

日本語の自発音声に見られる 節連鎖構造の分析

国立国語研究所 言語資源研究系

丸山 岳彦

2011年12月17日

「複文構文の意味の研究」ワークショップ

1

発表の流れ

- 問題の所在
- 本発表の目的
- 『日本語話し言葉コーパス』
- 節境界ラベル
 - 連用節の種類と出現数
 - 「文末」が現れるまでに含まれる語数、節数
 - 節境界ラベルの連鎖パターン
- まとめ

2

背景

- 文法研究の方法論
 - a. 母語話者の内省、文法性判断に基づく研究
 - b. 言語資料に基づく研究
- 文法研究の対象
 - 現代日本語（共通語）
 - 現代日本語（方言）
 - 古典
 - 話し言葉

3

問題の所在

- 日本語の話し言葉に現れる節連鎖構造
(Clause-linkage structure)
 - 連用節
 - 連体節
- 自発的な話し言葉では、連用節（副詞節、並列節）が何重にも連鎖することによって、発話が長くなることが多い
 - どのような種類の節が、どのように連鎖しているのか？

4

現象（1）



(F えっと) 今日 お話しさせていただく 内容なんですけれども (F えーっとー) (F ま) 特に珍しい ことではない と思うのですが (F あのー) 自分は (F あの) 体は (F えー) 元から 強い 方ではなかったのですが (F あの) いわゆる 病気が 多い 病気という ことは した ことが なくて (F その)(F え) (D い) 一か月程 ずっと 寝たきり と言うか 家におりまして (F あの) 病気を しておりますもので 自分にとって は 生まれて 初めての ことだったので (F えー) その こと について お話しさせていただきたい と思います

5

現象（2）



で 年が 明けまして 今年に なりましてから (F え) 徐々に 良くなって きて そのまま 病院には 行っていない かったんですけれども その (F ま) 家で 何日か 寝ておりました で (F あのー) 仕事の方に 復帰いたしましたら (F その) 一日 復帰いたしましたら その 日の 夕方から また 体調が 良く 悪く なりまして (F あの)(F えー)(F まー) 寒気ですとか (F えー) 発熱が 酷くて (F ま) また ぶり返したのか と思って (F あのー) 家に 戻って 寝ていたんですが (F あの) 熱が もう 長く 久しぶりに 四十度 ぐらい 出まして (F あのー) (D つ) 家に あった 強い 解熱剤 を 飲んで いたんですが 全く 効かずに (F え) 土日 だったもので 病院に 行けませんでした その 強い (D くす) お薬を 飲んでも (F あの) もう 生まれて 初めての こと なんですが 熱で もう 眠れずに (F あの) うわ言とかを もう 言って しまっていて もう もう 暑くて 言うか 寒気は する んですが 暑くて 寝ても いられない という 感じで

6

久野(1978)『談話の文法』

従来の言語学研究(中略)は、厳密に、科学的に定義できる構文法規則の研究が偏重され、(中略)公式化のしにくい概念に基づいた規則の研究は、「談話部門」というごみ箱に入れられて、無視される傾向が強かった。併し、構文法的規則で、すっきりと割り切れる言語事象は極めて稀で、多くの場合、構文法規則と、種々の談話法規則が、相関し合っており、極めて複雑な言語事象を作り出している。(中略)構文法規則を立てる前に、まず、その言語事象の総体を深く研究し、その事象と関連してくる談話法的要因を隔離した上で、残りの部分に構文法的規則づけを行なうことが必要である。

久野(1978)『談話の文法』 p.307

7

本発表の目的

- 日本語の自発音声における従属節の出現と、その連鎖の実態について、大規模な自発音声コーパスを用いて分析する。
 - 連用節の種類、生起頻度、生起順序
 - 話し言葉の自発性と、節連鎖構造の出現傾向

8

コーパス

- 『日本語話し言葉コーパス (CSJ)』
 - 現代日本語の自発音声を取録
 - 約661時間、772万語分の音声データ
 - 合計3,302講演
 - 1,418名の話者(男性947名、女性471名)
 - 主として2種類の独話で構成
 - 学会講演：学術学会での口頭発表(フォーマルなスタイル)
 - 模擬講演：日常的なトピックに関するスピーチ(カジュアルなスタイル)

