214 |
| 10 | at the same time        | 0.1168 | 60 | should not be allowed       | 0.0214 |
| 11 | think it s a            | 0.1150 | 60 | the other hand i            | 0.0214 |
| 12 | on the other hand       | 0.1104 | 64 | will be able to             | 0.0212 |
| 13 | while they are in       | 0.1058 | 65 | have to worry about         | 0.0195 |
| 14 | don t have to           | 0.1011 | 65 | that it should be           | 0.0195 |
| 15 | with the statement that | 0.0941 | 65 | they are going to           | 0.0195 |
| 16 | don t think that        | 0.0896 | 68 | and things like that        | 0.0191 |
| 17 | i think that s          | 0.0885 | 68 | to learn how to             | 0.0191 |
| 18 | it s it s               | 0.0818 | 68 | to take care of             | 0.0191 |
| 19 | i believe that it       | 0.0807 | 71 | and it s very               | 0.0177 |
| 20 | they don t have         | 0.0804 | 71 | disagree with the statement | 0.0177 |
| 21 | and i think it          | 0.0781 | 71 | in the real world           | 0.0177 |
| 22 | the people around them  | 0.0708 | 71 | so i think that             | 0.0177 |
| 23 | should be able to       | 0.0673 | 71 | the people who are          | 0.0177 |
| 24 | a lot of people         | 0.0631 | 71 | they should be able         | 0.0177 |
| 25 | it s very important     | 0.0622 | 71 | when i was a                | 0.0177 |
| 26 | if you don t            | 0.0621 | 78 | i mean it s                 | 0.0175 |
| 27 | it s good for           | 0.0605 | 78 | it helps them to            | 0.0175 |
| 28 | think it s very         | 0.0572 | 80 | don t want to               | 0.0164 |
| 29 | you don t have          | 0.0566 | 81 | as i was saying             | 0.0159 |
| 30 | i do believe that       | 0.0481 | 81 | s a good idea               | 0.0159 |
| 31 | i also think that       | 0.0451 | 81 | that it would be            | 0.0159 |
| 32 | it s bad for            | 0.0446 | 81 | think it s really           | 0.0159 |
| 33 | that it s important     | 0.0398 | 85 | be able to go               | 0.0143 |
| 34 | in the united states    | 0.0396 | 85 | don t believe that          | 0.0143 |
| 35 | you know it s           | 0.0391 | 85 | don t have a                | 0.0143 |
| 36 | be able to have         | 0.0372 | 85 | i agree that it             | 0.0143 |
| 36 | i do think that         | 0.0372 | 85 | i think it would            | 0.0143 |
| 36 | i don t really          | 0.0372 | 85 | it s really important       | 0.0143 |
| 36 | it s a good             | 0.0372 | 85 | no i don t                  | 0.0143 |
| 36 | think it should be      | 0.0372 | 85 | t want to be                | 0.0143 |
| 36 | think that it s         | 0.0372 | 85 | the person who is           | 0.0143 |
| 42 | but i don t             | 0.0368 | 85 | you re going to             | 0.0143 |
| 43 | not be able to          | 0.0345 | 95 | and i don t                 | 0.0128 |
| 43 | they re going to        | 0.0345 | 96 | because it helps them       | 0.0127 |
| 45 | s very important for    | 0.0322 | 96 | because they need to        | 0.0127 |
| 46 | think it s good         | 0.0319 | 96 | if you want to              | 0.0127 |
| 47 | believe that it is      | 0.0299 | 96 | people who do not           | 0.0127 |
| 47 | it would be a           | 0.0299 | 96 | should be allowed to        | 0.0127 |
| 49 | in all public places    | 0.0297 | 96 | that they have to           | 0.0127 |
| 50 | a good idea for         | 0.0276 | 96 | there should be a           | 0.0127 |
| 50 | yes i think it          | 0.0276 | 96 | what they want to           | 0.0127 |
| 52 | it s going to           | 0.0273 |    |                             |        |

注. 96位が複数項目あったため、96位までの103項目を記載している。

注. 分析では少数点第9位までの差分値を求めているが、本資料では少数点第4位までの値を掲載。

資料3: SD における学習者の過剰使用語連鎖リスト(差分値の絶対値が大きい順) .....

