

科学研究費助成事業 研究成果報告書

平成 26 年 5 月 17 日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23720282

研究課題名(和文) 英語の高頻度 POS グラムおよびフレーズフレーム表現における韻律パターン認識の分析

研究課題名(英文) Prosodic pattern recognition of high-frequency POS-grams and Phrase-frames

研究代表者

村尾 玲美 (MURAO, Remi)

名古屋大学・国際言語文化研究科・准教授

研究者番号：80454122

交付決定額(研究期間全体)：(直接経費) 1,500,000 円、(間接経費) 450,000 円

研究成果の概要(和文)：本研究では、日本人英語学習者が句を形成する品詞連鎖を韻律的なパターンとして認知することができるかどうかを調査した。品詞連鎖を高頻度と低頻度の条件に分けて韻律パターンマッチング課題を行ったところ、上級と初級の英語学習者の間には品詞連鎖条件に関わらず正解率に差が見られなかった。英語母語話者との間には正解率に有意な差が見られ、品詞連鎖条件による違いも見られた。以上の結果から、品詞連鎖を韻律パターンとして認知できるかどうかは TOEIC で測られるリスニング力とリーディング力に影響する要素ではないこと、言語習得において重要な要素とは言えないこと、母語話者との処理の仕方に違いがあること、が示唆された。

研究成果の概要(英文)：This study examined whether Japanese learners of English could recognize the prosodic pattern of a string of part-of-speech (POS-gram) which constitute a phrasal category. The results of prosodic pattern-matching task showed that there was no difference in the performance between advanced and lower level learners in recognizing the prosodic pattern of high and low frequency POS-grams. Native speakers of English (NS) outperformed English learners (NNS) especially on the phrasal categories which appeared frequently in the NS corpus but not in the NNS corpus. Interestingly, non-phrasal categories were recognized at the same level between NS and NNS. The results suggest that although NS and NNS differ in their performance of recognizing prosodic pattern of a phrasal category, it is not a crucial factor that affects listening and reading comprehension skills which divides advanced level learners from lower level learners by their TOEIC scores.

研究分野：人文学

科研費の分科・細目：言語学・外国語教育

キーワード：英語教育 第二言語習得 音声言語認識 プロソディ コーパス

1. 研究開始当初の背景

英語に限らず、あらゆる言語の母語話者は第一言語を獲得するにあたって、連続した音声から韻律的にまとまったチャンクを抽出し、韻律表象の形成を行いながら言葉の意味を獲得していくことが指摘されている (Peters, 1985)。従って、母語話者の心的辞書には高頻度表現の韻律表象が数多く蓄えられていると考えられる。村尾(2006)の実験では、英語母語話者のみならず、日本人の上級英語学習者も、高頻度表現(例: Do you know what I mean?)の音素情報を劣化させた音声を聞かせた場合、メロディやリズムの手がかりを用いて表現を認識できることが証明された。一方、英語音声の聞き取りに不慣れな学習者は、韻律情報を利用して表現を検索することができなかつたため、心的辞書における韻律表象の形成が未発達であることが指摘された。聞き取りに韻律情報を利用できない学習者は、個々の音素の聞き取りに頼ってしまうため、少しでも雑音が入ったり音声が劣化したりすると認識が困難になり、認知的な負荷がかかる結果となる。

母語話者の心的辞書には、固定された高頻度表現 (fixed expression) の韻律表象のみならず、他の単語と置き換え可能なスロットを持つ、半固定的な高頻度表現も存在すると考えられる (Wray, 2002)。Nattinger & DeCarrico (1992) の用語を借りるならば “phrasal constraints” や “sentence builders” に当てはまる表現であり、コーパスの用語ではフレーズフレームと呼ばれている。例えば “a__ago” という表現では、スロット部分は day, week, year に置き換えが可能である。このような半固定的な高頻度表現は反復利用されるため、言語使用において経済的であり、言語処理においてもパターン化された部分に注目する必要がなくなるため、聞き取りの負荷を軽減すると考えられる (Schmitt, 2010)。しかしながら先行研究では、半固定的な高頻度表現の韻律表象が言語使用者の心的辞書に存在するか否かについて明らかになっていない。韻律表象が形成されているのであれば、これらの表現を韻律パターンで認識することができるため、聞き手はスロット部分の聞き取りにのみ注意の配分を行い、負荷の低い聞き取りが可能になる。

