

構文解析の観点からの対話文に特有な現象の分類

3B-6

中野幹生 島津 明 小暮 潔

NTT 基礎研究所

1. はじめに

話し言葉により人間と自然に会話できる計算機を作るためには、自然な対話文を解析して、意味を求める方法を構築する必要がある。そのためには、まず、対話文が書き言葉の文とどのように異なるかを調べる必要がある。そこで、問い合わせや仕事の依頼の電話対話、野球中継、ラジオ電話相談の書き起こしデータを分析し、対話文特有の現象の調査を行っている³⁾。これらの対話は、あらかじめ状況を設定して行われたものではなく、自然に行われた対話である。

次に、発見された現象を分類整理し、個々の現象について対処法を考える必要がある。この時、どのように現象を分類するかが問題となる。現象の分類と対処法の構築は分離して行えるものではなく、処理の枠組をどうとるかに応じて分類が決まる。話し言葉には、一見規則が見られないように見えるが、よく観察すると、一定の規則性が見られる。我々は、書き言葉の文法を構築しなおすことによって話し言葉を扱う立場で分析を進めている。竹沢ら⁵⁾は、状況を設定して行われた対話の調査⁴⁾を基にして、対話特有の現象を扱う音声認識用の文法を構築しているが、網羅的であるかどうか、また、どのように意味関係を求めるのか明かでない。

本稿では、文法枠組として、統語構造と意味表現の関係が明確な単一化文法を用い、この観点からの対話文特有の現象の整理を試みる。文法の拡張法として、語彙項目の追加、素性制約の緩和、句構造規則の追加、文の概念の拡張の四つを考え、これらの拡張を書き言葉の文法に施すことにより対処できる現象について述べる。今までに調べられている現象^{3), 4)}は、これらの四つに分類が可能である。

2. 分類の基とする文法

本稿でベースとして用いる文法は、日本語句構造文法(JPSG)¹⁾をベースにしたもの³⁾である。ベース文法で用いる句構造規則は、下位範疇化規則(必須格の名詞句が動詞句に係る場合などに用いられる規則)、隣接規則(助詞が名詞句や動詞句に、助動詞が動詞句に後続する現象を記述する規則)、付加規則(修飾の規則)、および、連体修飾のための規則、連用修飾のための規則等がある。この文法を拡張して話し言葉を扱う方法を以下に述べる。

A Clasification of Phenomena Peculiar to Dialogue Sentences from the Viewpoint of Syntactic Analysis
Mikio Nakano, Akira Shimazu, and Kiyoshi Kogure
NTT Basic Research Laboratories

3. 対話文に特有な現象の分類

語彙項目の追加により扱える現象

語彙項目を増やすことによって扱える現象を考える。対話文に特有な単語の存在が、この分類に入る。

対話文に特有な単語を、どのような統語カテゴリに入るかで下位分類する。なお、本稿では統語カテゴリは、同じ統語的素性制約をもつ単語の集まりという意味で用いる。

- 次に来る単語に係るが統語的情報を変えないもの³⁾

「えっと」「あの一」などのつなぎ語がこのカテゴリに入る。つなぎ語は、修飾の規則により、次の語にかけてまとめて扱う。つなぎ語から得られる意味的情報は、談話解析部で扱う。

例: 正門のところに、えっと、守衛さんのいるところがあるので

- 名詞句(+助詞)、動詞句(+助詞)に後接し、統語意味情報を変えないもの³⁾

「ですね」「ね」などの間投助詞(間投詞的用法の終助詞ともいう)がこのカテゴリに入る。これらの語は、句のあとに挿入される。

例: そこからですね、研究所に来るまでの、...

- 書き言葉に存在するカテゴリに収まるもの

話し言葉に特有であるが、統語的に、書き言葉の単語と同じカテゴリに分類できるものがある。「んです(のです)」などの書き言葉の語句と音韻的に異なる語句は、対応する書き言葉の語句と同じカテゴリに入る。また、話し言葉特有の助詞も相当する書き言葉の助詞と同じカテゴリに入る。例えば、「って」は、引用をあらわす助詞「と」、同格を表す「という」、提題助詞「は」と同じカテゴリに入れることができる⁶⁾。また敬語や対話にしか現われない終助詞もここに分類することができる。

素性制約の緩和により扱える現象

フィルタとして用いられている素性の制約を緩和することにより扱える現象を考える。

- 意味素性制約(選択制限)の緩和

これにより、などの換喩や隠喩が扱える。

例: 東京支社が電話してきた

また、「所用時間が、だいたい30分くらいで行けます。」のようなねじれ文、すなわち、最初の方と後の方がうまくつながらない文も解析可能になるが、意味関係がうまく求められない。このような場合は、後に述べる文の概念の

拡張により、二つの発話からなっているとし、意味関係は、談話解析部で扱う。

規則の追加により扱える現象

句構造規則を追加することによって扱える現象をここに分類する。

● 名詞句を助詞句とみなす規則

PP → NP

の形の規則により、名詞句を助詞のついた句として扱うことができる。これによって扱える現象は、助詞の省略である。

例: あの、また話戻しますけども、...