| 順位 | 項目                      | 差分値     | 順位  | 項目                    | 差分値     |
|----|-------------------------|---------|-----|-----------------------|---------|
| 1  | i i don t               | -0.8941 | 51  | want to continue my   | -0.0495 |
| 2  | i don t like            | -0.7983 | 52  | uh uh i i             | -0.0492 |
| 3  | umm i don t             | -0.5313 | 52  | want to go to         | -0.0492 |
| 4  | i went to the           | -0.4104 | 54  | so it is very         | -0.0467 |
| 5  | so i want to            | -0.3936 | 54  | uh i can t            | -0.0467 |
| 6  | uh i don t              | -0.3856 | 54  | uh uh uh i            | -0.0467 |
| 7  | i don t have            | -0.3659 | 57  | don t i don           | -0.0461 |
| 8  | when i was a            | -0.3641 | 58  | a lot of things       | -0.0440 |
| 9  | i i want to             | -0.3638 | 58  | but i can t           | -0.0440 |
| 10 | uh i want to            | -0.3546 | 58  | so i want you         | -0.0440 |
| 11 | i want to i             | -0.3260 | 58  | uh yes uh i           | -0.0440 |
| 12 | i think it is           | -0.3084 | 62  | i don t go            | -0.0412 |
| 13 | i want to continue      | -0.2781 | 63  | uh i uh i             | -0.0407 |
| 14 | want to i want          | -0.2581 | 64  | i was a child         | -0.0385 |
| 15 | i can i can             | -0.2411 | 65  | yes yes yes yes       | -0.0372 |
| 16 | i want you to           | -0.2081 | 66  | when i when i         | -0.0367 |
| 17 | to i want to            | -0.2034 | 67  | it is it is           | -0.0364 |
| 18 | i went to your          | -0.1961 | 68  | but i want to         | -0.0362 |
| 19 | i i went to             | -0.1811 | 69  | are a lot of          | -0.0355 |
| 20 | don t want to           | -0.1304 | 69  | there are a lot       | -0.0355 |
| 21 | umm i want to           | -0.1203 | 71  | because i want to     | -0.0340 |
| 22 | i don t i               | -0.1069 | 71  | don t have enough     | -0.0340 |
| 23 | agree with this opinion | -0.1011 | 71  | i want to teach       | -0.0340 |
| 24 | i think she is          | -0.0983 | 74  | don t like uh         | -0.0336 |
| 25 | i want to go            | -0.0960 | 74  | i want to study       | -0.0336 |
| 26 | uh i think uh           | -0.0930 | 76  | uh i i don            | -0.0317 |
| 26 | when i went to          | -0.0930 | 76  | want i want to        | -0.0317 |
| 28 | i don t want            | -0.0920 | 78  | if i if i             | -0.0315 |
| 29 | i want to do            | -0.0893 | 79  | and i want to         | -0.0294 |
| 30 | but i don t             | -0.0855 | 79  | i like japanese food  | -0.0294 |
| 30 | i agree with this       | -0.0855 | 79  | want you to return    | -0.0294 |
| 32 | with my family and      | -0.0821 | 79  | we can we can         | -0.0294 |
| 33 | i want to work          | -0.0815 | 79  | yes yes uh i          | -0.0294 |
| 34 | don t think so          | -0.0783 | 84  | i went to a           | -0.0280 |
| 35 | i agree with it         | -0.0747 | 85  | yeah yeah yeah yeah   | -0.0275 |
| 36 | i i can t               | -0.0744 | 86  | and it is very        | -0.0273 |
| 36 | i want to uh            | -0.0744 | 87  | i want to umm         | -0.0272 |
| 38 | so i have to            | -0.0723 | 87  | there are many people | -0.0272 |
| 39 | t i don t               | -0.0707 | 89  | i like i like         | -0.0254 |
| 40 | i don t agree           | -0.0679 | 90  | it is not good        | -0.0252 |
| 41 | so i can t              | -0.0614 | 90  | to i have to          | -0.0252 |
| 42 | umm when i was          | -0.0614 | 90  | uh it s not           | -0.0252 |
| 43 | uh when i was           | -0.0584 | 90  | yes yes yes uh        | -0.0252 |
| 44 | i have to i             | -0.0582 | 94  | don t agree with      | -0.0233 |
| 45 | i can t i               | -0.0553 | 94  | i don t uh            | -0.0233 |
| 45 | i have i have           | -0.0553 | 94  | i i have to           | -0.0233 |
| 47 | i want i want           | -0.0524 | 94  | i think so too        | -0.0233 |
| 48 | i want to be            | -0.0522 | 94  | she is very strong    | -0.0233 |
| 49 | i have to earn          | -0.0517 | 94  | uh i agree with       | -0.0233 |
| 50 | uh i went to            | -0.0495 | 100 | i i didn t            | -0.0231 |

