

令和 4 年 6 月 1 日現在

機関番号：13901

研究種目：挑戦的研究(萌芽)

研究期間：2020～2021

課題番号：20K21809

研究課題名(和文)日本語述語形式の体系的整理に基づく文末述語解析・生成システムの実現

研究課題名(英文)Automatic Analysis and Generation of Japanese Complex Predicates based on the Complete List of Predicate Components

研究代表者

佐藤 理史(Sato, Satoshi)

名古屋大学・工学研究科・教授

研究者番号：30205918

交付決定額(研究期間全体)：(直接経費) 4,800,000円

研究成果の概要(和文)：日本語では、文末に文法機能を担う多くの要素が出現し、「書かれていたにちがいません = 書く + れる(受身) + テいる(状態) + タ(過去) + にちがいない(推測) + ます(丁寧)」のように複雑な述語表現を形成する。本研究では述語を構成する要素として、助詞137語、判定詞1語、助動詞113語、述語接尾辞74語、複合辞301語を列挙し、その文法的性質を定義することにより、このような複雑な述語表現を自動生成するシステムを実現した。

研究成果の学術的意義や社会的意義

本研究は、機械的な生成の観点から日本語の複雑な述語表現に対する見取り図を与えたという意味において、学術的意義がある。特に、接続型の導入により、活用形を陽に指定しなくとも文法的に正しい述語表現を生成できるようになった。本研究で実現した生成システムを利用することにより、たとえば、書き言葉から話し言葉への自動変換、キャラクタ設定に応じた発話文の自動生成などが可能となる。

研究成果の概要(英文)：Japanese language has very complex predicates such as "Kaka-re-tei-ta-nichigai-arimasen" that are composed by "kaku (write)", "reru (passive)", "iru (aspect)", "ta (past)", "nichigainai (must)", and "masu (polite)". This study has compiled the complete list of these components, which are 137 particles, one copula, 133 auxiliary verbs, 74 predicate suffixes, and 301 compound auxiliary verbs. By defining the grammatical properties of conjugation type and connection type for each component, we have realized a system that automatically generates such complicated predicates.

研究分野：自然言語処理

キーワード：自然言語処理 機能語 文末述語生成 文末述語解析

1. 研究開始当初の背景

日本語では、文末に文法機能を担う多くの要素が出現する。次の例文では、動詞「書く」に5つの要素が付加されて、複雑な述語表現を形成している。

彼の半生は小説に書かれていたにちがいません

書く + れる(受身) + テいる(状態) + タ(過去) + にちがいない(推測) + ます(丁寧)

このような複雑な述語表現の存在は、日本語の大きな特徴の一つであるにもかかわらず、これを自動認識する技術は確立されていない。その大きな理由は、日本語学において、どの範囲を文の述語とみなすかということに関して、統一的な規定が存在しないことにある。一方、日本語の構文解析では、文節の係り受け解析が主流であり、このような複雑な述語表現の範囲を決定し、解析しようという研究はほぼないと言ってよい。

たとえば、日本語文の解析によく用いられる Juman+KNP は、「書かれていたにちがいません」を以下のように3文節に解析する。

彼の / 半生は / 小説に / 書かれていたに / ちがい / ありません

この解析結果では、どの範囲が述語であるかは不明あり、かつ、「～にちがいない」という助動詞相当表現が認定されていないため、この文が推測の意味(モダリティ)を持つことが認識できない。

2. 研究の目的

日本語の文末述語表現の形式を体系的に整理し、文末述語解析システムと文末述語生成システムの2つのシステムを実現することを目的とする。この研究の挑戦的意義は、(1) 文末の述語形式全体を体系的に整理すること、および、(2) それに基づき解析と生成の両方のシステムを実現し、日本語文法の新しいあり方を示す点にある。日本語では、文末に文法機能を担う多くの要素が出現し、「書かれていたにちがいません = 書く + れる(受身) + テいる(状態) + タ(過去) + にちがいない(推測) + ます(丁寧)」のように複雑な述語形式を形成する。このような複雑な述語形式の存在は、日本語の大きな特徴であるにもかかわらず、これを自動認識する技術は確立されていない。本研究では、文末の述語形式を上記の例のような構成要素に分割し、それぞれの構成要素の役割を同定する解析システム、および、構成要素から複雑な述語形式を生成するシステムを実現する。

3. 研究の方法

文末述語表現を、核となる述語に付加要素として接尾辞、助動詞、助動詞的に働く複合辞、終助詞等が付加されたものと捉える。これらの構成要素を列挙し、その文法的振舞いを規定すれば、文末述語形式は機械処理が可能な形で明確に定まる。そのために必要な体系的整理を、以下の4ステップで進める。

