

## 日本語文末表現の文体的単調さの解消

河野香織, 黒澤義明, 相沢輝昭

広島市立大学情報科学部

{kaori,kuro,aizawa}@nlp.its.hiroshima-cu.ac.jp

## 1 はじめに

日本語の書き言葉における文末表現は、例えば「～だ。～だ。～だ。」のように、同じ音が続いて単調になりがちとなる。このような“単調さ”は、日本語の基本構文では述語が文の最後に置かれることに起因する。述語は通常、基本形の語尾がウ段である動詞、「ダ」や「タ」等の助動詞、基本形語尾が「イ」である形容詞、と決まった音で締め括られるからである。特に叙述文では「～デアル」や「～ダ」が多く、更に過去形であれば「タ(ダ)」で終わるしかない。このような単調さについては、『文章読本』をはじめとした多くの文献で言及されている<sup>1</sup>。

文末に関する研究は数多く存在するが、その多くはテンスやアスペクト、モダリティといった観点からの文末の機能的なものに関する研究であり、文末の単調性、及びその解消を目指した研究は見受けられない。本研究は、文末における単調さを解消する方法について考察を進めていくものである。尚その際、意味解析等の諸解析は行わない。表層表現と最小限の意味情報(辞書)のみを用いて、高速に処理する手法について考察する。

以下に考察を進めるに当たっての留意点を挙げる。単調さの解消は、文章の推敲の一手法であり、他人の文章も対象となる。したがって、書き手の個性という意味での文体を損ねないよう気をつける<sup>2</sup>。

単調さについての考察対象は、口語体・常体・書き言葉に限定した。口語体とは、現在一般的に用いられている言葉のことであり、常体とはデスマス調ではない、いわゆる敬語を用いない文体のことである。書き言葉としてあるのは、「～ダ」「～ヨネ」等の会話文における単調さは考えないことを意味している。

<sup>1</sup>[外山 84, 丸谷 77] 他。<sup>2</sup>テンス・アスペクト・モダリティは変えない立場である。

## 2 単調さの尺度

単調さについて考察を進めるに当たり、どのような文章が単調であるのか、その尺度が必要である。その尺度の指標として、前節の留意点を満たし、内容に偏りのない朝日新聞のコラム『天声人語』(’97年6月～12月)を採用した。

## 2.1 尺度の指標 1

まず、『天声人語』ではどの程度同じ文末が連続しているのかを、形態素・音・母音に着目してその連続回数を調査した。結果を表1に示す。

表1 天声人語における同一文末の連続回数

種類 \ 連続数	2	3	4	5	6	7	8 ≤
形態素	249	78	29	7	7	1	2
音	381	104	33	8	8	2	2
母音	565	218	89	32	13	12	6

連続回数が増える毎に、ある一定回数まで出現頻度はほぼ単調減少している。この結果から、同一文末の連続回数が増すと共に単調さも増加し、閾値に相当する連続回数があると予想される。文章に対する評価は個人差があると考えられるため、尺度の決定は本節の仮説を元に、次節の調査により決定する。

## 2.2 尺度の指標 2

## [調査]

前節で述べたように、『天声人語』のデータでは、一定以上の連続回数を持つ文章の出現頻度が減少している。そこで、同一文末の連続回数を操作する調査を行い、連続回数の増加が読み手の単調さ評定に及ぼす効果について検討する。

## [方法]

被験者: 大学生 100 名。

刺激: 同一文末連続回数の要因を操作するため、『天声人語』より5日分の記事を選び、各々の記事に対し文末を修正し、文末連続回数の異なる5種類の文章を作成した。なお、5種類の文末連続回数とは、同種の文末音が4, 5, 6, 7回連続する4

種類の文章、及び文末音が3回以上連続しない比較対象(以下Cと略す)の、計5種類を指し、上記要因の5水準を表す。

手続き: 調査に先立ち、5日分の異なる記事(5文章)から構成された冊子を作成した。この5文章は、連続回数5水準全てが出現するように選ばれており、連続回数は被験者内要因である。