注. 分析では少数点第9位までの差分値を求めているが, 本資料では少数点第4位までの値を掲載。

資料4: SDにおける学習者の過少語連鎖の差分値リスト(差分値の絶対値が大きい順).....

| 順位 | 項目                    | 差分値    | 順位 | 項目                     | 差分値    |
|----|-----------------------|--------|----|------------------------|--------|
| 1  | i think it s          | 1.9706 | 40 | s going to be          | 0.0434 |
| 2  | it s it s             | 0.8493 | 40 | the same time i        | 0.0434 |
| 3  | we had to leave       | 0.3794 | 40 | well i think it        | 0.0434 |
| 4  | i don t think         | 0.3651 | 40 | yes i do it            | 0.0434 |
| 5  | i am going to         | 0.2385 | 40 | it s like a            | 0.0406 |
| 6  | i would like to       | 0.2212 | 57 | t think it s           | 0.0354 |
| 7  | i don t really        | 0.1951 | 58 | i didn t know          | 0.0348 |
| 7  | yes i do i            | 0.1951 | 59 | as you can see         | 0.0325 |
| 9  | think it s a          | 0.1734 | 60 | as you know i          | 0.0325 |
| 10 | it s kind of          | 0.1707 | 60 | don t know how         | 0.0325 |
| 11 | so i think it         | 0.1539 | 60 | i am trying to         | 0.0325 |
| 12 | don t think it        | 0.1340 | 60 | i ve been to           | 0.0325 |
| 13 | at the same time      | 0.1328 | 60 | it s a good            | 0.0325 |
| 14 | no i don t            | 0.1153 | 60 | not going to be        | 0.0325 |
| 15 | but i think it        | 0.0976 | 60 | or something like that | 0.0325 |
| 15 | i do it s             | 0.0976 | 60 | they are going to      | 0.0325 |
| 15 | it s a very           | 0.0976 | 60 | would like to continue | 0.0325 |
| 18 | thank you very much   | 0.0917 | 60 | so i think that        | 0.0275 |
| 19 | don t have a          | 0.0830 | 70 | they should be able    | 0.0271 |
| 20 | it was it was         | 0.0813 | 71 | and i don t            | 0.0251 |
| 21 | my friend and i       | 0.0734 | 72 | i have to pay          | 0.0246 |
| 22 | don t know what       | 0.0677 | 73 | a bit of a             | 0.0244 |
| 22 | don t think i         | 0.0677 | 74 | a it s a               | 0.0244 |
| 22 | i am not a            | 0.0677 | 74 | a little bit more      | 0.0244 |
| 22 | think it would be     | 0.0677 | 74 | and it s just          | 0.0244 |
| 22 | to be able to         | 0.0677 | 74 | and it s very          | 0.0244 |
| 27 | a lot of the          | 0.0650 | 74 | and it was it          | 0.0244 |
| 27 | i d like to           | 0.0650 | 74 | and they don t         | 0.0244 |
| 27 | it s going to         | 0.0650 | 74 | and we would like      | 0.0244 |
| 27 | we would like to      | 0.0650 | 74 | are going to be        | 0.0244 |
| 31 | a lot of people       | 0.0616 | 74 | because it s a         | 0.0244 |
| 32 | should be able to     | 0.0569 | 74 | but i am not           | 0.0244 |
| 32 | uh huh uh huh         | 0.0569 | 74 | but it s very          | 0.0244 |
| 34 | because i don t       | 0.0546 | 74 | can t afford to        | 0.0244 |
| 35 | as long as they       | 0.0542 | 74 | couldn t stand the     | 0.0244 |
| 35 | i used to go          | 0.0542 | 74 | didn t have any        | 0.0244 |
| 35 | t know how to         | 0.0542 | 74 | do it s a              | 0.0244 |
| 38 | have a lot of         | 0.0483 | 74 | don t i am             | 0.0244 |
| 38 | they don t have       | 0.0483 | 74 | don t think that       | 0.0244 |
| 40 | am not going to       | 0.0434 | 74 | had to leave the       | 0.0244 |
| 40 | and i think i         | 0.0434 | 74 | has a lot of           | 0.0244 |
| 40 | going to be a         | 0.0434 | 74 | i didn t see           | 0.0244 |
| 40 | have yes i did        | 0.0434 | 74 | i do have a            | 0.0244 |
| 40 | i am more comfortable | 0.0434 | 74 | i do i have            | 0.0244 |
| 40 | i am not going        | 0.0434 | 74 | i go to a              | 0.0244 |
| 40 | i have yes i          | 0.0434 | 74 | i think you know       | 0.0244 |
| 40 | i think it would      | 0.0434 | 74 | i want to get          | 0.0244 |
| 40 | i think that s        | 0.0434 | 74 | i was able to          | 0.0244 |
| 40 | i was working at      | 0.0434 | 74 | i was in the           | 0.0244 |
| 40 | i was working in      | 0.0434 | 74 | i would i would        | 0.0244 |
| 40 | it s a it             | 0.0434 | 74 | i would prefer to      | 0.0244 |
| 40 | s a it s              | 0.0434 | 74 | if you have a          | 0.0244 |

