

# 話し言葉を対象とした文末解析と表現文型の同定

柳 将吾 佐藤 理史 夏目 和子 宮田 玲 小川 浩平

名古屋大学大学院工学研究科

yanagi.shogo@i.mbox.nagoya-u.ac.jp

## 1 はじめに

日本語の文末には、テンス・アスペクト・モダリティ・否定などの文法的機能が集中して現れる。たとえば、

- (1) 彼は面白い小説を書いていたにちがいません

という文では、動詞「書く」に、「ている(アスペクト)」、「た(テンス)」、「にちがいない(モダリティ)」「ます(敬体)」が接続し、文末述語を構成している。日本語文において、文末述語の範囲を同定し、その構成要素を同定することは、文の意味を解釈するために不可欠である。

佐野らは、書き言葉を対象に、このような文末解析を実行するシステム Panzer を実現した [1]。Panzer は、与えられた文の文末述語の範囲を決定し、その範囲を内容語と機能語の列に分解する。これらの要素はいずれも HaoriBricks3 (HB3) [2] の Brick (文を組み立てるための部品) として出力されるため、解析結果から元の表層文を復元することが可能である。Panzer は、すでに、書き言葉・話し言葉変換 [3] や、文の平易化 [4] に利用されている。

本論文では、小説の会話文を中心とした話し言葉に対して、Panzer と同等の機能を提供するシステム Panzer/U と、それを利用した発話文の表現文型の同定について述べる。一般に、話し言葉の解析は、書き言葉の解析より難しいと考えられている。そのひとつの理由は、話し言葉に見られるくだけた表現に起因する。Panzer/U では、くだけた表現にも比較的広く対応した「現代話し言葉 UniDic [5]」(以下、UniDic と略記) を形態素解析に利用する。

## 2 音変化表現と形態素解析

小説の会話文や漫画などでは、話し言葉に現れる音変化が文字として現れる。これを宮崎ら [6] は音変化表現と呼んでいる。具体例を以下に示す。

- (2) 彼にちがいねえ (ちがいない)

- (3) 本当に走んのか (走るのが)

例文 (2) は、Juman++ でも UniDic/MeCab でも適切に解析できるが、例文 (3) の解釈は、次のように分かれる。

- (4) a. 走(走る)/ん(ぬ)/の(のだ)/か (Juman++)  
b. 走ん(走る)/の(準体助詞)/か (UniDic)

(4a) のように「走ん」の「ん」を「ぬ」と解釈するのは明らかに誤りであり、(4b) のように「走ん」を「走る」の音変化表現と解釈するのが適切である。ここでは 1 例を示したが、UniDic は、音変化表現に対して比較的頑健である。

音変化表現を UniDic で解析した結果は、おおよそ、次のように分類できる。

1. 独立した語として認識される  
例：書い/てく (助動詞てく)
2. ある語の異形として認識される  
例：すんごい (すごい [終止形-一般])、  
うるせえ (うるさい [終止形-一般])
3. 活用形として認識される  
例：走ん/のだ (走る [終止形-撥音便])
4. 正しく解析できない

しかしながら、いずれの場合も、「音変化が起きている」ことが明示的に出力されるわけではない。そのため、音変化表現の検出には、4 節に示すような検出規則が必要となる。

## 3 Panzer/U の概要

Panzer/U の構成を図 1 に示す。基本的な構成は Panzer [1] を踏襲している。Panzer/U は、前処理(形態素解析)の結果として得られる形態素列に対して、文末述語の範囲を決定し、その範囲を HB3 の Brick 列(内容語と機能語列)に変換する。オリジナルの Panzer では、前処理として、形態素解析と文節解析(Juman+KNP)を用いているが、Panzer/U の前処理

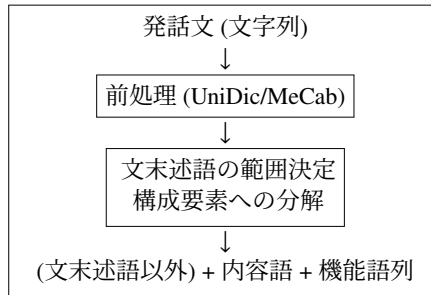


図1 Panzer/U の構成

は、UniDic/MeCab による形態素解析のみで、文節境界の決定は、組み込み規則によって行う。以下に、Panzer/U の実行例を示す。ここで、a が入力、b と c が出力であり、b は内容語、c は機能語列を表す。下線は音変化に関わる Brick [7] を表す。