被験者には、上記の冊子を配布し、各々の文章に対する単調さを7段階で評定するよう求めた。

結果: 評価1～7に値1～7を割り振り、連続回数毎に平均をとったものを表2に示す。

表2 連続回数毎評定平均値

連続回数	4	5	6	7	C
評価	4.14	4.40	3.90	4.19	3.78

被験者内計画に基づき、1要因の分散分析を行った<sup>3</sup>。その結果、連続回数の主効果が認められた( $F(4,396)=165.44$ ,  $p < .10$ )<sup>4</sup>。したがって、文末の連続回数が単調さの判断に影響を与えていると考えられる。

次に、4種類の連続水準と非連続の水準間の組み合わせについて、TukeyのHSD検定による平均値の差の検定を行ったところ、連続回数5回と非連続文との水準間に有意差が認められた( $5-c=3.23 > \text{HSD 臨界値 } 3.11$ )。しかし、他の組み合わせに有意差は見られなかった( $4-c=1.81$ ,  $6-c=0.52$ ,  $7-c=2.07 < 3.11$ )。

上記の結果、連続回数5回の文章は単調であると判断可能である。しかし、『天声人語』のデータから予想された6回、7回の高水準には有意差が認められなかった。一定回数以上の繰り返しは読み手にリズム感を与える等、単調さを減少させる方向に働く可能性があるとも考えられる。この点については、さらに検討が必要である。

以上のことから、本研究では暫定的に、同一文末が5回連続する文章を、単調な文末を持つ文章であると考えることとする。

### 3 非単調化の手法

単調である文を非単調な文へと推敲する手法について考察する。

同一文末が5回連続する文が単調な文章である。したがって、連続する文中のいくつかを違う

<sup>3</sup>なお、分析に先立ち、評定値のT尺度化を行っている[岩原65]。

<sup>4</sup>被験者の主効果は、有意であった( $F(99,396)=119.56$ ,  $p<.01$ )。なお、本研究は有意水準を10%に設定している。

母音の文末に変換することにより、その連続性が断ち切れ、単調さは解消され则认为られる。

文末の変換手法として考え得るものを以下に挙げる。

1. 同義語動詞の変換
2. 類意表現同士の交換
3. 同意接尾辞の変換
4. サ変動詞の名詞化(用言の体言化)
5. 過去形から現在形への転換
6. 複数の文を一文に統合
7. 倒置法

ここでは1章で述べた留意点を考慮し、3.、4.の2手法を採用する。これらは、意味解析を必要とせず、変換によってニュアンスが大きく変化しない。尚、同意接尾辞の交換とは、例えば、「～デアル」を「～ダ」に変換することであり、サ変動詞の名詞化とは、サ変動詞から「スル」「シタ」を削除し、サ変名詞化することである。

### 4 システムの構築と評価

単調さの尺度を元に、同一文末が5回連続する箇所を単調箇所として検索し、非単調化を行うシステムを作成した<sup>5</sup>。入力された文章の文末に対してその情報を数値化したタグ情報<sup>6</sup>を得る。その情報を元に単調箇所を検索する。単調箇所の候補に対しては、「ダ」がマ行五段動詞などの過去形ではないか、「シタ」がサ行五段動詞の過去形ではないか等を辞書<sup>7</sup>で確認した後、変換を施す。

変換する位置に関しては、文末の変換手法が限られているため、変換可能な場所は変換する方針を採る。しかし単調箇所すべてに同意接尾辞の変換が行われる場合、違う母音の単調箇所が出来る可能性があるため、一度変換されるとその最も近い同意接尾辞を持つ文末に対し、同変換はしないようフラグを立てて回避するものとする。また単調箇所前後の文末情報も考慮する。変換箇所が他にない場合を除き、変換後の母音が前後の母音と同じであれば、変換しないものとする。

<sup>5</sup>一度、変換が成された後は、ユーザーの好みにより、尺度の設定変更が可能である。

<sup>6</sup>タグの数値は、ア段には10代、イ段には20代、ウ段には30代、…(以下省略)と割り振る。

<sup>7</sup>本システムで使用する辞書はすべて、[EDR]から作成した。

図1に変換例を示す。

《変換前》	タグ: 12 11 12 12 12 50
いま世界で広く採用されているのはキリスト紀元、つまり西暦だ。キリストが生まれた年を起算の年＝紀元とした。イスラム教徒が使うのはヒジュラ紀元だ。教祖ムハンマド（マホメット）が迫害を受け、メッカから逃れた西暦622年が起算点だ。ヒジュラは「聖遷」の意味だ。キリスト紀元は、6世紀にローマの神学者が考案したもの。	
↓	
《変換後》	タグ: 31 11 12 31 12 50
いま世界で広く採用されているのはキリスト紀元、つまり西暦である。キリストが生まれた年を起算の年＝紀元とした。イスラム教徒が使うのはヒジュラ紀元だ。教祖ムハンマド（マホメット）が迫害を受け、メッカから逃れた西暦622年が起算点である。ヒジュラは「聖遷」の意味だ。キリスト紀元は、6世紀にローマの神学者が考案したもの。	