資料4: SD における学習者の過少語連鎖の差分値リスト(差分値の絶対値が大きい順).....

| 順位 | 項目                   | 差分値    | 順位 | 項目                  | 差分値    |
|----|----------------------|--------|----|---------------------|--------|
| 74 | in the first place   | 0.0244 | 74 | think it s okay     | 0.0244 |
| 74 | is going to be       | 0.0244 | 74 | think it should be  | 0.0244 |
| 74 | it s a really        | 0.0244 | 74 | to finish our meals | 0.0244 |
| 74 | it s just a          | 0.0244 | 74 | to pay for my       | 0.0244 |
| 74 | it s just the        | 0.0244 | 74 | uh let me see       | 0.0244 |
| 74 | it was kind of       | 0.0244 | 74 | was working in a    | 0.0244 |
| 74 | it was really nice   | 0.0244 | 74 | we weren t able     | 0.0244 |
| 74 | know i don t         | 0.0244 | 74 | well i i think      | 0.0244 |
| 74 | like i don t         | 0.0244 | 74 | well i think that   | 0.0244 |
| 74 | maybe once a month   | 0.0244 | 74 | weren t able to     | 0.0244 |
| 74 | me and my friend     | 0.0244 | 74 | yes i have yes      | 0.0244 |
| 74 | no i didn t          | 0.0244 | 74 | you can you can     | 0.0244 |
| 74 | right now so i       | 0.0244 | 74 | you don t have      | 0.0244 |
| 74 | s it s a             | 0.0244 | 74 | you know i am       | 0.0244 |
| 74 | s kind of a          | 0.0244 | 74 | you know i think    | 0.0244 |
| 74 | so no i don          | 0.0244 | 74 | you re going to     | 0.0244 |
| 74 | so that s why        | 0.0244 | 74 | you very much sir   | 0.0244 |
| 74 | that s going to      | 0.0244 | 74 | you want me to      | 0.0244 |
| 74 | the number of people | 0.0244 |    |                     |        |

注. 74位が複数項目あったため, 74位までの141項目を記載している。

注. 分析では少数点第9位までの差分値を求めているが, 本資料では少数点第4位までの値を掲載。