

日本語母語話者と学習者による定式表現の 産出過程の研究

蘇 振 軍

(2018年10月4日受理)

Production of Japanese Formulaic Language by Native and Nonnative Japanese Speakers

Zhenjun Su

Abstract: This study examined the production of two types of Japanese formulaic language, i.e., collocations and idioms, through a spoken dictation task in which native speakers and advanced second-language (L2) learners of Japanese listened to a sentence with or without a formulaic string, performed an intervening task, and spoke the sentence aloud. The accuracy of production and recall time was measured as the dependent variables. The results indicated that the native speakers' production was nearly perfect; however, the L2 learners produced many nonnative strings for collocations and idioms, and their idiom production accuracy was poorer than that for collocation. The speakers' recall time showed that both types of formulaic language facilitated language processing for native speakers and L2 learners. In addition, L2 learners processed collocations and idioms in the same manner; however, native speakers processed collocation faster than they did idioms. These results indicate that native speakers store formulaic strings as chunks in their mental lexicon, but learners store smaller units (i.e., words and morphemes) than chunks.

Key words: formulaic language, collocation, idiom, language processing, mental representations

キーワード：定式表現, コロケーション, 慣用句, 言語処理, 心内表象

1. はじめに

第二言語学習者（以下、L2学習者）は上級者であっても、文法的には正しいが、目標言語¹としては不自然な表現を産出することが少なくない。そのうち、「知識を勉強する」や「気分が嫌だった」のような語と語の組み合わせに現れる共起語選択の誤り（大曾・滝沢, 2003; 佐治, 1992）や、「長い髪がある」という誤用に現れるように学習者は文法知識を用いて構築可能な様々な文を創造する。一方、母語話者は文法に

よって作り出すというより、その一部の語の固まりを場面や文脈に応じて繰り返して使用することで言語を産出する（畑佐, 2012; Pawley & Syder, 1983）。この母語話者が繰り返して使用する語の固まりは定式表現（formulaic language）と呼ばれ、一般に take a break（休憩する）や pass the buck（責任転嫁する）のような複数の単語が組み合わさって、1つのまとまった意味や機能を表す表現を指す（Wray, 2009）。近年、英語を対象とした第一言語（以下、L1）研究では、この定式表現の役割について様々な検討がなされている。例えば、コーパス言語学の分野では、構成語どうしの特徴を基に、大規模コーパスで高頻度に現れる表現を抽出し、その特徴を分析している。その結果、英語では30%から60%の表現が定式表現と見なされることや、定式表現が文を構成する重要な要素となってい

本論文は、課程博士候補論文を構成する論文の一部として、以下の審査委員により審査を受けた。

審査委員：畑佐由紀子（主任指導教員）、金 愛蘭、白川博之、宮谷真人

ることがわかっている (Altenberg, 1998)。

また、言語心理学の分野においては、定式表現は母語話者の長期記憶に固まりで貯蔵され、統語処理がなされないため、文法と語彙によって構築された表現(以下、自由結合)よりも速く処理されることが実験的に検証されている (Conklin & Schmitt, 2012)。それゆえ、話し手と聞き手の処理負荷を軽減することができると考えられている。これに対し、L2を対象とした研究でも、定式表現が学習者のL2の自然さ、流暢さ及び適切さに影響する可能性が指摘されている (甄, 2009)。その一方で、学習者は定式表現に関する知識が乏しく、語彙と文法知識を用いて文を産出する傾向があるため、上級者でも母語話者が使用しない不自然な文を産出することが指摘されている (Wray, 2008)。したがって、学習者が定式表現をどのように処理・産出するかを解明することは、L2の処理過程と習得の解明に重要な示唆をもたらすと考えられるが、日本語を対象とした研究は極めて少ない。

そこで、本稿では母語話者と学習者がどのように定式表現を産出するのか、また、定式表現の種類によって産出過程が異なるのかについて検討する。これにより、定式表現がどのように母語話者と学習者の心内辞書に貯蔵されているのか、そして、母語話者と学習者の間に相違があるのかを探る。

2. 先行研究と本研究の目的及び研究課題

2.1 先行研究の概観と問題の所在

英語の定式表現研究では、その種類や機能、処理過程、習得についてコーパス言語学、心理学、脳科学など様々な分野で研究がなされている (Biber, 2009; Wray, 2009)。コーパス言語学では、母語話者コーパスを基に、典型的な定式表現の種類には、(1) kick the bucket (死ぬ) のように句の意味が構成語 (kick, the, bucket) から推測できず、固まりとして慣用的に意味を表す慣用語、(2) commit suicide (自殺する) のように構成語から句全体の意味推測ができ、2つの実質語が慣用的に共起するコロケーション、(3) that is to say (すなわち) のようにコーパスで高頻度に出現し、実質語と機能語で構成される語彙束 (lexical bundles) が挙げられている (Biber, 2009; Wray, 2002, 2008)。

心理学的研究では、こうした定式表現は統語処理を必要とせず、固まりとして心内辞書に貯蔵され、処理されているという固まり仮説 (holistic hypothesis) が立てられている (Conklin & Schmitt, 2012)。この

背景には、コーパスはあくまでも二次的な言語データであり、コーパス言語学に基づく分析のみでは定式表現がどのように処理されるか解明できないと考えられていることがある。よって、定式表現の本質を明らかにするためには、母語話者の心内辞書に定式表現がどのように貯蔵されるかを検証することが有効だと考えられている (Wray, 2002, 2009)。