言語使用者の心的辞書内に、どのような韻律表象が形成されているかについて、固定的および半固定的な高頻度表現の他に、「高頻度の品詞の連なり」(POS グラム)も抽象的な韻律パターンとして存在している可能性が考えられる。BNC を検索すると、最も高頻度に出現する 5 品詞の連なりは、「冠詞 + 名詞 + OF + 冠詞 + 名詞」であり、具体的には “the end of the year” や “the rest of the world” といった表現であった。高頻度の品詞の連なりは、文中でまとまった句の働きをしている場合が多く、一定のリズムとメロディを持っていると考えられる。高頻度 POS グ

ラムを持つ一定のリズムとメロディが心的実在性を持つのであれば、聞き取りの際に、韻律情報を利用したパターン認識が可能となる。

2. 研究の目的

本研究では、固定的な個別の高頻度表現だけではなく、韻律的なパターンを形成すると考えられる半固定的な表現および品詞の連なりという表現の「型」を、新たなテスト項目として導入し、韻律パターン認識能力のデータを収集する。収集したデータの分析により、韻律表象の形成とリスニング能力との関連を明らかにする。

3. 研究の方法

3.1 実験項目

名古屋大学の杉浦正利氏によって編纂された英語母語話者コーパス (NICE 2.0: NS) と日本人英語学習者コーパス (NICE 2.0: NNS) を使用し、実験項目の抽出を行った (NICE の最新版は 2013 年 11 月 25 日公開の NICE2.2)。NICE 2.0 における英語母語話者コーパスの総語数は 117,296 語で、日本人英語学習者コーパスの総語数は 114,550 語である。5 単語の連なりから成るフレーズフレームを収集するにあたり、まず 5 単語連鎖表現 (5-gram) を抽出し、5 単語のうち 4 単語が同じで 1 単語が可変 (スロット) となる表現を抽出した。その結果、5 単語のうち 3 語目がスロットとなる頻度 2 以上のフレーズフレームは 36 表現抽出されたが、36 表現のうち頻度が 3 以上あったのは 7 表現のみであり、“I think * is very” という表現ひとつのみが最大頻度の 4 であった。BNC でフレーズフレームを検索すると、高頻度表現は “in the * of the” や “at the * of the” のようにスロット部分に内容語が入るものが多く、使い回しのきく表現が検索されるが、114,550 語規模の学習者コーパスでは頻度の高いフレーズフレームを抽出するには小さすぎ、スロット部分に内容語が入る高頻度表現は抽出できなかった。学習者コーパスの規模の限界という問題から、本研究ではフレーズフレームを実験項目として扱うことを見送ることとした。

POS グラム (品詞連鎖) であれば、フレーズフレームよりも抽象化の度合いが高いため、ある程度の頻度が確保されると考えられる。そこで、TreeTagger によって品詞タグ付けされた英語母語話者コーパス (NICE 2.0: NS_parsed) と日本人英語学習者コーパス (NICE 2.0: NNS_parsed) から 5 品詞の連鎖を抽出し、NNS・NS とともに高頻度の品詞連鎖、NS のみ高頻度の品詞連鎖、NNS のみ高頻度の品詞連鎖、NNS・NS とともに低頻度の品詞連鎖の 4 条件に分類した (表 1)。品詞連鎖は句読点を含まない、連続した連鎖のみを抽出した。

表1. 条件別の品詞連鎖頻度 (NS/NNS)