- (5) a. 張り切ってるわね  
b. 張り切る  
c. テ助動詞る, 終助詞わ, 終助詞ね
- (6) a. まさか、遭難じゃないだろなあ  
b. 遭難  
c. じゃない, だろう, ウ脱落, 終助詞なあ
- (7) a. こんなに暗いのに、車にはねられたらどうすんだ！  
b. どうする  
c. ル脱落, 連体助動詞んだ, 感嘆符
- (8) a. おまえ、日本語しゃべれるじゃねえか  
b. しゃべれる  
c. じゃない, アエ交替, 終助詞か, 句点

HB3 は益岡・田窪文法 [8] とそれに準拠する Juman 体系から発展した羽織文法に基づくため、HB3 の Brick と Juman の形態素の差異は比較的小さい。その一方で、UniDic の短単位との差異は、かなり大きい。そのため、Brick 列への変換の過程で、この差異を吸収する必要がある。

短単位列から Brick 列への変換は、おおよそ、次の 3 種類に分類できる。

1. 文脈を考慮しない変換
  - (a) 1 つの短単位 → Brick (列)  
例：終助詞ね → 終助詞ね
  - (b) 複数の短単位 → Brick (列)  
例：書い + 助動詞た  
→ 書く + タ形  
書い + 接続助詞て  
→ 書く + タ系連用テ形

## 2. 文脈を考慮した変換

例：(食べ) 終助詞な → 終助詞な命令  
(食べる) 終助詞な → 終助詞な禁止

タイプ (1a) は単純な変換である。タイプ (1b) は単位の差異を吸収する変換で、Panzer/U ではこのタイプの変換が多数存在する。例に示したような、活用語の直後の助動詞や接続助詞を、活用語尾 (活用形) とする変換は、その典型である。

タイプ (2) の文脈に依存した変換は、短単位 (列) が、複数の変換候補を持つ場合である。たとえば、HB3 では命令を表す「終助詞な命令」と、禁止を表す「終助詞な禁止」を区別する。そのため、「終助詞な」の変換先は、直前の語の活用形を参照しなければ定まらない。

なお、Panzer/U の開発には、『舟を編む』 [9]、および、『風が強く吹いている』 [10] (0 章から 5 章の途中まで) から抜き出した発話文、約 3000 文を開発用データとして利用した。

## 4 音変化表現の検出

Panzer/U の実装に先立ち、音変化表現を生成する仕組み (Brick) を HB3 に導入した [7]。HB3 における音変化表現の扱いには、次の 5 種類がある。

1. 内容語として扱う  
例：やっぱし
2. 音変化形 (異形) を作り出すコマンドで扱う  
例：ウ脱落, ル脱落, アエ交替
3. 活用形として扱う  
例：条件音便形, 条件音便縮約形, 連用ウ音便形
4. 機能語として扱う  
例：テ助動詞る, テ縮約接尾辞とく
5. テンス有標 (タ) の拡張として扱う  
例：タあ

解析を担う Panzer/U は、入力文に含まれる音変化表現を検出し、表層文字列に復元可能な Brick (列) に変換する必要がある。以下では、2 節に示した 4 つの場合の概要を示し、付録に実装した検出規則の一覧を示す。なお、文末解析では、文末の機能語とその直前の内容語 (多くは、活用語) を扱うため、実装した検出規則は、これらに関わる音変化を対象としている。

**1. 独立した語として認識される場合** この場合の変換は容易で、あらかじめ用意した変換テーブルで、Brick に変換する。対象となるのは機能語で、具

体例には「てる・てく・とる・とく・ちまう・ちゃう」などがある。

**2. ある語の異形として認識される場合** 活用しない内容語の音変化、および、活用する内容語の語幹の音変化<sup>1)</sup>は、新たな(表記の)内容語として扱う。つまり、音変化は明示的に検出しない。それ以外の語に対しては、検出規則によって音変化を検出し、対応する Brick に変換する。