固まり仮説を検証した研究には様々な実験方法が用いられているが、主に break the ice のような定式表現と、break the cup のような定式表現の実質語を入れ替えた自由結合の反応時間や産出割合を比較することにより、定式表現が言語処理にもたらす効果と心内表象について研究がなされている。以下、実験方法ごとに先行研究をまとめる。

まず、文が聴覚的に提示された後に、何らかの介入課題を挟んでから再生させるという口頭再生課題がある。実験参加者は材料を聞き、それを理解しながら覚えようとする。介入課題をかけることで作動記憶にある情報をそのままリピートすることができなくなる。介入課題を経た後でも、流暢に再生できた表現は長期記憶に固まりで貯えられているというロジックである。Schmitt et al. (2004) はこのような課題を用い、母語話者と学習者を対象に、語彙束に焦点を当てて固まり仮説を検討した。その結果、母語話者にとっては、全ての語彙束は固まりで産出されたわけではなく、意味的透明性が高い語彙束のほうが流暢に再生できた。この結果を基に、Schmitt et al. は母語話者の心内辞書で意味的透明性が高い定式表現が固まりとして貯蔵されていると結論付けた。また、再生できた表現の長さは一般に一回で作動記憶に貯蔵されうるとされる文の長さよりも長かった。このことから、定式表現はその処理過程で内部構造の分析を要しないため、一度に処理される言語資源数を下げることができ、作動記憶内での処理負荷を下げることができると主張した。一方、学習者においては、上級者であっても完全に再生できた文が非常に少なく、独自に構築した表現がほとんどであった。また、実質語をキーワードとして、文を再構築する特徴がみられた。このことから、Schmitt et al. は、L2においては定式表現が固まりとして心内辞書に貯蔵されておらず、産出が困難であると述べた。

産出課題のほか、理解課題としてフレーズの文法性判断課題も多く用いられる (Jiang & Nekrasova, 2007; Swinney & Cutler, 1979)。この課題では、定式表現が固まりで心内辞書に貯蔵されているのであれば、表現が提示されると、すぐに心内辞書の当該項目が照合されるから、句の統語処理が必要ではなくな

る。その結果、定式表現の方が自由結合より反応時間が短く、正答率も高くなるというロジックである。Swinney & Cutler (1979) はこのような課題を用い、慣用句を対象に実験を行った結果、自由結合より判断速度が速いことがわかった。また、李 (2012) は日本語の「名詞+を+動詞」からなるコロケーションを材料に、コーパスに出現した頻度や、語と語の共起関係の強さを表す相互情報量を操作した実験を行った。その結果、母語話者も学習者も頻度の低いコロケーションよりも頻度の高いコロケーションの方を速く処理していることが分かった。

さらに、産出を伴う理解課題として自己ペース読み上げ課題がある。このような課題では、文が提示されると同時に、文を理解しながら声を出して読み上げるため、産出を伴う理解課題と言える。この課題には単語単位と、定式表現単位、文単位という3つのパターンがある (Tremblay et al., 2007; 李・賈, 2013)。英語母語話者を対象とした Tremblay et al. (2007) と、英語学習者を対象とした李・賈 (2013) では、同様の結果が得られ、定式表現単位と文単位の読み上げ課題は、自由結合よりも処理速度が速く、固まり仮説を支持する結果となった。

この他に、脳内活性化範囲を用いた固まり仮説の検証もなされている。定式表現が固まりとして処理される場合、統語処理がなされないため、統語処理を要する自由結語よりも脳内活性化範囲が狭くなるはずである。そこで、梁・松野・杉浦 (2008) は英語のコロケーションを対象に脳内活性化部位を測る実験を行った。その結果、母語話者はコロケーションの活性化範囲が自由結合より小さかったのに対し、学習者はコロケーションも自由結合も活性化範囲が広がった。趙・安永・小島・林 (2016) は日本語のコロケーションを対象に脳活性化部位を測定し、同様の結果を得た。

以上、英語を対象とした研究結果をまとめると、L1では定式表現が分析されずに固まりとして心内辞書に貯蔵される可能性が高いことが分かった。しかし、L2では定式表現が速く処理されるという報告はあるものの、固まりで処理されているか否かについては疑問が残る。また、日本語の研究からもコロケーションが速く処理され、母語話者が固まりとして記憶している可能性が示唆されたが、学習者がコロケーションをどのように処理・産出しているのかは解明されていない。加えて、以下のような課題も残されている。

まず、先行研究では、コロケーション、慣用句、語彙束が個別に検討されているものと、定式表現の種類を区別せずに検討されているものがある。定式表現の種類によって文法構造の規範性や語彙の共起強度、意

味的透明性が異なることから (Wray, 2008)、反応時間が短いからといって、すべての定式表現が固まりで処理されると仮定することはできない。例えば、慣用句のように構成語から句の意味が推測できない句は固まりとして考えやすいが、「注目を集める」のように意味的透明性が高いコロケーションは語の結合度が強いいため、固まりではなくとも、前の語が呈示されると後ろの語が活性化され、処理が速くなると解釈することも可能である。そこで、コロケーションと慣用句の頻度を統制し、意味的透明性を操作して定式表現の処理に与える影響を検討することで固まり仮説を検証する必要があると考える。

次に、先行研究では、主に理解課題によって固まり仮説が検証されている。しかし、理解実験のみで心内辞書に定式表現が固まりで貯蔵されていると主張するのは不十分だと思われる。定式表現が固まりで貯蔵されているのであれば、産出においても自由結合よりも、心内辞書に速くアクセスし、且つ正確に産出できるはずである。しかし、産出課題を用いた研究は Schmitt et al. (2004) のみである。この研究では再生成績のみを分析対象としており、定式表現に関する明確な結果が得られていない。そこで、再生成績と再生に要した時間 (以下、課題遂行時間) の両方を検討することで、固まり仮説をより厳密に検討することができると考える。