9

CSJのアノテーション情報

- 音声データに、多くの研究用情報を付与。
 - 転記テキスト：発話・音声イベントの忠実な書き起こし。
 - 形態論情報：発話を「語」に分割し、その種別を記述した情報。
 - 節境界ラベル：従属節の終端境界に、ラベルを付与した情報。
 - 印象評定データ：その講演が聞き手に与える印象に関する情報。

10

節境界ラベル

- 従属節の終端境界に、ラベルを自動付与。
- ラベルは49種類。
- [絶対境界]、/強境界/、<弱境界>

絶対境界：文末、文末候補、と文末
強境界：並列節カ、並列節ケレドモ、並列節ケレド、並列節ケドモ、並列節ケド、並列節>、ヨウニ節
弱境界：条件節タラ、条件節タラバ、条件節ト、条件節ナラ、条件節ナラバ、条件節レバ、理由節カラ、理由節カラニハ、理由節カラ-助詞、理由節ノデ、タリ節、タリ節-助詞、テ節、テハ節、テモ節、テカラ節、テカラ節-助詞、テ節-助詞、トカ節、トカ節-助詞、ノニ節、連用節、引用節、引用節-助詞、引用節トノ、トイウ節、間接疑問節、間接疑問節-助詞、連体節テノ、並列節タノ、並列節デ、並列節ナリ、フィラー文、感動詞、接続詞、接続詞C、接続詞L、接続詞CL、接続詞M

11

節境界ラベルの例(1)

1. (F-えっと)今日お話しさせていただく内容なんですけれども/並列節ケレドモ/
2. (F-えっとー)(F-ま)特に珍しいことではないと<引用節>
3. 思うのですが/並列節カ/
4. (F-あのー)自分は(F-あの)体は(F-えー)元から強い方ではなかったのですが/並列節カ/
5. (F-あの)いわゆる病気がらしい病気ということはしたことがなくて/テ節/
6. て(F-その)(F-え)(D-い)一か月程ずっと寝たきりと言うか<文末候補>
7. 家におりまして/テ節/
8. (F-あの)病気をしておりましたもので<並列節デ>
9. 自分にとっては<テハ節>
10. 生まれて<テ節>
11. 初めてのことであったので<理由節ノデ>
12. (F-えー)そのことについて<テ節>
13. お話しさせていただきたいと<引用節>
14. 思います[文末]

12

節境界ラベルの例（2）

1. で年が明けまして/テ節/
2. 今年になりましてから<テカラ節>
3. (F-え)徐々に良くなってきて<テ節>
4. そのまま病院には行っていなかったんですけども/並列節クレドモ/
5. その(F-ま)家で何日か寝ておりました/テ節/
6. で(F-あの一)仕事の方に復帰いたしましたら<条件節タラ>
7. (F-その)一日復帰いたしましたら<条件節タラ>
8. その日の夕方からまた体調が凄く悪くなりまして/テ節/
9. (F-あの)(F-えー)(F-まー)寒気ですとか<トカ節>
10. (F-えー)発熱が酷くて<テ節>
11. (F-ま)また振り返したのかなと<引用節>
12. 思っ<テ節>
13. (F-あの一)家に戻って<テ節>
14. 寝ていたんですが/並列節ガ/

13

分析の手順

1. 節境界ラベルの出現数を集計 → 分析1
2. 節境界ラベル **[絶対境界]** の直後で発話を分割
3. 各発話中の語数を計測 → 分析2
4. 各発話中の節境界ラベル数を計測 → 分析3
5. 節境界ラベルが連鎖するパターンを集計 → 分析4

14

分析1：節境界ラベルの分布

- 対象データ：CSJ（人手で解析済み）
 - 学会講演 987講演（3,309,760語）
 - 模擬講演 1,715講演（3,636,368語）
- 節境界ラベルの出現数

	絶対境界	強境界	弱境界	合計
学会講演	84,047	45,055	108,859	237,961
	35.3%	18.9%	45.7%	100.0%
模擬講演	82,794	67,950	178,361	329,105
	25.2%	20.6%	54.2%	100.0%