たとえば、「すげえ」は「すごい」の異形として認定される。しかし、その活用形は「終止形一般」であり、活用形だけからは音変化が起きているかどうかはわからない<sup>2)</sup>。そのため、検出規則を適用して、音変化表現かどうかを決定する。

**3. 活用形として認識される場合** 活用形「仮定形-融合」は、それだけから音変化が起きていることがわかるが、「美しけりゃ」も「美しきゃ」も「仮定形-融合」と認定されるため、どのような音変化が起きているか一意には決定できない。そのため、検出規則を適用し、音変化表現の種類を同定する。

**4. 正しく解析できない** UniDic を用いても、すべての音変化表現が正しく解析できるわけではない。いくつかの典型的な誤りに対しては、誤りを修正して音変化表現として検出したり、音変化が起きている表現に修正したりする。以下に例を示す。

(9) 食べりゃ / あ / いい (あ [感動詞])

→ 食べる, 条件音便形, ア長音化, ばいい

(10) 食べ / て / やん / の (やる [連体形-撥音便])

→ 食べる, テ助動詞やがる, 連体省略撥音形, 終助詞の

(11) 食べる / ねえ (ない [終止形-一般])

→ 食べる, 終助詞ねえ

## 5 Panzer/U の性能

以下のデータを用いて、Panzer/U による文末解析の性能評価を行なった。

**開発用データ** システム開発に使用した発話文の一部。『舟を編む』[9]の1章から100文(D1)。『風が強く吹いている』[10]の2章と5章からそれぞれ100文(D2, D3)。

**評価用データ** 『風が強く吹いている』の7章から100文(T1)。『氷菓』[11]の3章と5章からそれぞれ100文(T2, T3)。

1) 語幹末尾の母音の音変化を除く。

2) UniDic では、一つの活用形に対応する表層表記が複数存在する。

表1 Panzer/U の性能評価

	OK	NG	?	計
D1	93 (19)	6 (3)	1 (0)	100 (22)
D2	91 (20)	7 (0)	1 (1)	100 (21)
D3	91 (34)	4 (0)	1 (0)	100 (34)
T1	89 (19)	6 (0)	2 (0)	100 (19)
T2	99 (25)	1 (0)	0 (0)	100 (25)
T3	92 (14)	5 (3)	3 (0)	100 (17)

これらの文は、出典において引用記号(鉤括弧)で囲われた部分に出現し、かつ、文末に述語(動詞・形容詞・判定詞)が存在する文である。

Panzer/U の出力の正誤判定結果を表1に示す。括弧内の数字は、文末に音変化表現が含まれている文数(内数)を表す。なお、出力の正誤判定は、それほど自明ではないため<sup>3)</sup>、正誤判定が難しい場合は「?」とした。

この表より、おおむね90%の発話文が正しく解析されていることがわかる。発話文の23%(138/600)には文末に音変化表現が含まれているが、それらの文でも精度が特に低いわけではない。

しかしながら、この結果をもって、音変化表現を含む文末表現が高い精度で解析できるようになったと判断するのは早計である。文末述語に音変化表現を含む138文中、97文は「じゃない・んだ(んです)・てる」のみを含む文であった。つまり、高い頻度で現れる音変化表現を正しく解析できているだけで、多くの種類の音変化表現を正しく解析できるかどうかは不明である。そのため、たとえば宮崎ら[6]が示した音変化パターンに基づいて例文集を作成し、これらがどの程度正しく解析できるかを確かめる必要がある。

一方、音変化表現を含む文の解析に失敗した原因のほとんどは、形態素解析の誤りであった。たとえば、末尾への促音(「っ」)の付加の具体例である「それは、本当ですかっ?」は、UniDicでは以下のように形態素解析され、「すかっ」は副詞と判定される。

(12) それ / は / 、 / 本当 / で / すかっ / ?