最後に、先行研究では、母語話者と学習者の両方を対象とした研究が少ない。そのため、母語話者と学習者がどのような表現を固まりで処理し、どのような表現を統語的に処理するか、そしてこの両者の処理過程にはどのような類似点と相違点があるのかは明らかにされていない。母語話者と学習者を比較することで、学習者の定式表現の処理過程を明らかにでき、L2における語彙、文法指導に対する有効な情報が提供できると考える。

2.2 本研究の目的及び研究課題

以上を踏まえ、本研究では日本語母語話者と学習者を対象に、コロケーションと慣用句を用いて意味的透明性を操作し、定式表現が日本語の処理・産出に与える影響を、産出課題を用いて検討する。これにより、母語話者と学習者の定式表現の心内表象を探ることを目的とする。コロケーションに関しては様々な定義があるが、本稿では国広 (2007) の「連語」の定義を基に、単語どうしの結びつきは固定しているが、全体の意味が解釈可能なものをコロケーションと定義する。構成語の意味総和が句全体の意味となる表現であるため、意味的透明性の高い定式表現と言える。一方、慣用句は句の意味が構成語の意味から推測できない語の

組み合わせである(国広, 1985)。ゆえに、意味的透明性の低い定式表現と言える。そして、自由結合は先行研究に倣い、コロケーションと慣用句の実質語を入れ替えた自然な日本語とする。

以上のことを総合すると、本研究の研究課題は以下の通りである。

【研究課題1】 母語話者と学習者はコロケーションと慣用句という定式表現を、自由結合より容易に再生するのか、母語話者と学習者の間に相違があるのか。

【研究課題2】 母語話者と学習者において、意味的透明性の高低は定式表現の再生に影響を与えるのか、母語話者と学習者の間に相違があるのか。

以上の課題を通して、母語話者と学習者が定式表現をどのように処理するのか、そして定式表現がどのように心内辞書に貯蔵されているのかを考察する。

3. 実験

3.1 実験参加者

参加者は母語話者20名と、中国語を母語とする日本語上級学習者30名であった。母語話者の内訳は大学院生が18名、大学生が2名(男性11名、女性9名)であり、平均年齢は23歳であった。学習者は大学院生が29名、大学院を卒業したばかりの社会人が1名であり(男性10名、女性20名)、平均年齢は25歳であった。本実験に参加した時点で、平均日本語学習歴は7.9年であり、日本滞在歴は平均で3.2年であった。全員が日本語能力試験N1を取得していた。

3.2 実験計画

母語話者と学習者において、ターゲット項目の再生成績と課題遂行時間をそれぞれ従属変数として、 $2 \times 2 \times 2$ の3要因計画を用いる。第1の要因は実験参加者であり、母語話者と学習者の2水準である。第2の要因は表現タイプであり、定式表現の有無で、定式表現と自由結合の2水準である。第3の要因は定式表現のタイプで、2種の定式表現を対象とし、コロケーションと慣用句の2水準である。すべての計画において、実験参加者は参加者間変数、句のタイプは参加者内変数であった。

3.3 材料

以下の基準に基づき、コロケーション文と慣用句文を9文ずつ、計18文の課題文、コロケーションの統制文と慣用句の統制文を9文ずつ計18文、合計36文作成した。

(1) コロケーションは、まず『現代日本語書き言葉均衡コーパス』(以下、BCCWJ)における出現頻度が上位20位以内の動詞と形容詞を選択し²、これらの

単語からなるコロケーションを『研究社日本語コロケーション辞典』(柏崎・藤村・鈴木, 2012)から20個選定した。(2) 慣用句は基幹慣用句リスト(呉, 2016)から、複数のジャンルに共通に現れ、且つ使用頻度が高いものを20個選定した。(3) (1)と(2)には重複があり、「念を押す」のような表現がコロケーション、慣用句の両方に含まれていた。そこで、Wray (2008)が設定した定式表現の母語話者の判定基準を基に、日本語教育を専攻する母語話者13名(大学院生6名、大学生7名)に、選定した40の定式表現について、表現の構成語から句全体の意味をどの程度推測できるかを、「不透明」、「やや不透明」、「どちらともいえない」、「やや透明」と「透明」の5段階で評価させ、それぞれに-2点、-1点、0点、1点、2点を付与した。そして、平均点が1点以上の項目をコロケーション(例: 注目を集める)、-1点未満の項目を慣用句(例: 棚に上げる)に分類した。平均点が-1点以上1点未満の項目は材料から外した。その結果、22個の表現を除外し、コロケーションと慣用句それぞれ9個を材料とした。さらに、国立国語研究所とLago言語研究所が開発したNINJAL-LWP for BCCWJを用いて、各コロケーションと慣用句の頻度を抽出した上で、独立標本 t 検定を行い、有意差がないことを確認した($t = 1.89, df = 16, p = .077, r = .43$)。(4) 自由結合はコロケーションと慣用句の目的語、または主語を入れ替えて作成した。入れ替えた語と述語で構成した自由結合が自然な日本語であることを、『研究社日本語コロケーション辞典』(柏崎他, 2012)と『日本語作文作成支援システムなつめ』(東京工業大学留学生センター仁科研究室, 2012)で確認した。また、入れ替えた語と定式表現に含まれていた語の頻度はBCCWJを用いてすべて統制した。それぞれの頻度について、独立標本 t 検定を行ったところ、有意差はなかった($t = 0.19, df = 34, p = .989, r = .03$)。さらに、入れ替えた語のモーラ数も定式表現に含まれていた語のモーラ数と同じになるよう統制した(例: 腰(コシ)が低い、価値(カチ)が低い)。(5) 選定したコロケーションと慣用句、自由結合を用いて文を作成し、日本語教育を専攻する母語話者(大学院生1名、教師1名)に文の自然さと正確さを確認してもらった。また、統制文は文法構造と長さを課題文と同様に統制した。文の長さは両者共に平均29モーラ、最も長い文は40モーラで、最も短い文は22モーラであった。そして、各文のレベルについて、日本語読解学習支援システムのリーディングチュウ太(川村・北村, 1998)で判定し、「語彙レベルがやさしい」であることを確認した。(6) 以上のように作成した課題文と統制文を用い、東京圏