(1) 構成要素の列挙

「～にちがいない」や「～ずにはいられない」などの助動詞相当表現(複合辞)の列挙が大きな課題となる。このような複合辞は、少なくとも数百種類存在し、モダリティ(文によって表された事態に対する書き手や話し手の判断や認識)を表出する重要な役割を担っている。さらに、核となる述語となりうる述語慣用句(「気になる」や「二の足を踏む」など)の列挙も必要である。

(2) 活用型の付与

終助詞を除くほとんどの構成要素は活用する。それぞれの要素がどのように活用するかを類型化して、その情報を活用型として付与する必要があるが、形態素解析用の体系では不十分であるため、これを拡張する必要がある。たとえば、形態素解析用の体系では「詮無い」も「限りない」もイ形容詞型の活用と定義されるが、丁寧体の形式が異なる(「詮無いです」、「限りありません」)ため、活用型を分ける必要がある。

(3) 接続型の付与

各構成要素に対し、それがどのような構成要素のどのような形式(活用形)の後ろに接続できるかを規定した情報を接続情報と呼び、これを類型化したものを接続型と呼ぶ。たとえば、助動詞「だろう」は、一般に述語の終止形・タ形に接続するが、判定詞型の終止形に接続する場合は、「本だろう、明確だろう」のように「だ」が省略される。助動詞「べきだ」は、動詞の終止形のみ接続し(「書くべきだ」)、タ形には接続できない。このように、構成要素の接続には、個別の現象が多い。これらを網羅的に整理する必要がある。

(4) ID付与

語形、活用型、接続型の3種類の情報に基づいて、構成要素の同一性を定義し、IDを付与する。たとえば、名詞に接続する「です」(「本です」)と形容詞に接続する「です」(「寒いです」)

は、活用型が異なる（後者はタ形を取らない）ので異なる ID を付与する。

4. 研究成果

(1) 核となる述語に付加される要素の列挙

助詞 137 語、判定詞 1 語、助動詞 113 語、述語接尾辞 74 語、複合辞 301 語を列挙した。そのそれぞれに、ID と語形を定義し、活用する語には活用型を、活用する語に接続する語には接続型を定義した。活用型は表 1 に示す 57 種類を定義し、活用形の総数は 54 種類となった。一方、接続型は表 2 に示す 123 種類を定義した。付加要素はすべて HaoriBricks3 のブリック（文を組み立てる部品）として定義されており、これらを組み合わせることによって、複雑かつ多様な文末述語形式を容易に生成できるようになった。

表 1 活用型一覧

母音動詞型	イ形容詞型	判定詞型ナ形容詞	特殊型無活用
母音動詞型いる	イ形容詞型ナ可能	判定詞型ナ形容詞語幹	特殊型助動詞です
母音動詞型くれる	イ形容詞型ない	判定詞型ナ形容詞ノ	特殊型接尾辞です
子音動詞型カ行	イ形容詞型ない複合	判定詞型ナ形容詞ノ語幹	特殊型ます
子音動詞型ガ行	イ形容詞型接尾辞ない	判定詞型ナ形容詞同じだ	特殊型ません
子音動詞型カ行促音便	イ形容詞型よい	判定詞型ナ形容詞語幹同じ	特殊型ませぬ
子音動詞型カ行促音便ゆく		判定詞型だ	特殊型だろう
子音動詞型サ行		判定詞型助動詞	特殊型やろう
子音動詞型サ行らす		判定詞型助動詞ノ	特殊型でしょう
子音動詞型タ行		判定詞型ことだ	特殊型っす
子音動詞型ナ行		判定詞型伝聞そうだ	特殊型っしょ
子音動詞型バ行		判定詞型接尾辞や	特殊型うる
子音動詞型マ行			特殊型ざる
子音動詞型ラ行			特殊型し
子音動詞型ラ行イ			特殊型なり
子音動詞型ラ行やがる			特殊型ぬ
子音動詞型ワ行			特殊型べし
子音動詞型ワ行文語音便			
サ変動詞型			
ザ変動詞型			
カ変動詞型くる			
カ変動詞型来る			

(2) 音変化表現の文法における扱いの決定

話しことばでは、「書いてしまう」が「書いちゃう」のように変化する場合がある。これを音変化表現とよぶ。この音変化表現を扱う方法をして、(a) 新たな内容語として扱う（例：「すっごい」）、(b) 音変化コマンドで扱う（例：「ばっちし」）、(c) 新たな活用形として扱う（例：食べにゃあかん）、(d) 新たな機能語として扱う（例：「書いたげる」）、(e) テンス有標として扱う（例：「読んだあ」）の 5 つの方法を設定し、既存研究で示されいた 137 種類中の 128 種類のそれぞれに対し、どの方法で扱うかを定めた。これにより、音変化表現を含む文末述語形式のほとんどが、HaoriBrick3 で生成できるようになった。