出現頻度が低い音変化表現は、UniDicでも正しく解析できないことが多いように思われる。このことから、多くの種類の音変化表現を含む例文集を作成して、発話文解析ツールの開発や評価に使用することが重要となろう。

3) 特に、内容語が複合語の場合、どこを文節境界と認定するかが難しい。

## 6 表現文型の同定

Panzer/U の開発目的は、発話文の文末解析の実現にとどまらない。文末解析結果を利用して、発話文を『発話文表現文型辞書』[12]の文型に結びつけ、発話意図や話し方の特徴の情報を取得可能とすることを最終目標としている。

『発話文表現文型辞書』は、小説の発話文生成を念頭に、「ある目的(発話意図)で発話する時、ある話し方で表すならば、この文型を使う」という情報を提供する辞書である。この辞書は、発話文生成での使用を想定して設計されたが、発話文解析において、与えられた発話文をこの辞書のエントリに結びつけることができれば、その発話文の発話意図や話し方特徴が判明する。これを実現すべく、『発話文表現文型辞書』の改定と Panzer/U を利用した表現文型の同定の実現を並行して進めている。

辞書の新しい版では、文末述語を、次のような形式でモデル化する。

### (13) 中核要素 + 形式 + 主要部 + 末尾部

ここで、中核要素は、この述語の中核となる部分(動詞、形容詞、名詞+判定詞)を表す。一方、形式は、この中核要素の活用型、および、テンス・否定の有無を、それぞれ以下のようなアルファベット1文字で表したものである。

#### 1. 活用型

V(動詞型)、A(イ形容詞型)、C(判定詞型)

#### 2. テンス・否定

b(なし)、t(テンスあり)、n(否定あり)、k(テンス・否定あり)

末尾部は、文末述語の最後尾に位置する終助詞(列)を、主要部は、それ以外の要素(機能語列)を表す。以下に、具体例を示す。

- (14) a. 食べる方がいいよ  
b. 食べる,方がいい,よ  
c. 食べる, Vb,方がいい,よ  
d. 食べる, Vb, 文型\_方がいい\_よ
- (15) a. 行かない方がいいわ  
b. 行く,ない,方がいい,わ  
c. 行く, Vn,方がいい,わ  
d. 行く, Vn, 文型\_方がいい\_わ

ここで、a は入力、b は Panzer/U の出力、c は文末述語のモデルに沿った分解であり、d が表現文型の同

定結果として出力する形式である(一部、情報を省略した)。

Panzer/U の出力から、表現文型を同定するために、まず、それぞれの表現文型がどのような機能語(Brick)列から構成されるかを定義しておく。実際の同定処理は、Panzer/U の出力を前から見ていき、中核要素と形式に対応する要素(内容語、判定詞、タ、ない)を取り除き、残った機能語 Brick 列に基づき文型を決定すればよい。

文型「方がいいよ」「方がいいわ」は、いずれも発話意図「忠告」の文型として辞書に定義されている。さらに、前者は、特に話し方特徴を持たないと定義されているのに対し、後者は「女性的・親しみ」という話し方特徴を持つと定義されている。表現文型が同定され、辞書のエントリが一意に定まれば<sup>4)</sup>、このような情報が得られることになる。

## 参考文献

- [1]佐野正裕, 佐藤理史, 宮田玲. 文末述語における機能表現検出と文間接続関係推定への応用. 言語処理学会第26回年次大会発表論文集, pp. 1483–1486, 2020.
- [2]佐藤理史. HaoriBricks3: 日本語文を合成するためのドメイン特化言語. 自然言語処理, Vol. 27, No. 2, pp. 411–444, 2020.
- [3]柳将吾, 佐藤理史. ウィキペディアから抽出した人物エピソードの話し言葉への変換. 言語処理学会第26回年次大会発表論文集, pp. 437–440, 2020.
- [4]Taichi Kato, Rei Miyata, and Satoshi Sato. BERT-based simplification of Japanese sentence-ending predicates in descriptive text. In *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 242–251. Association for Computational Linguistics, 2020.
- [5]Unidic, (2021-1 閲覧). <https://unidic.ninjal.ac.jp>.
- [6]宮崎千明, 佐藤理史. 発話テキストへのキャラクター性付与のための音変化表現の分類. 自然言語処理, Vol. 26, No. 2, pp. 407–440, 2019.
- [7]佐藤理史, 柳将吾, 夏目和子. HaoriBricks3 による音変化表現の生成. 言語処理学会第27回年次大会発表論文集, 2021(発表予定).
- [8]益岡隆志, 田窪行則. 基礎日本語文法-改訂版-. くろしお出版, 1992.
- [9]三浦しをん. 舟を編む. 光文社, 2011.
- [10]三浦しをん. 風が強く吹いている. 新潮社(新潮文庫), 2009.
- [11]米澤穂信. 氷菓. KADOKAWA(角川文庫), 2001.
- [12]夏目和子, 佐藤理史. 発話文表現文型辞書の設計と編纂. 言語資源活用ワークショップ2019発表論文集, pp. 295–312, 2019.