出身の日本語母語話者に録音してもらった。

3.4 装置

本研究の実験プログラムは Superlab Pro (Cedrus 社製 Version 5.0) を用いて作成した。また、実験には、パーソナルコンピュータ (Epson Endeavor NJ3900E) と周辺機器を使用した。実験参加者の発話を、参加者の許可を得た上で IC レコーダー (Panasonic RR-X350) を用いて録音した。

3.5 手続き

実験は個別に行われた。練習試行を 3 試行行った。練習試行により実験参加者が実験方法を理解したことを確認した後、本試行を 36 試行行った。全ての文はランダムに聴覚呈示されるように設定した。各練習試行と本試行は以下のような流れで行った。まず、コンピュータ画面の中央に注視点が呈示され、指定したキーを押すと、次の画面で文が聴覚的に提示された。提示が終わり、キーを押すと、次の画面に引き算、または足し算課題が呈示された。課題に答え、キーを押すと、次の画面で聞いた文を口頭で再生するよう指示した。再生が終わり、キーを押すと、次の文を準備する段階である注視点のある画面が再び呈示された。

本試行終了後、未知単語、未知コロケーション、未知慣用句の有無の確認や、日本語学習歴、日本滞在歴に関する質問紙調査を行った。

4. 再生成績の結果と考察

母語話者は課題文も統制文もほぼ完全に再生できたのに対し、学習者は完全に再生できた文が非常に少なく、ほとんど自分なりに再構築した文である。また、言いよどみが多かった。定式表現に関する知識の有無を把握するため、本節では母語話者と学習者の定式表現及び自由結合の再生成績について検討する。

4.1 再生成績の結果

表 1 は採点基準と用例を示したものである。ターゲット項目の再生において、助詞抜きと助詞の入れ替え、述語の正確な活用の場合には 2 点、ターゲット項目の意味を保持したままの述語の入れ替えを 1 点とした。また、述語がなかなか思い出せず、言いよどんで再生できた場合は 2 つの構成語が共起に関する知識が乏しいと判断し、1 点にした。最後に、再生できていなかった項目と定式表現が未知語として確認された項目は 0 点とした。そのうち、未知語率を算出した結果、コロケーションは 1.48%、慣用句は 27.04% であった。

この基準に基づいて採点した結果 (各項目の平均得点と各点数に対応する人数を表 2 に示す)、母語話者の再生成績は 1.91 点 (95.5%) ~ 1.98 点 (99.0%) であり、

コロケーションも慣用句もほぼ満点を獲得した。これに対し、学習者はターゲット項目の成績にばらつきがあり、母語話者との再生成績は不等分散であった。そのため、実験参加者を独立変数とする分散分析はできなかった。そこで、学習者のデータのみを用いて課題句と統制句の成績に対して、2 (表現タイプ: 定式表現, 自由結合) × 2 (定式表現タイプ: コロケーション, 慣用句) の分散分析を行った。その結果、表現タイプの主効果 ($F(1,8)=0.91, p=.368, \eta^2=.10$)、定式表現タイプの主効果 ($F(1,8)=2.15, p=.181, \eta^2=.21$)、表現タイプ × 定式表現タイプの交互作用 ($F(1,8)=0.61, p=.456, \eta^2=.07$) はいずれも有意な差がみられなかった。

表 1 採点基準及び用例

	採点基準	用例
2 点	完全に再生できた項目	念を押す
	助詞抜き	気持ち悪い
	助詞の入れ替え	調子のいい
	述語の活用	役に立ちます
1 点	述語の入れ替え	検索をする
	言いよどんで再生した項目	注目をす、す、す る? 集める?
0 点	再生不可と未知項目	

4.2 考察

本実験では、定式表現に関する知識の有無を把握するため点数付けして再生成績を算出した。その結果、母語話者はほぼ満点を獲得したことがわかった。本実験で用いた文の長さは最も長いものが 40 モーラであり、母語話者にとってやや短く、再生が容易であったことが原因であると考えられる。

一方、学習者の再生成績は定式表現と自由結合の間に有意差がみられなかった。これは定式表現に学習者が再生できない項目が多かったため、定式表現の成績が下がったことが原因として考えられる。自由結合は構成語の入れ替えが最も自由であるがゆえに、1 点を取った人数が定式表現より多かった。例えば、「検索をかける」という表現は 9 割程度の学習者が「検索する」を産出した。一方、定式表現には多くの学習者が 0 点を取った項目が含まれていた。例えば、「理解を得る」というコロケーションには、「理解がえられる」や「理解をもらえた」などのようにミス結合が多く、30 人中 20 人が 0 点を取った。また、慣用句は構成語の入れ替えがコロケーションと比べると少なく、再生できたか、できなかったかの 2 パターンがほとんどであった。慣用句のうち、「虫がいい」と「腰が低い」は未知語率がそれぞれ 86.67%、53.33% と高く、30 人