(3) 構成要素から複雑な述語形式を生成するシステムの実現

上記に示した構成要素の列挙と HaoriBricks3 のブリック実装により、構成要素から複雑な述語形式を生成するシステムが実現され、本研究の目的の半分が達成された。

(4) 述語形式を構成要素に分割してそれぞれの要素の役割を同定する解析システムの実現

解析システム Panzer を改良・拡張し、話し言葉対応版 Panzer/U を実装した。Panzer は与えられた文から文末述語を切り出し、それを構成要素に分割するシステムである。さらに、その

発展版である Panzer2 では、処理のモジュール化を進め、複数の形態素解析システムに対応できるアーキテクチャを採用した。同時に、処理過程のステップ化を進め、出力をどのレベルまで正規化するかを制御できる方式とし、これに対応するために HaoriBrick3 に新たなブリックを追加した。Panzer2 は動作確認には至ったが完成には至らなかったため、今後も改良を継続する予定である。

表 2 接続型一覧

イ形容詞型語幹形接続	接続助詞に連用接続	テ助動詞接続	副助詞挿入可能接続
イ形容詞型連体修飾接続	接続助詞まで接続	動詞型意志形接続	文末形接続
イ形容詞型連用形接続	接続助詞まま接続	動詞型已然形接続	並列節連用接続
イ形容詞型連用修飾接続	接尾辞ありません接続	動詞型終止形接続	ボイス接尾辞子音動詞接続
已然形接続	接尾辞う接続	動詞型終夕形接続	ボイス接尾辞母音動詞接続
敬語接尾辞接続	接尾辞がる接続	動詞型夕形接続	ボイス接尾辞る接続
敬語接尾辞なさる接続	接尾辞しめる接続	動詞型夕接形接続	ボイス接尾辞れる接続
終止形接続	接尾辞する接続	動詞型チャ形接続	無条件接続
終止助動詞ことだ接続	接尾辞っばい接続	動詞型テ形接続	名詞接続
終止助動詞でない接続	接尾辞です接続	動詞型未然形接続	連体修飾接続
終止助動詞なり接続	接尾辞ない語構成接続	動詞型連体修飾接続	連体修飾ダ消失可能接続
終止助動詞べきだ接続	接尾辞ない接続	動詞型連用形接続	連体助動詞引用接続
終止助動詞まい接続	接尾辞ぬ接続	取立助動詞ある接続	連体助動詞ごとし接続
終助詞い接続	接尾辞ます接続	取立助動詞いらっしゃる接続	連体助動詞接続
終助詞接続	接尾辞ません接続	取立助動詞いる接続	連体助動詞ナ引用接続
終助詞ダ消失可能接続	接尾辞めだ接続	取立助動詞ござる接続	連体助動詞ナ接続
終助詞ダ消失接続	接尾辞ん接続	取立助動詞する接続	連体助動詞ナノ接続
終助詞の接続	節末の接続	取立助動詞ない接続	連体助動詞ノ接続
終夕形接続	節末もの修飾接続	取立助動詞なる接続	連体助動詞のだ接続
終夕形ダ消失接続	夕系条件形接続	取立助動詞二ござる接続	連体助動詞ばかりだ接続
終夕助動詞接続	夕形接続	取立助動詞二ない接続	連体接続
終夕助動詞ダ消失接続	タリ助動詞接続	ない接続	連用形接続
述語接続	チャ形接続	内容語接続	連用修飾接続
述語だ省略可能接続	丁寧助動詞接続	ナ形容詞連体修飾接続	連用助詞接続
条件音便形接続	丁寧助動詞っす接続	ナ形容詞連用修飾接続	連用接尾辞すぎる接続
条件形接続	丁寧助動詞でしょう接続	ぬ接続	連用接尾辞接続
条件節しか接続	丁寧助動詞です接続	判定詞型消失夕形接続	連用接尾辞そうだ接続
助動詞ろう接続	テ形接続	判定詞型チャ形接続	連用節連用接続
接続助詞接続	テ縮約接尾辞接続	判定詞型テ形接続	連用チャ形接続
接続助詞と接続	テ助動詞いい接続	判定詞接続	連用テ形接続
接続助詞ながら接続	テ助動詞ござる接続	副助詞述語接続	

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計3件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 佐藤理史, 柳将吾, 夏目和子
2. 発表標題 HaoriBricks3 による音変化表現の生成
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 柳将吾, 佐藤理史, 夏目和子, 宮田玲, 小川浩平
2. 発表標題 話し言葉を対象とした文末解析と表現文型の同定
3. 学会等名 言語処理学会第27回年次大会
4. 発表年 2021年

1. 発表者名 佐藤理史
2. 発表標題 多様な形式の発話文の合成を可能とする辞書とツール
3. 学会等名 シンポジウム「日常会話コーパス」VII
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------