NS・NNS 高頻度	冠+名+前+冠+名(263/145) 名+前+冠+形+名 (185/91) 前+冠+名+前+名 (144/95) 冠+形+名+前+名 (140/50) 形+名+前+冠+名 (120/81) 冠+名+前+形+名 (88/71)
NS 高頻度 NNS 低頻度	to 不+動+代名+形+名(24/6) to 不+冠+名+前+名 (22/1) 前+動 ing+冠+形+名 (28/3) 冠+形+接+形+名 (22/3) 前+冠+名+to 不+動 (48/5)
NS 低頻度 NNS 高頻度	代名+助+動+前+動 (0/76) 代名+be+形+前+動 (0/43) 前+動+前+冠+名 (0/44) 代名+助+動+冠+名(25/102) 形+名+接+形+名 (9/44) 形+名+前+形+名 (19/51)
NS・NNS 低頻度	動+副+動+冠+形 (4/4) 前+代名+名+前+ing (4/4) 名+前+代名+名+前 (4/4) 前+冠+名+形+前 (4/4) 冠+名+接+代名+動 (4/4) 副+動+副+前+冠 (4/4)

なお、ここで注意したいのは、NICE2.0:NNS は合計 342 の作文が 241 名の日本人英語学習者によって書かれているのに対し、NICE2.0:NS は合計 200 の作文が 28 名の英語母語話者によって書かれている点である。母語話者コーパスにおいては特に、書き手の癖がデータに影響を及ぼしている可能性があり、それぞれのコーパスに出現する品詞連鎖の頻度を基に英語母語話者と日本人英語学習者の特徴として直接比較するには問題があると考えられる。このような問題を含んでいることを考慮した上で、結果の解釈をする必要がある。

からは句を形成する品詞連鎖を 6 種類ずつ選択し、からは句を形成しない品詞連鎖を 6 種類選択した。各条件 6 種類の品詞連鎖に当てはまる具体的な単語連鎖をコーパスから 2 種類ずつ抽出し、合計 48 の実験項目を用意した。具体的な単語連鎖の抽出にあたっては、短音節語のみからなる単語連鎖と多音節語を含む単語連鎖の数が各条件で同じになるように考慮した。次節で説明する、

韻律パターンマッチング課題の選択肢を作成するにあたり、いくつかの項目を削除したため、条件間の項目数が変わり、最終的には 6 種類の品詞連鎖と 11 個の単語連鎖、は 5 種類の品詞連鎖と 9 個の単語連鎖、とは 6 種類の品詞連鎖と 10 個の単語連鎖の合計 40 項目となった。実験項目の音声ファイルを作成する際には、自然なプロソディで発音を録音するため、40 個の実験項目をそれぞれ文中に埋め込み、アメリカ人英語母語話者に一文ずつ音読させた。その後音声編集ソフトでターゲットの単語連鎖のみを切り取り、実験項目の音声を作成した。

3.2 実験参加者と実験方法

実験参加者は英語母語話者 8 名、上級英語学習者 10 名 (TOEIC 平均 933 点)、初級英語学習者 14 名 (TOEIC 平均 418 点) である。実験項目の音声にはローパスフィルタをかけ (cut-off point 400Hz)、音素情報を劣化させることで、表現のプロソディのみが手掛かりとなるようにした。実験では音素情報を劣化させた刺激 (例: as a way of life) を音声提示した後、2 種類の 5 単語連鎖を選択肢として視覚提示し、刺激と同じ品詞連鎖から成る表現 (例: in a court of law) か、異なる品詞連鎖から成る表現 (例: a great deal of time) のいずれに近い音声だったかを判断させた。この韻律パターンマッチング課題を行うには、実験参加者は視覚呈示された二つの表現を頭の中で音声として再現し、その再現された 2 表現のプロソディのどちらが、刺激提示として聞こえてくる表現のプロソディとよりよくマッチするかを判断する必要がある。品詞連鎖の持つ内容語と機能語の強弱リズムを、学習者が頭の中で正しく再現できていれば、マッチングは可能となる。

正解の選択肢は、刺激と同じ品詞連鎖に当てはまる単語連鎖を NICE コーパスから抽出することで作成した。不正解の選択肢を作成するためには、まず刺激の品詞連鎖の強弱リズムを内容語と機能語の観点から分析した。次に、内容語と機能語が一部反転するような品詞連鎖を考え、その品詞連鎖に当てはまる具体的な単語連鎖をコーパスから抽出した。正解と不正解の選択肢は音節数を揃え、多音節語を含む場合、単語のアクセント位置も揃えるようにした。また、発音がわからないような難しい単語を含まないように配慮した。