● 助詞を無視するための規則

XP → X P

(X は任意の単語。XP は X と同じ統語意味情報をもつ。)

という規則により、助詞を無視することができる。名詞句の後の助詞が無視されるため、上記の名詞句を助詞句とみなす規則を用いて、格助詞の誤用を扱うことができる。

● 名詞句を判定詞句 (名詞句 + だ)²⁾ とみなす規則

判定詞句 → 名詞句

の形の規則により、判定詞「だ」の省略を扱うことができる。

例: 時間は 15 分くらい (だ)。

● 繰り返しを扱う規則

$X'_2 \rightarrow X_1 (Y) X_2$

(X_1 と X_2 は統語的、意味的または音韻的に類似、Y は「いや」「じゃない」などの打ち消しの語で、存在してもしなくてもよい。)

の規則で、統語的、意味的、音韻的に似た語句の繰り返しを扱うことができる。 X'_2 と X_2 の統語情報は同じである。このような規則で扱える現象としては、言い直し、言い替え (同格)、言い淀み、繰り返しがある。例えば、繰り返しは、 X_1 と X_2 が音韻的に同じである。

例: それやっちゃうと、全部、全部 かわっちゃうでしょ?

また、次の言い直しは、 X_1 と X_2 が同じ統語カテゴリで、かつ、意味的にも同じ数量のカテゴリに入る。

例: 一つの、一人の 対話項目にしてしまうと云ったでしょ。

● 語順の変更を扱う規則

付加規則、下位範疇化規則の右辺 (子カテゴリの並び) の順序を変更した規則を用いることにより扱える現象として、倒置がある。

例: 左揃えだよ確か。

● その他特殊な規則の追加

その他特殊な規則を追加することによって扱える現象としては、数え上げや綴りを説明する句の存在がある。

例: あのスターのところに 1,2,3 で番号つけて....

文の概念の拡張により扱える現象

対話では、書き言葉ほど文の境界が明確でなく、また、必要な述語が必ずしも存在するわけではない。したがって、書き言葉の文法では文と認められない語句を発話の単位と認めることにより、断片的な発話を扱うことができる。

● 名詞句 (+ 助詞) を単位とすることで扱える現象

駅の窓口で「横浜」と言って横浜までの切符を買うときなどの名詞句文、「鈴木さん」と人に呼びかける呼びかけ、「値段は?」などの名詞句 + 助詞だけが存在し述部がない現象などが扱える。

● 動詞句 (+ 助詞) を単位とすることで扱える現象

接続助詞での中止が扱える。

例: これから家を出ますので。

● 感動詞を単位とすることで扱える現象

「あっ」「うーん」などの感動詞一語からなる孤立した発話が扱える。

● 接続詞を単位とすることで扱える現象

「で」、「それですね」といった、発話と発話を結ぶ接続詞が扱える。

4. おわりに

対話文特有の現象を、文法をどのように拡張して扱うかという観点から分類できることを示した。この拡張した文法を用いると、前処理が必要でないため、入力を逐次的に解析することが可能になる。

参 考 文 献

- 1) T. Gunji. *Japanese Phrase Structure Grammar*. Reidel, Dordrecht, 1987.
- 2) 益岡, 田窪. 基礎日本語文法 - 改定版 -. くろしお出版, 1992.
- 3) 中野, 島津, 小暮. 対話文の文法構築に向けた分析. 情報処理学会自然言語処理研究会報告, (107):35-42, 1995.
- 4) 竹沢, 田代, 森元. 音声言語データベースを用いた自然発話の言語現象の調査. 人工知能学会研究会資料 SIG-SLUD-9403-3, pp. 13-20, 1995.
- 5) 竹沢, 田代, 森元. 自然言語の言語現象と音声認識用文法. 情報処理学会研究報告 SLP-6, pp. 27-34, 1995.
- 6) 国立国語研究所 (編). 現代日本語の助詞・助動詞. 秀英出版, 1951.