

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 14 日現在

機関番号：62618

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24520523

研究課題名(和文) 自発的な話し言葉に見られる節連鎖構造の研究

研究課題名(英文) A corpus-based analysis of multiple clause linkage structure in Japanese

研究代表者

丸山 岳彦 (MARUYAMA, Takehiko)

大学共同利用機関法人人間文化研究機構国立国語研究所・大学共同利用機関等の部局等・准教授

研究者番号：90392539

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：本研究の目的は、自発的な話し言葉に現れる「多重的な節連鎖構造」の実態を、自発音声コーパスに基づいて明らかにすることである。自発的な話し言葉では、連用節が何重にも連なることによって、長大な発話が形成されることがある。これは日常の話し言葉で頻繁に観察される現象であるが、そこでどのような形式の連用節がどのような組み合わせによって用いられているのかは、明らかになっていない。その実態を、『日本語話し言葉コーパス』(CSJ)を用いて定量的に分析する。また、『現代日本語書き言葉均衡コーパス』(BCCWJ)を用いて同様の分析を実施することにより、話し言葉と書き言葉の比較・対照を行う。

研究成果の概要(英文)：This study will quantitatively and qualitatively describe the multiple clause linkage structure in contemporary spoken and written Japanese, using two large-scale Japanese corpora, Corpus of Spontaneous Japanese (CSJ) and Balanced Corpus of Contemporary Written Japanese (BCCWJ). The following two research questions are posed: 1. What types of clauses are connected in what order to form the multiple clause linkage structure? 2. What factors bear on the variable use of the multiple clause linkage structure?

研究分野：日本語学

キーワード：コーパス 話し言葉 節連鎖構造

## 1. 研究開始当初の背景

2000年代に入って以降、国立国語研究所の主導により日本語コーパスの開発と公開が進められてきた。2004年に『日本語話し言葉コーパス』(CSJ)が、2011年に『現代日本語書き言葉均衡コーパス』(BCCWJ)が公開され、大規模な日本語コーパスに基づく実証的な文法研究が個人レベルで実施できる環境が整った。

この状況を受け、CSJを用いて話し言葉の文法現象を定量的に分析することを計画した。分析項目として、自発的な話し言葉に特徴的に現れる「多重的な節連鎖構造」を取り上げることとした。教育的な観点からは「だらだら文」などと称され、規範的な文法意識からは避けられる傾向にある「多重的な節連鎖構造」が、実際の発話行動の中でどのように分布しているのか、という点を調査・記述することとした。

## 2. 研究の目的

本研究の目的は、現代日本語記述文法の立場から、自発的な話し言葉に現れる「多重的な節連鎖構造」の実態を、大規模な自発音声コーパスに基づいて明らかにすることである。自発的な話し言葉では、連用節が何重にも連なることによって、長大な発話が形成されることがある。これは、述語の活用形やさまざまな接続助詞によって、形式的な文末表現を使わずに発話を継続することができる日本語の特徴であるが、自発的な話し言葉の中で、どのような形式の連用節がどのような組み合わせによって用いられているのかは、これまで明らかにされていない。その実態を、『日本語話し言葉コーパス』(CSJ)を用いて定量的に分析する。また、『現代日本語書き言葉均衡コーパス』(BCCWJ)を用いて同様の分析を実施することにより、話し言葉と書き言葉の比較・対照を行う。

## 3. 研究の方法

多重的な節連鎖構造とは、例えば、以下のようなものである。CSJから例を挙げる。

初めての海外旅行が中国だって言ったら  
もう色々な人に渋いとか色々言われた  
んですけど 中国元々興味もあったんで  
楽しみだなど思いながら行ってきたんで  
すが でまず中国にどうして中国に行くことになったかっていうお話をちょっとしたいんですが その当時ですね中国からの留学生が来てまして...