4) 表現文型の同定ができたからといって、かならずしも表現文型辞書のエントリが一意に決定できるわけではない。なぜなら、同一の文型が複数の表現意図に対して使用できる(定義されている)からである。

## 付録 音変化表現の検出規則一覧

### 活用形=語幹-一般

1. (品詞 = 形容詞  
後続要素 = 取立助動詞なる) → 連用形 + ク脱落  
例：うれしくなる

### 活用形=終止形-一般

2. (表層形末尾 = え  
基本形末尾 = Oい) → オエ交替  
例：すげえ
3. (表層形末尾 = え  
基本形末尾 = Aい) → アエ交替  
例：うるせえ
4. (表層形末尾 = え) → 末尾長音小書化  
例：うるせえ
5. (表層形末尾 = Iい  
基本形末尾 = Uい) → ウイ交替  
例：わりい

### 活用形=終止形-撥音便

6. (活用型 = 助動詞-ヌ) → 終止音便形  
例：食べん
7. (活用型 ≠ 助動詞-ヌ) → ル撥音化  
例：走んぞ

### 活用形=終止形-促音便

8. (基本形末尾 = る) → 基本形 + ル音促音化  
例：入れっか
9. (基本形末尾 = い) → 基本形 + イ音促音化  
例：いっか

### 活用形=未然形-撥音便

10. → 未然形 + ラ撥音化  
例：煽んない

### 活用形=連用形-一般

11. (活用型 = 下一段サ行  
表層形の末尾 = し) → セシ交替  
例：食べさして

### 活用形=連体形-撥音便

12. → ル撥音化  
例：走んのだ

### 活用形=連用形-促音便

13. (活用型 = 上一段  
後続要素 = て・た) → イ音促音化  
例：おっこってしまう
14. (基本形 = 歩く  
後続要素 = て・た) → イ音促音化  
例：歩ってしまう
15. (基本形末尾 = る  
後続要素の先頭 = か) → ル促音化  
例：食べっか

### 活用形=連用形-ウ音便

16. (表層形末尾 = う) → 連用ウ音便テ形  
例：もろうて
17. (表層形末尾 ≠ う) → 連用ウ音便省略テ形  
例：もろて

### 活用形=連体形-省略

18. → ル脱落  
例：食べてんだ

### 活用形=仮定形-融合

19. (品詞 = 形容詞  
表層形末尾 ≠ りゃ) → 条件音便縮約形  
例：美しきゃ
20. (活用型 = 助動詞ナイ  
表層形末尾 ≠ りゃ) → 条件音便縮約形  
例：なきゃ
21. ((上記以外)) → 条件音便形  
例：食べりゃ

### 活用形=意志推量形

22. (表層形末尾 = っ) → ウ促音化  
例：走ろっか
23. (表層形末尾 ≠ う | っ) → ウ脱落  
例：走ろか

### 副助詞って

24. (直前の品詞 = 形容詞  
直前の活用形 = 連用形) → テ形 + テ促音挿入  
例：あたたかくって

### 助動詞た

25. (表層形 = たあ) → タあ  
例：たべちあったあ

### 「ない」

26. (活用形 = 終止形-一般  
表層形 = ね) → 語末長音化 + エ脱落  
例：いいんじゃね

### 「っす」

27. (活用型 = 助動詞デス) → 助動詞っす