正解の選択肢に刺激とは異なる単語連鎖を用いたのには二つ理由がある。一つは、刺激にローパスフィルタをかけても完全には音素情報を消し去ることはできないからである。聞こえ度の高い音素を手掛かりとしてマッチングに成功するという可能性を減らすため、本研究では品詞連鎖は同じまま、異なる単語連鎖を用いた。二つ目の理由は、本研究の目的は品詞連鎖の韻律パターン処理能力を調査することであるため、品詞連鎖に当てはまる具体的な単語の並びは多様であ

る必要があるからである。また、本研究では問題は二者択一とし、「わからない」という選択肢は設けなかった。これは刺激と正解の単語連鎖が異なり、頭の中で再現した音声と刺激が完全に合致することはないため、「どちらでもない」という意味で「わからない」を選択する学習者が増える可能性を排除するためである。本研究では回答への確信度は調査していないが、チャンスレベルの回答であるかどうかを見極める方法の一つとして導入することも考えられる。

よく使用する品詞連鎖の韻律をパターンとして処理しているのであれば、英語母語話者はとの品詞連鎖に比べ、との品詞連鎖を正しく認識するのに対し、学習者はとに比べ、とを正しく認識するという仮説が成り立つ。

3.3 分析方法

実験では正答率と音声再生回数と回答までの時間を測定したが、本稿では正解率の結果のみを報告する。分析は正解率を従属変数とし、実験グループ(母語話者・上級者・初級者)と四つの品詞連鎖条件を独立変数とする、二要因分散分析混合計画を行った。

4. 研究成果

4.1 結果

図1は実験グループによる条件ごとの正解率を示した図である。二要因分散分析の結果、実験グループと条件の交互作用が有意であった($F_{(6,87)} = 2.602, p < .05, \eta_p^2 = .152$)。単純主効果の検定を行ったところ、NNS・NSともに低頻度の品詞連鎖については実験グループ間に有意な差が見られなかったのに対し、その他の条件については母語話者と学習者の間に有意な差が見られた。以下に実験グループの単純主効果の結果をまとめる。

- : 母語話者 > 上級者 = 初級者
- : 母語話者 > 上級者 = 初級者
- : 母語話者 = 上級者 = 初級者
- 母語話者 > 初級者
- : 母語話者 = 上級者 = 初級者

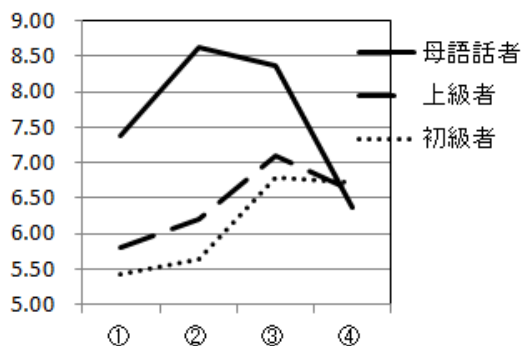


図1. 実験グループによる条件ごとの平均

各実験グループにおける条件間の差については、母語話者はとがに比べ有意に高い点であったのに対し、学習者はがに比べ有意に高い点であった。初級学習者は仮説に反してよりもの方が有意に点が高いという結果となった。以下に条件の単純主効果の結果をまとめる。

- 母語話者 : >
- 上級者 : >
- 初級者 : >

4.2 考察

4.2.1 仮説を支持する結果

実験の結果、(NSのみ高頻度の品詞連鎖)は母語話者による認識は高く、学習者は低いという仮説を支持した。また、学習者は(NNSのみ高頻度の品詞連鎖)の認識が高いという結果も仮説通りであった。の認識に関しては母語話者も高いという結果を得た。これは、の表現が、本研究で用いた母語話者のエッセイライティングのコーパスにおいては多く出現しなかっただけであり、他のジャンルの書き言葉や話し言葉では頻出する品詞連鎖であったためと考えられる。表1からも分かる通り、には代名詞を含む表現が多く含まれていることから、学習者がエッセイライティングに一人称を多用していることが見て取れる。これらの結果から、母語話者も学習者も、よく使う品詞連鎖は韻律的なパターンとして身に付けている可能性が示唆された。