連用修飾節を導く接続助詞「たら」「けど」「んで」「が」「まして」などが再帰的に連鎖することにより、発話がいつまでも終わらず、多重的な節連鎖構造が生じている。このよう

な節連鎖構造の実態を捉えるためには、コーパス中に出現した連用節の終端境界の位置と種類を網羅的にアノテーション(情報付与)する必要がある。

CSJには、すでに「節境界ラベル」がアノテーションされている。このラベルを用いることで、CSJにおける節連鎖構造を抽出することができる。一方、BCCWJには「節境界ラベル」がアノテーションされていないため、新たに付与する必要がある。

CSJ・BCCWJに対してアノテーションされた節境界ラベルの連鎖パターンを分析することにより、現代日本語の話し言葉・書き言葉における節連鎖構造を定量的に記述することができる。

## 4. 研究成果

(1) 研究期間中に発表した研究成果は、以下のとおりである。

雑誌論文 1件  
国内学会発表 6件  
国際学会発表 5件  
図書(編著書・共著) 2件

いずれもCSJおよびBCCWJを用いて節連鎖構造を分析した研究成果、または、そのアノテーション結果を利用した研究成果である。特に、国際学会における研究発表では、日本語コーパスの整備状況とアノテーションの実施状況を紹介し、その結果を用いた分析例を示すことで、自発音声コーパスの国際共同研究プロジェクト(イタリア、フランス、ロシア、ブラジル、イスラエル、日本)に貢献した。

(2) 上述のように、CSJには、すでに「節境界ラベル」がアノテーションされている。このラベルを用いることで、CSJにおける節連鎖構造を抽出することができる。一方、BCCWJには「節境界ラベル」がアノテーションされていないため、新たに付与する必要がある。そこで、以前に開発した節境界検出プログラム「CBAP」(丸山ほか、2004)を用いて、BCCWJへの節境界ラベルのアノテーションを実施した。CSJとBCCWJに節境界ラベルを付与した結果の例を、以下に示す。  
// で囲まれた部分が節境界ラベルである。

そのルクラの村なんです /並列節ガ/ 飛行場は本当に山の中にありまして /テ節/ しかも飛行場が砂利道というところでしたので /理由節ノデ/ 実際に着陸する時は本当に心臓が止まりそうになったんですけども /並列節ケレドモ/ (F え) 着いてみたら /条件節タラ/ 意外にそういった宿泊施設などで (F えー) 賑わった小さな村でした /文末/ (CSJ:S01F0151)

イスラエルからの報道によると /条件節ト / 同国最大の都市テルアビブで十五日夜パレスチナ自治区ガザからのイスラエル軍撤退とユダヤ人入植地撤去を求める /連体節/ 集会が開かれ /中止節/ 十五万人以上が参加した。 /文末/ 最大野党の労働党などの呼びかけによる /連体節-形式名詞/ もので /デ節/ 二千年秋にパレスチナ紛争が激化して以来 /時間節イライ/ 最大の規模となった。 /文末/ (BCCWJ:PN4f 00018)

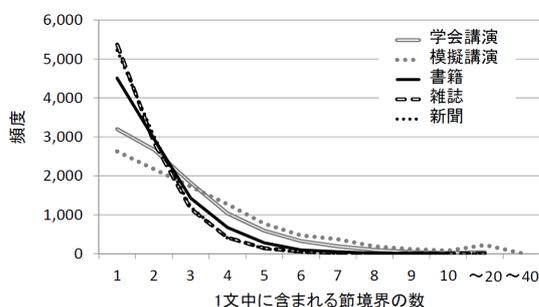
ただし、CSJとBCCWJに付与されている節境界ラベルの体系が異なるため、比較・対照するラベルを限定することにした。分析対象とした節境界ラベルの一覧を、表1に示す。

表1 分析対象とした節境界ラベル

分類	節境界ラベル
文末表現	文末
並列節	ガ節、ケレドモ節、ケレド節、ケドモ節、ケド節、シ節
理由節	カラ節、ノデ節
条件節	タラ(バ)節、ト節、ナラ(バ)節、レバ節
その他	中止節、デ節、テ節、引用節、トイウ節

(3) CSJ・BCCWJそれぞれにアノテーションされた節境界ラベルを集計した結果、話し言葉と書き言葉の間で、節境界の連鎖数に大きな違いがあることが確認された。さらに、話し言葉における改まった発話スタイル(学会講演)とくだけた発話スタイル(模擬講演)の間で、また、書き言葉における書籍・雑誌・新聞というレジスターの違いによって、節境界の分布に差が見られた。集計の結果を図1に示す。

図1 節境界の連鎖数



(4) さらに、節境界ラベルの連鎖パターンを各レジスター間で検討した。通常はテキスト中における語彙の多様性を測る指標として用いられる「TTR(トークン比)」を、節連鎖パターンの多様性を測る指標として利用した。