表2 学習者における再生成績及び、点数ごとの人数

課題句		成績	2点	1点	0点	統制句	成績	2点	1点	0点	
コロケーション	迷惑をかける	1.97	29	1	0	検索をかける	1.20	12	12	6	
	調子がいい	1.80	27	0	3	食べ物がいい	1.50	21	3	6	
	口がうまい	1.70	24	3	3	水がうまい	1.73	24	4	2	
	力を入れる	1.57	23	1	6	名前を入れる	1.80	27	0	3	
	気持ちが悪い	1.53	23	0	7	政府が悪い	1.93	29	0	1	
	注目を集める	1.43	20	3	7	アドバイスを集める	1.73	25	2	3	
	影響を受ける	1.33	20	0	10	相談を受ける	1.63	24	1	5	
	眠りが深い	0.83	12	1	17	悩みが深い	1.23	18	1	11	
	理解を得る	0.63	9	1	20	支援を得る	0.93	9	10	11	
	平均	1.42					1.52				
	SD	0.41					0.31				
慣用句	役に立つ	1.87	28	0	2	庭に立つ	1.10	16	1	13	
	間に合う	1.80	27	0	3	指に合う	1.30	19	1	10	
	耳が遠い	1.73	25	2	3	夏が遠い	1.90	28	1	1	
	口を挟む	1.33	20	0	9	足を挟む	1.67	24	2	4	
	鼻が高い	1.23	18	1	11	価値が高い	1.30	19	1	10	
	念を押す	1.20	18	0	12	ベルを押す	1.43	21	1	8	
	棚に上げる	0.87	13	0	17	陸に上げる	0.97	14	1	15	
	腰が低い	0.73	11	0	19	価値が低い	1.63	24	1	5	
	虫がいい	0.33	5	0	25	パンがいい	1.70	24	3	3	
	平均	1.23					1.43				
	SD	0.49					0.38				

中19人から25人が再生できなかった。これらのことから、学習者は定式表現、特に慣用句に関する知識が乏しいこと、共起関係に関する不十分な知識を用いて語と語を組み合わせることで日本語を産出していることがわかる。よって、学習者はL2を産出する際に、語彙体系の中から単語を選び、文法知識を用いて組み合わせるという自由選択原理 (Sinclair, 1991) にしたがっていると見えよう。

また、成績が1.33点（2点が20人）以上であったコロケーションは「迷惑をかける」、「調子がいい」、「口がうまい」、「力を入れる」、「気持ちが悪い」、「影響を受ける」、「注目を集める」の7項目であった。一方、慣用句は「役に立つ」、「間に合う」、「耳が遠い」、「口を挟む」の4項目であった。本研究で用いたコロケーションの平均モーラ数は慣用句より長いにもかかわらず、慣用句よりも再生されやすかったことが言える。

さらに、1.7点（2点が24人）以上の項目には、3つのコロケーション（「迷惑をかける」、「調子がいい」、「口がうまい」）と3つの慣用句（「役に立つ」、「間に合う」、「耳が遠い」）が含まれていた。これらの項目は初級から習う項目であり（秋元・有賀, 1996）、教

室内インプットとして多い表現である。そのため、学習者にとって身近であるがゆえに、再生成績が最も高かったと思われる。さらに、これら6つの定式表現文の再生状況を確認したところ、ターゲット項目はそのまま流暢に再生されていたが、ターゲット項目の前後の文は学習者が自分なりに再構築した表現となっていた。したがって、これらの6つの定式表現は学習者の心内辞書に語彙化されて固まりで貯蔵されている可能性があると思われる。この結果は、学習者はほとんど再生できず、単語をキーワードとして再構築したという結果が得られた Schmitt et al. (2004) とは異なっている。

以上、再生成績を用いて定式表現に関する知識の有無について質的な分析を中心に検討した。その結果、母語話者は満点を獲得したのに対し、学習者にとっては固まりとして処理されている可能性が高い定式表現とそうでないものがあることがわかった。そこで、次節では再生成績が高く、固まりとして処理される可能性が高い定式表現及び、それに対応する自由結合がどのように処理・産出されていたのかについて課題遂行時間を用いて検討する。

5. 課題遂行時間の結果と考察

5.1 課題遂行時間の分析と結果

本節では定式表現の処理・産出パターンを検討するためにターゲット項目の課題遂行時間を比較する。表2で示したもののうち、課題句と統制句の再生成績が両方ともに高い項目を選定した。その結果、「調子がいい」、「口がうまい」、「力を入れる」、「気持ちが悪い」と「注目を集める」というコロケーション(平均点1.94)とこれに対応する統制句(平均点1.95)をそれぞれ5項目選定した。慣用句は「間に合う」、「耳が遠い」、「口を挟む」、「鼻が高い」と「念を押す」という項目(平均点1.96)とこれに対応する統制句(平均点1.91)もそれぞれ5項目選定した。頻度の影響を排除するため、コロケーショングループと慣用句グループの頻度について独立標本 t 検定を行い、両群に有意差がないことを確認した($t=0.82$, $df=8$, $p=.433$, $r=.28$)。これらの項目と統制句を完全に再生できた学習者は18名であり、これに合わせて母語話者も18名を選定した。録音した音声データから上記の項目を切り出し、課題遂行時間を算出した。本研究で用いたコロケーションは慣用句よりやや長いいため、課題遂行時間を1モーラ単位で算出した。