4.2.2 仮説を支持しない結果

学習者は仮説に反し、(NNS・NSともに低頻度の品詞連鎖)の認識が他と比べ高かった。Thorp & Trehub (1989)によると、言語獲得過程にある6ヶ月から9ヶ月の乳児は、音のまとまりよりもピッチやポーズによって境界が変わる部分に敏感に反応する。本研究で用いたの項目は句としてのまとまりを為さない品詞連鎖であるため(例: the paper, and they look)ピッチやポーズによる音韻変化がより顕著であり、マッチングの手がかりが多かったため、他と比べると認識しやすかったと考えられる。興味深いのは、の正解率が母語話者と学習者で同じである点である。

をコントロールと考えるなら、母語話者はよりも句を成す品詞連鎖を正しく認識したのに対し、学習者(初級者)は、品詞連鎖というまとまりとして韻律パターンを認識するよりも、音韻変化が顕著な部分に反応して認識していると考えられる。

また、学習者は仮説に反し、(NNS・NSともに高頻度の品詞連鎖)が有意に低く、母語話者も予測より低い結果となった。の項目を見ると、高頻度であるがゆえに品詞のバリエーションが少なく、冠詞、名詞、形容詞、

前置詞の四種類しか含まれていないことが分かる。単語認知の研究では、高頻度語は音韻的に競合する候補語である“Neighborhood density”が高く、認知が難しくなると言われている(Garlock, Walley & Metsala, 2001)。品詞連鎖も同様に、高頻度であれば音韻的に競合する表現が増え、その分認識が困難になることが予想される。一つの解釈としては、高頻度品詞連鎖の韻律パターンは習得はされているものの、類似したパターンが多いために、今回のような韻律パターンマッチング課題で選択するのは困難であったと考えられる。二つ目の解釈としては、高頻度品詞連鎖の韻律パターンは習得されていないという考えである。言語習得における頻度効果については Ellis (2002)が述べているが、Peters (1985)が述べているように、こどもは言語獲得において韻律的に顕著なまとまり(prosodically salient chunks)に注目して心的辞書に保持していく。言語の連鎖が暗黙的に心的辞書に保持されるには高頻度であるだけでなく、音韻的に顕著である必要があり、この項目はその条件に満たなかった可能性も考えられる。

4.3 全体的傾向

学習者が他の条件に比べ、をを高く認識したのは仮説通りであったが、正解率は全体的に低かった。TOEIC900点を超える上級者でも高頻度品詞連鎖が持つプロソディを区別できなかったことから、品詞連鎖を韻律的なパターンとして捉える能力は言語習得において重要な要素ではないと考えられる。また、上級者と初級者に差が見られなかったことから、この能力は TOEIC が測定するリーディング力やリスニング力に直接影響する能力ではないと考えられる。しかしながらまた、母語話者とは点数の差が見られ、交互作用も見られたため、TOEIC で測られていない能力となんらかの関係がある可能性は否めない。

4.4 結論

本研究では、日本人英語学習者が句を形成する品詞連鎖を韻律的なパターンとして処理することができるかどうかを調査した。品詞連鎖を高頻度と低頻度の条件に分けて韻律パターンマッチング課題を行ったところ、上級英語学習者と初級英語学習者の間には品詞連鎖条件に関わらず正解率に差が見られなかった。一方、英語母語話者との間には正解率に有意な差が見られたのみならず、品詞連鎖条件による違いが見られた。以上の結果から、を品詞連鎖を韻律パターンとして処理できるかどうかは言語習得において重要な要素ではないこと、TOEIC で測られるリスニング力とリーディング力に影響する要素ではないこと、母語話者との処理の仕方に違いがあること、が明らかとなった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計1件)

村尾玲美(2013). 「英語高頻度品詞連鎖における韻律パターン認識」『言語文化論集第XXXV巻(第2号)』pp. 99-108.

〔学会発表〕(計1件)

村尾玲美(2013年8月13日). 英語高頻度品詞連鎖における韻律パターン認識. 全国英語教育学会北海道大会.

6. 研究組織

(1)研究代表者

村尾玲美 (名古屋大学・准教授)

研究者番号：80454122