すなわち、節連鎖パターンの異なりを総文数で割ることにより、TTRを求めた。この値が低いほど、同一の節連鎖パターンが繰り返し現れることを示し、逆に値が高いほど、多様な節連鎖パターンが生じていることを示す。結果を表2に示す。

表2 節が連鎖するパターンの数と多様性

		総文数	パターン数	TTR
CSJ	学会	5,389	1,065	0.198
	模擬	4,494	1,542	0.343
BCCWJ	書籍	8,780	857	0.098
	雑誌	9,342	571	0.061
	新聞	11,898	497	0.042

TTRの値を見ると、多様な節連鎖パターンは、書き言葉よりも話し言葉の方に多く現れていることが分かる。さらに、学会講演よりも模擬講演の方でTTRの値が高いことから、特にくだけた発話スタイルにおいて「だらだら文」調の長い発話が多く生じていると考えられる。一方、書き言葉では、新聞のTTRが最も低い。新聞記事は比較的統制された文体を持っており、節連鎖パターンの多様性が抑制されているためと考えられる。

(5) ここで示した研究成果は、自発的な大規模音声コーパスや、さまざまなレジスターを備えた書き言葉コーパスが利用できるようになって初めて可能になった文法研究の例である。特にBCCWJの完成以降、コーパスを用いた日本語文法研究の数が飛躍的に増えてきたが、コーパスに付与されたアノテーション情報を積極的に利用した研究や、自らコーパスにアノテーションを実施してその結果を集計するという研究の事例は、あまり見られないようである。本研究の成果により、利用者が積極的にアノテーションを実施し、その結果を分析して定量的な分析を行うという、コーパスに基づく文法研究の方向性・可能性の一つを示すことができたものと思われる。

<引用文献>

丸山岳彦・柏岡秀紀・熊野正・田中英輝 (2004) 「日本語節境界検出プログラムCBAPの開発と評価」『自然言語処理』Vol.11, No.3. 39-68. 言語処理学会。

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)

丸山岳彦 (2014) 『日本語話し言葉コーパス』に基づく挿入構造の機能的分析』『日本語文法』14 巻 1 号, 88-104. 日本語文法学会. 査読あり.

〔学会発表〕(計 11 件)

丸山岳彦・佐藤理史・夏目和子 (2016) 「現代日本語における節の分類体系について」『言語処理学会第 22 回年次大会予稿集』1113-1116. 言語処理学会. 2016 年 3 月 10 日. 東北大学(宮城県・仙台市) Maruyama, Takehiko (2015)

“Two-Level Utterance Units: Cognitive and Communicative Aspects of Spontaneous Speech”. IX LABLITA and IV LEEL International Workshop: Units of Reference for Spontaneous Speech Analysis and their correlations across languages. August, 6, 2015. Belo Horizonte (Brazil)

丸山岳彦 (2013) 「現代日本語の従属節に現れるモダリティ形式の分布」『第 4 回コーパス日本語学ワークショップ 予稿集』299-308. 国立国語研究所. 2013 年 9 月 6 日. 国立国語研究所(東京都・立川市)

Maruyama, Takehiko (2013) “Analysis of Parenthetical Clauses in Spontaneous Japanese”. Proceedings of DiSS 2013 The 6th Workshop on Disfluency in Spontaneous Speech, 45-48. August, 23. 2013. Stockholm (Sweden)

〔図書〕(計 2 件)

丸山岳彦 (2014) 「現代日本語の連用節とモダリティ形式の分布 — BCCWJ に基づく分析」益岡隆志・大島資生・橋本修・堀江薫・前田直子・丸山岳彦編 『日本語複文構文の研究』399-425. ひつじ書房.

丸山岳彦 (2014) 「現代日本語の多重的な節連鎖構造について — CSJ と BCCWJ を用いた分析」石黒圭・橋本行洋編 『話し言葉と書き言葉の接点』93-114. ひつじ書房.

〔産業財産権〕

出願状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
出願年月日：  
国内外の別：

取得状況(計 件)

名称：  
発明者：  
権利者：  
種類：  
番号：  
取得年月日：  
国内外の別：

〔その他〕  
ホームページ等

6. 研究組織

(1) 研究代表者

丸山 岳彦 (Maruyama, Takehiko)  
国立国語研究所・言語資源研究系・准教授  
研究者番号：90392539

(2) 研究分担者

( )

研究者番号：

(3) 連携研究者

( )

研究者番号：