さらに、再生時間に用いた10項目の定式表現の再生状況を確認した結果、全ての項目は非常に流暢に再生されていた。しかし、母語話者と学習者における再生時間のローデータも対数変換した値も等分散でないため、実験参加者を独立変数とする分散分析はできなかった。だが、母語話者は全ての項目においても2種の定式表現においても(64.21ms)、明らかに学習者より速く再生していた(図1、母語話者はNS、学習者はNNS、コロケーションはCol、慣用句はKan、コロケーション統制句はColT、慣用句統制句はKanT)。そこで、母語話者と学習者のデータを対数変換した値を用いてそれぞれ2(表現タイプ:定式表現、自由結合)×2(定式表現タイプ:コロケーション、慣用句)の分散分析を行った。その結果、母語話者においては表現タイプの主効果($F(1,17)=17.03$, $p<.001$, $\eta^2=.50$)、定式表現の主効果($F(1,17)=39.89$, $p<.001$, $\eta^2=.70$)、表現タイプ×定式表現タイプの交互作用($F(1,17)=7.27$, $p=.015$, $\eta^2=.30$)はいずれも有意であった。単純主効果検定を行った結果、コロケーションも($F(1,17)=24.99$, $p<.001$, $\eta^2=.60$)、慣用句も($F(1,17)=34.83$, $p<.001$, $\eta^2=.67$)、自由結合より有意に速く再生されていた。そして、コロケーションは慣用句よりも有意に速く再生された($F(1,17)=4.91$, $p=.041$, $\eta^2=.22$)。コロケーション統制句は慣用句統制句よりも速く処理されて

いた($F(1,17)=19.26$, $p<.001$, $\eta^2=.53$)。一方、学習者においては、表現タイプの主効果は有意であった($F(1,17)=15.43$, $p=.001$, $\eta^2=.48$)が、定式表現タイプの主効果($F(1,17)=0.83$, $p=.374$, $\eta^2=.05$)、表現タイプ×定式表現の交互作用($F(1,17)=0.09$, $p=.767$, $\eta^2=.01$)は、いずれも有意な差がみられなかった。定式表現は自由結合より速く再生されたことが示されている。

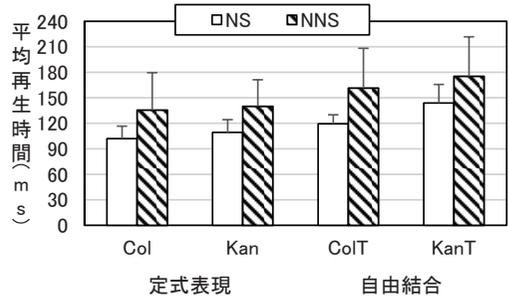


図1 母語話者と学習者におけるターゲット項目の課題遂行時間及び標準偏差

5.2 考察

課題遂行時間を比較した理由は定式表現が再生を促進するのか、意味的透明性が定式表現の再生に影響を与えるのかを検討することにより、母語話者と学習者がどのように定式表現を処理するのか、どのように心内辞書に貯蔵されているのかを検討するためであった。

分析の結果、母語話者でも学習者でも、定式表現は自由結合より速く再生されていた。この結果について、自由結合は構成語の共起関係が弱く、各構成語が心内辞書に個別に貯蔵され、統語処理を通して再構築されたため時間がかかったと考えられる。一方、コロケーションと慣用句は構成語の共起関係が強いため、1つの固まりとして心内辞書から取り出したか、若しくは統語処理が行われてもアクセス速度が速くなったかといった2通りの可能性が考えられる。

また、意味的透明性については、母語話者はコロケーションを慣用句よりも速く再生していた。これは以下の理由によると思われる。コロケーションは構成語の意味総和が句の意味となっている。だが、慣用句は固まりとして処理される場合は慣用的な意味が処理され、統語処理がなされる場合は、構成語から類推される意味が処理されるという2つの可能性を含む表現である。この場合、もし固まりで処理されないならば、統語処理がなされ、処理速度が遅延する。しかし、慣用句の再生はコロケーションより遅かったが統制句より速かったことから、コロケーションも慣用句も固まりで心内辞書から取り出したものの、慣用句の文字通りの意味が再生に妨害した可能性が考えられる。ま

た、コロケーションもコロケーション統制句も、慣用句と慣用句統制句より有意に速く再生されていたことから、コロケーションとコロケーション統制句の文脈は慣用句と慣用句統制句の文脈より再生されやすかった可能性も否めない。一方、学習者においては、コロケーションと慣用句の間に課題遂行時間の差がなかったことから、意味的透明性は学習者の再生に影響を与えないことがわかった。このことから、課題遂行時間に用いたコロケーションは慣用句のように固まりで処理された可能性がある。

以上、課題遂行時間を比較した結果、母語話者か学習者に関わらず、定式表現が自由結合より再生されやすく、言語産出を促進すること、母語話者では意味的透明が言語産出に影響を与えていたが、学習者では影響がなかったことが明らかとなった。次節では、再生成績と課題遂行時間の分析結果を基に、母語話者と学習者がどのように定式表現を処理・産出するのか、定式表現がどのような心内表象を有するのかについて考察する。

6. 総合考察

本研究では、意味的透明性の異なる定式表現が母語話者と学習者の言語産出を促進するかどうかについて、口頭再生課題を用いて検討した。実験に用いた定式表現は母語話者にとっては頻度の高い表現であった。統制句に入れ替えた語の頻度は定式表現の構成語との間に差が出ないように統制されていた。また、課題文と統制文の文法構造も統制し、すべての文が実際に起こり得る自然な文であることを母語話者に確認した。そのため、これらの要因による影響を排除したと言える。