図1 非単調化を施した例

1文目の語尾「だ」が「である」に変換される<sup>8</sup>。2文目は「した」が候補の対象となるが、辞書<sup>9</sup>の検索からサ変動詞ではないと判断され、変換は行われぬ。1文目で同意接尾辞に変換されていることから、3文目の「だ」は変換されず、4文目の「だ」が「である」に変換される。前の変換から5文目は変換されず、全変換結果を出力し、終了する。

## 5 実験からの考察

システムの問題点を挙げるためには、できるだけ多く、様々なタイプの文章を入力する必要がある。その一つとして、尺度の指標として用いた'97年6月から12月までの『天声人語』をすべてシステムに通した。

結果、全27箇所が単調箇所として検索され、内20箇所(74.0%)が変換された。ここで明らかになった問題点を以下に挙げる。

まず、サ変動詞の名詞化に用いる辞書の不備が問題となる。日本語は漢字かなの書き換えが自由であるため、難しい漢字を使う動詞は、平仮名で書かれている場合が多い。例えば「醸す(かも

<sup>8</sup>「んだ」「いだ」ではないことから、マ行五段動詞などの過去形ではないことが分かる。

<sup>9</sup>サ変動詞であるか検索する辞書を指す。

す)」というサ行五段動詞<sup>10</sup>などである。EDR辞書では、一部平仮名での記載もあるが、漢字のみ記載された項目も多く、作成した辞書が平仮名で書かれた文には対応できない。しかし、安易に漢字を平仮名に対応させた辞書を作成すると、別の問題が発生してくる。この他、送り仮名による語表記のゆれに対応しきれていないという問題がある。

次に、変換された箇所に関する問題点を挙げる。変換後の文には、文法的に間違っていないが、変換しない方が適切と考えられる箇所が見られた。例えば「戦争による」または「社会的」障害者の優先は、兵役に関係する。」のように主部が長く述部が短い場合、「する」を削るとバランスが悪いという印象を与える等である。これは主観の問題であると考えられるが、文章を扱う上では無視できない問題である。このように、文中のバランスも考慮すべきであると考えられる。

実験においては74%の率で単調箇所が変換された。しかし、上述の問題を考慮すると、変換率は低くなると考えられる。3章に挙げた変換手法をすべて用いるなら、改善は望まれるが、それでも単調さが解消できない場合、モダリティの変換規則を適用することが考えられる。

## 6 モダリティからの考察

日本語の文末にはモダリティの文法表現が集中している。したがって、文末の変更はモダリティに影響を及ぼすことになる。ここでは、[仁田91]の日本語のモダリティの体系に従って考察する。

### 言表事態めあてのモダリティ

本研究では、考察の対象をひとまず口語体の常体の書き言葉に絞った。そこで、言表事態めあてのモダリティとして主に考えなければならないのは〈判断〉のモダリティである[仁田91]。

#### (A) 書き手の把握・推し量り作用を表すもの

スル形で表される断定と、スルダロウ形・スルマイ形で表される推量とがある。これらは、書き言葉を構成する文の中心的存在と言える。断定の文末には、動詞と形容詞の終止形がくる。推量の文末は助動詞「ウ」「マイ」の終止形がくる。

#### (B) 推し量りの確からしさを表したもの

<sup>10</sup>システムに通した文中に「物議をかもした。」とあったが、「かもした」は、サ行五段ではなくサ変動詞と見做されて「した」が削除され、「物議をかも。」と変換された。