→ 模擬講演に多く出現

15

分析1：節境界ラベルの分布

- 頻出する節境界ラベル（上位10位）

学会講演		模擬講演	
75,554	31.8% [文末]	69,474	21.1% [文末]
40,943	17.2% <テ節>	54,702	16.6% <テ節>
19,839	8.3% <引用節>	33,944	10.3% <引用節>
14,007	5.9% /並列節ガ/	15,281	4.6% /テ節/
8,431	3.5% <連用節>	12,928	3.9% <理由節ノデ>
8,293	3.5% /テ節/	12,451	3.8% <並列節テ>
6,896	2.9% /並列節クレドモ/	12,049	3.7% [文末候補]
6,777	2.8% /条件節ト/	11,236	3.4% /並列節ガ/
6,514	2.7% <並列節テ>	11,230	3.4% /並列節クレドモ/
6,343	2.7% [文末候補]	10,422	3.2% /並列節ケド/

16

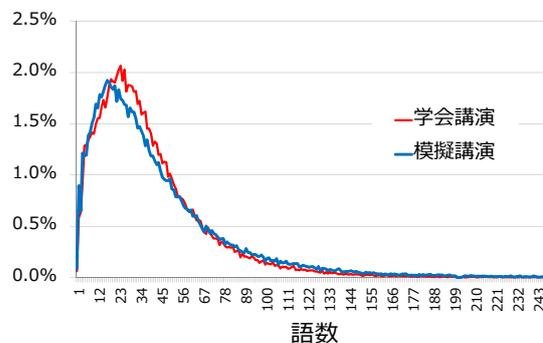
分析2：[絶対境界]までの語数

- 発話を **[絶対境界]** の位置で分割
 - 学会講演：84,047発話
 - 模擬講演：82,794発話
- 各発話中に含まれる語数を計測

	学会講演	模擬講演
平均値	39.36	43.89
中央値	32	33
第1四分点	19	18
最大値	477	1,192

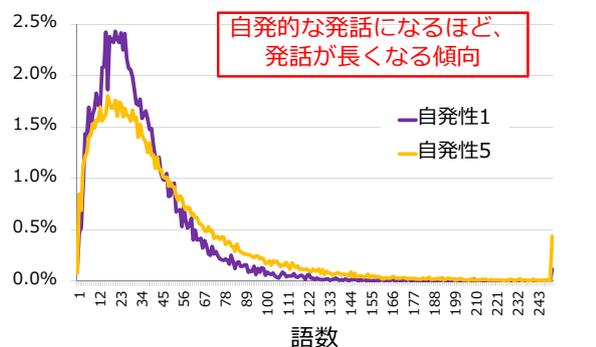
17

分析2：[絶対境界]までの語数



18

分析 2 : [絶対境界]までの語数

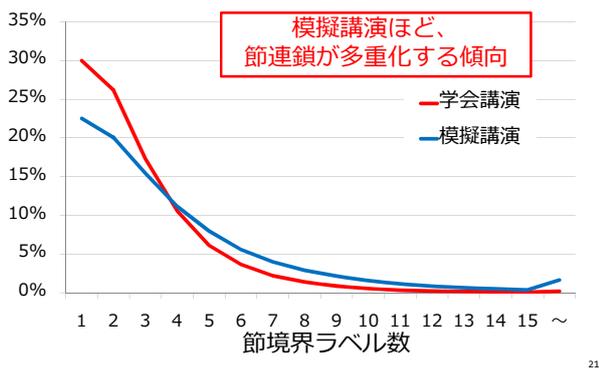


分析 3 : 節境界ラベルの接続数

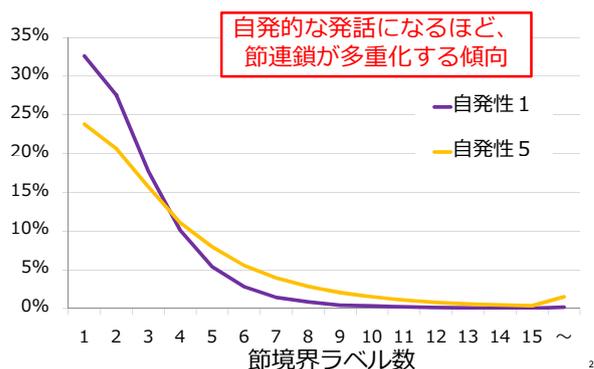
- 発話を [絶対境界] の位置で分割
 - 学会講演 : 84,047発話
 - 模擬講演 : 82,794発話
- 各発話中に含まれる節境界ラベル数を計測