その上で、課題1については定式表現と自由結合の再生成績と課題遂行時間を比較した。その結果、母語話者でも学習者でも、定式表現と自由結合の間に再生成績の差はなかった。しかし、課題遂行時間においては定式表現のほうが自由結合より有意に速く再生された。これは理解課題を用いた Jiang & Nekrasova (2007) と Tremblay et al. (2007)、李 (2012) の結果と一致しており、定式表現が言語産出を促進すると言える。また、産出課題を用いた Schmitt et al. (2004) は再生成績だけを分析対象としたが、本研究では課題遂行時間も分析し、母語話者だけでなく学習者も定式表現を自由結合より容易に心内辞書にアクセスすることができていた。それゆえ、自由結合より言語処理の速度が速くなり、流暢に再生されていた。

また、課題2で設定した意味的透明性の高低は定式

表現の処理に与える影響についても再生成績と課題遂行時間を用いて分析した。その結果、母語話者においても学習者においても、再生成績では意味的透明性が影響を与えなかったが、課題遂行時間では異なっていた。母語話者では、コロケーションが慣用句より速く再生されていたが、両方共に流暢に再生されていた。よって、定式表現は母語話者の心内辞書に固まりで貯蔵されており、固まりとして産出されたことで課題遂行時間が短かったと言える。しかし、コロケーションは慣用句より再生されやすかった。この原因としては慣用句の処理過程において慣用的な意味だけではなく文字通りの意味が活性化され、産出が遅くなったか、コロケーションの文脈が慣用句の文脈より再生されやすかった可能性が考えられる。

一方、学習者においては、コロケーションと慣用句の再生成績も課題遂行時間も違いがなかった。再生成績に用いた18項目の定式表現の再生状況を確認した結果、コロケーションは構成語の入れ替えが慣用句より比較的緩いがゆえに、より多くのバリエーションがみられた。未知語の割合も慣用句よりはるかに少なかった。これらの結果から、学習者はL2の文法と語彙知識を利用して自由選択原理に従って文を産出することが言える。しかし、課題遂行時間の分析に用いた定式表現が自由結合より速く再生されていたことから、流暢に運用できる程度に習得している定式表現に関しては、学習者も母語話者のように文を細かく分析せずに言語を処理できるようになると考えられる。

しかし、本研究でも学習者の処理速度は母語話者に著しく劣っていた。また、再生成績から学習者のコロケーションと慣用句に関する知識が乏しいことが明らかであった。これは Wray (2002) が提唱する L2 心内辞書モデル (図2) とも合致する。Wray によると、心内辞書は固まりの形で貯蔵されている定式表現、複合語が含まれる語彙及び、形態素という3つの枠組みから構成されている (図2: 3つの四角形)。母語話者は豊富なインプットを受け、多語単位表現を固まりとして心内辞書に直接貯蔵する (図2: ①の破線矢印、矢印の太さは処理傾向の強さを示す)。一方、学習者は教室でのインプットに依存し、多語単位表現に遭遇した時に、常に単語や形態素を分析して処理する傾向が強い (図2: ②の実線矢印)。そのため、学習者のL2心内辞書には定式表現が少なく、語彙や形態素のようなより小さい単位の表現が貯蔵されている。本研究の結果から、学習者は上級になっても本研究で扱った全ての定式表現を固まりとして処理できていたとは言えず、心内辞書にある定式表現の数が限定的であり、Wray のモデルを支持する可能性が高いと思われる。

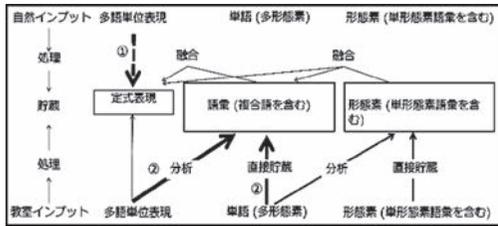


図2 L2心内辞書モデル (Wray, 2002 より引用, 一部改変)

その一方で、本研究では学習者にとって再生成績の高い定式表現のほうが自由結合より産出が速かった。このことから、学習者にも定式表現の獲得は処理速度を上げ、適切な言語運用を支えることができると考えられる。しかし、金・鄧 (2014) は中国の教科書では慣用句が重要視されていないことを、李 (2014) は「中国の教科書は新出単語を母語話者がよく使うコロケーションの形で提示していない」ことを指摘している。よって、L2教育では初期段階から定式表現を積極的に導入することが重要だと考える。どの表現を導入するかに関しては、学習者の心内辞書について研究を積み重ねる必要があるが、本研究から、特に慣用句は固まりとして初期から導入し、積極的に使わせる必要があると考える。また、母語話者コーパスで高頻度かつ複数のジャンルで広く使用される定式表現を優先的に導入することで、自然なインプットに出現する可能性の高い定式表現を指導でき、学習者の言語処理を促すことにつながると言えよう。

7. まとめと今後の課題

本研究の結果から、学習者はコロケーションと慣用句に関する知識が乏しいため、L2知識によって語と語を組み合わせる傾向がある。しかし、高い成績を獲得した定式表現の課題遂行時間の比較では、母語話者、学習者を問わず、定式表現は自由結合より容易に産出され、処理効率が高いことがわかった。これらの結果から、日本語においても、母語話者の心内辞書には固まりで貯蔵されている定式表現が存在する可能性がある。学習者については Wray (2002) の L2心内辞書モデルが支持され、上級者であっても心内辞書に定式表現に関する枠が小さいため、処理効率が母語話者に劣っていることが示された。以上の結果を基に学習者に単語より大きな単位の言語情報を固まりとして指導し、教科書にも積極的に取り入れる必要があると述べた。