代表的な表現形式は「～ニチガイナイ」や「～カモシレナイ」である。

(C) 徴候の存在の下での推し量りをあらわすもの  
これには「ラシイ」「ソウダ」「ヨウダ」「ミタイダ」等の表現形式がある。

(D) 推論の様態に関するもの

「ハズダ」が代表的な表現形式である。

ここで、(A) 表現から (C) 表現等への書き換えを許せば、文末音は変化する。その結果、〈判断〉のモダリティが変わり、ひいては文の意味まで変わることになる。しかし、日本語の書き手は、〈判断〉のモダリティの転換にまで踏み込むことがある。例えば、未来の行動を表現する際、(A) 表現“彼がする”を (C) 表現“彼がするらしい”、“彼がするみたいだ”とあえて言い替える等である。

#### 発話・伝達のモダリティ

日本語のモダリティにはもう一種、発話・伝達のモダリティがあり、これが言表事態めあてのモダリティを包み込む形で文類型を定める [仁田 91]。

〈判断〉のモダリティを包み込む発話・伝達のモダリティは〈述べ立て〉と〈問いかけ〉であるが、標準的な書き言葉に現れるのは主に〈述べ立て〉である。このモダリティによって、(a) 現象描写文、(b) 判定文、(c) 疑いの文の3種の文タイプが定まる。また、このモダリティの付加によって文末音「タ」が加わることがあり、また疑いの文によって新しく「カ」等の文末音が加わることもある。

それでも尚、日本語文末は単調である。ゆえに書き手は〈述べ立て〉の文の間に、意志・希望・願望を示す〈表出〉の文を差し挟むことがあると考えられる<sup>11</sup>。この〈表出〉が「ヨウ」「マイ」「タイ」等の形態素を文末に加え、「ウ」・「イ」音を文末音に付け加える。

更に、疑いの文も〈表出〉の文も、特定の読み手の存在を必要としない。しかし、この枠を越えて、特定の読み手／聞き手に向かって直接に〈問いかけ〉たり〈働きかけ〉たりする文を書き言葉として使えれば、終助詞や命令形も使えることになり、文末の多様さは増す。

以上、文末の多様さ（むしろ貧困さ）とモダリティの関係を見てきた。3章の手法で単調さが解消できない場合には、以上のようなモダリティの転換にまで踏み込んだ変換が必要である。

<sup>11</sup>例えば、料理のレシピ等においてみられる。

## 7 おわりに

本研究では、文末表現の単調さの尺度を求め、非単調化の手法を提示し、それに基づいて非単調化を行うシステムを構築した。

今後の課題としてはまず、単調さの尺度に関する問題が挙げられる。今回は、『天声人語』とアンケートを用いて尺度を決定したが、この尺度は、『天声人語』の分析を中心に行ったため、今回対象とした文体すべてに適用できるとは限らない。別のサンプルとの比較検討が必要である。

また対象とする文への制約が多いため、変換手法が限られるという問題点があった。特に強力な手段として、過去形から非過去形への変換、複数の文を一文にまとめる手法を検討し、変換の手法を強化することも今後の課題として挙げられる。

更に、本研究は、他人が作成した文章を考察の中心とした。そこで設定した、文体を変えない等の制約は、自分で書いた文章の推敲を前提とした場合には取り除くことができる。書き手自身の作文を支援する方向から、3章で挙げた手法を再検討していくことも今後の大きな課題として挙げられる。

## 謝辞

アンケート調査に御協力いただいた本学の岩根先生、任先生に感謝致します。尚、本研究の一部は文部省科学研究費補助金基盤研究(C) (課題番号 09680376) による。

## 参考文献

- [外山 84] 外山滋比古, 日本 の文章, 講談社 (1984)
- [丸谷 77] 丸谷才一, 文章読本, 中央公論社 (1977)
- [岩原 65] 岩原信九郎, 新訂版 教育と心理のための推計学, 日本文化科学社 (1965)
- [仁田 91] 仁田義雄, 日本語のモダリティと人称, ひつじ書房 (1991)
- [EDR] 日本電子化辞書研究所, EDR Ver.1.5 第2章 日本語単語辞書, 日本電子化辞書研究所 (1995)
- [益岡他 97] 益岡隆志・仁田義雄・郡司隆男・金水敏, 岩波講座 言語の科学 5 文法, 岩波書店, 1997
- [河野 98] 河野香織, 日本語文末表現の文体的単調さ解消の研究, 広島市立大学情報科学部卒業論文 (1998)