	学会講演	模擬講演
平均値	2.83	3.98
中央値	2	3
第1四分点	1	2
最大値	37	140

分析 3 : 節境界ラベルの接続数



分析 3 : 節境界ラベルの接続数



分析 4 : 節境界ラベルの接続パターン

- 節境界ラベルのみを抽出して、その接続パターンを集計

- 学会講演 : 84,047発話 → 16,910パターン
- 模擬講演 : 82,794発話 → 29,735パターン (異なり)

→ 節連鎖構造のパターンは、模擬講演の方がはるかに多様。

分析 4 : 節境界ラベルの接続パターン

- 頻出する節境界の連鎖パターン (上位12位)

学会講演		模擬講演	
22,466	26.7% [文末]	15,221	18.4% [文末]
6,708	8.0% <テ節> [文末]	3,519	4.3% [文末候補]
2,404	2.9% [文末候補]	3,453	4.2% <テ節> [文末]
2,259	2.7% <引用節> [文末]	2,550	3.1% <引用節> [文末]
1,907	2.3% /並列節力/ [文末]	875	1.1% /並列節力/ [文末]
1,426	1.7% <連用節> [文末]	846	1.0% /テ節/ [文末]
1,235	1.5% /テ節/ [文末]	810	1.0% <並列節テ> [文末]
1,179	1.4% <テ節> <テ節> [文末]	742	0.9% <テ節> <引用節> [文末]
875	1.0% <テ節> <引用節> [文末]	700	0.8% <連用節> [文末]
842	1.0% <並列節テ> [文末]	662	0.8% <テ節> <テ節> [文末]
840	1.0% /並列節ケレドモ/ [文末]	618	0.7% <理由節ノテ> [文末]
737	0.9% /条件節ト/ [文末]	616	0.7% /並列節ケレドモ/ [文末]

分析4：節境界ラベルの接続パターン

□ [文末] のみを持つ発話 = 定形的

学会講演

296 以上です[文末]

92 すいません[文末]

79 以上で発表を終わります[文末]

68 ありがとうございます[文末]

67 よろしく願います[文末]

47 よろしく願います[文末]

42 以上で発表終わります[文末]

29 以上で終わります[文末]

28 どうもありがとうございました[文末]

21 次お願いします[文末]

模擬講演

381 以上です[文末]

252 これで終わります[文末]

114 終わります[文末]

87 以上で終わります[文末]

56 ありがとうございます[文末]

36 終わります[文末]

32 これで終わります[文末]

27 はい以上です[文末]

19 後ですね[文末候補]

18 以上終わります[文末]

25

分析4：節境界ラベルの接続パターン

□ 発話中の節境界の数が増えるに従い、語り（ナラティブ）の性格が強くなる。

- ガ、ケレドモ：前提の提示、話題の切れ目
- タラ、ト：話題の展開
- デシテ/マシテ、テ+接続詞：発話の切れ目

□ 話題の展開に伴う、発話の動的な構築

→ 自発的な話し言葉の特徴

26

分析4：節境界ラベルの接続パターン

1. 一週間ぐらいて<テ節>
2. 良くなったかなというところで(F-えー)ころ日常生活に復帰したんですけども/並列節ケレドモ/
3. そういたしましたら<条件節タラ>
4. また一日復帰したら<条件節タラ>
5. その日の夜に殆ど前と同じようにまた四十度ぐらいばって熱が出まして/テ節/
6. もうこれはどうなってんのかねと<引用節>
7. 思っ<テ節>
8. 次の日の朝に病院に行きましたら<条件節タラ>
9. (F-そのー)インフルエンザみたいな病気をした後一週間ぐらいたって<テ節>
10. そのようにまた高く発熱するっていうのは何か肺炎とかそういうの疑いがあるって<引用節>
11. 言われて<テ節>
12. その場でレントゲンを撮りまして/テ節/
13. 何かこう血を取りまして/テ節/

27

話し言葉の文法研究

- 自発音声コーパスに基づく定量的な研究
- モデル化しにくい研究
- 実際の言語運用における「文」と「発話」

□ 現実の発話を内省すればわかるように、あらかじめ、頭の中で<文>を構成してから話し始めるわけではない、頭に浮かぶことをつぎつぎと継ぎ足して、ひと通りの表出が終わったところが、結果的に<文>の末尾になる。(中略) <文>を確定したうえで構文を合理的に分析しようとする方法は、発話の実態に即していない。

(小松英雄『仮名文の構文原理 [増補版]』)

28