今後の課題として、まず本研究では、母語話者の

再生成績に天井効果が出た。今後この点を配慮し、母語話者の作動記憶に処理負荷をかける材料を用いた実験をする必要がある。また、本研究で用いた口頭再生課題は定式表現を産出する実験法である。今後、理解のみを求める文法性判断課題を用いた実験を行うことで、定式表現の処理過程の全体像を明らかにしていくことができる。最後に、本研究ではコロケーションと慣用句に焦点を当てたが、ほかの種類の定式表現についても検討が必要である。これにより、日本語学習者の L2心内辞書及び定式表現処理過程の全体像を把握することができると思われる。

【注】

- 1) 目標言語 (Target Language) とは学習目標となる第二言語を指す。
- 2) 『現代日本語書き言葉均衡コーパス』 (Balanced Corpus of Contemporary Written Japanese) は国立国語研究所が中心となって開発され、書籍、雑誌、新聞、yahoo 知恵袋、ブログ、法律、国会議事録といった様々な書き言葉のジャンルから無作為にデータを抽出し、編集されたコーパスである。本研究では国立国語研究所と Lago 言語研究所が開発した NINJAL-LWP for BCCWJ に掲載されている語彙表から上位20位以内の動詞と形容詞を選択した。

【参考文献】

- 秋元美晴・有賀千佳子 (1996). 『ペアで覚えるいろいろなことば: 初・中級学習者のための連語の整理』 武蔵野書院.
- 大曾美恵子・滝沢直宏 (2003). 「コーパスによる日本語教育研究—コロケーション及びその誤用を中心に—」 『日本語学』 22 (4), 234-244.
- 柏崎雅世・藤村知子・鈴木智美 (2012). 『研究社日本語コロケーション辞典』 姫野昌子 (監修), 研究社.
- 川村よし子・北村達也 (1998). 『日本語読解学習支援システム「リーディングチュウ太」』. <<http://language.tiu.ac.jp/index.html>> 2017年8月アクセス.
- 金華・鄧娟娟 (2014). 「中国の大学の日本語教育における慣用句の扱い」 『日本語教育研究』 60, 97-113.
- 国広哲弥 (1985). 「慣用句論」 『日本語学』 4 (1), 4-14.
- 国広哲弥 (2007). 「連語」 飛田良文主幹 『日本語学研究事典』 (171-172), 明治書院.
- 国立国語研究所・Lago 言語研究所 (2012). 『NINJAL-LWP for BCCWJ』. <nlb.ninjal.ac.jp>

- 2017年8月アクセス .
- 呉琳(2016). 「BCCWJ を用いた基幹慣用句の選定」『北海道大学大学院文学研究科研究論集』16, 99-113.
- 佐治圭三 (1992). 『外国人が間違えやすい日本語の表現の研究』 ひつじ書房 .
- 趙立翠・安永大地・小島治幸・林洪 (2016). 「中国人日本語学習者のコロケーション処理の特徴に関する探索的な実験研究 - 日本語母語話者と対照しながら -」『第27回第二言語習得研究会全国大会予稿集』41-46.
- 東京工業大学留学生センター仁科研究室 (2012). 『日本語作文作成支援システム「なつめ」」 .
<<https://hinoki-project.org/asunaro>> 2017年8月アクセス .
- 畑佐由紀子 (2012). 「語彙と習得 - 総論 -」畑佐一味・畑佐由紀子・百濟正和・清水崇文 (編著) 『第二言語習得研究と言語教育』(3-22), くろしお出版 .
- 李更春・賈冠桀 (2013). 「英語程式語心理表征模式研究」『西藏大学学报 - 社会科学版 -』28 (2), 174-180.
- 李文平 (2012). 「コロケーション処理の影響要因に関する研究」『電子情報通信学会技術研究報告 - 信学技報』112 (339), 65-70.
- 李文平 (2014). 「日本語教科書におけるコロケーションの取り扱いに関する一考察 - 中国の日本語教科書と現代日本語書き言葉均衡コーパスとの比較 -」『日本語教育』157, 63-77.
- 梁志銳・松野和子・杉浦正利 (2008). 「コロケーションの処理過程における脳内活性部位と心的負荷 - NIRS による脳機能測定法に基づく事例研究 -」『英語学習者のコロケーション知識に関する基礎的研究』平成17年度 - 19年度科学研究費補助 (基盤研究 (B)) 研究成果報告書 .
- 甄凤超 (2009). 「语块与外语口语流利度, 准确性及恰当性的相关研究 - 基于 COLSEC 语料库的实证研究 -」『中国外语教育』4 (2), 14-24.
- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. Cowie (Ed.), *Phraseology theory, analysis and application* (pp. 173-194). Oxford: Oxford University Press.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3), 275-311.
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61.
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91, 433-445.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication*. (pp.191-227). London: Longman.
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp.127-150). Amsterdam, the Netherlands: John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Swinney, D. A. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523-534.
- Tremblay, A., Derwing, B., & Libben, G. (2007). Are lexical bundles stored and processed as single units? *Working Papers of the Linguistics Circle of the University of Victoria*, 19, 258-279.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- Wray, A. (2009). Future directions in formulaic language research. *Journal of Foreign Languages*, 32, 2-17.