コ・ソ系指示詞の先行詞の同定

3R－9

古閑 敏真　呉 浩東　古郡 廷治
電気通信大学　情報工学科

1．はじめに

文はある関係をもって前後の文と連結している。その関係の一つに照応がある。照応関係は、文（章）中の２つの要素間に、指示と被指示の関係をもって成立する。

本研究では、コ・ソ系の照応詞の先行詞を同定する手法を提案し、その手法に従って行った実験結果を報告する。

2．コ・ソ系の照応詞

コ・ソ系の照応詞には、「これ」「それ」「これら」「それら」「ここ」「そこ」のような代名詞や、「この」「その」のような連体詞、「こんな」「そんな」のような形容動詞、「こう」「そう」などの副詞がある。連体詞や形容動詞の照応詞は、その直後に名詞を伴って使われる。

3．先行詞の同定法

ゼロ照応や代名詞照応の先行詞を同定する手法は、大別すると３種類ある。それらは、談話理論による方法、文法理論による方法、意味理論による方法である。

談話理論による手法は、Grosz らのセンターリング理論に基づくき、文中において話題の中心となる焦点をみつけることを通し、先行詞を同定する[1]。

文法(構文)理論を用いる手法は、分析対象となる文の表層構造をパターン化し、そのタイプから先行詞を同定する[2]。

意味理論を用いる手法は、辞書を使い、分析対象となる文の中の要素(単語)が持つ意味素性や、要素がとりうる選択制限情報を抽出して先行詞を同定する[3]。

コ・ソ系照応詞の先行詞の同定方法には、主に談話理論による方法と意味理論による方法がある。

(1) 机の上に、一冊の本がある。
　　それは、素晴らしい本だ。

(1)では、照応詞「それ」に対する先行詞の候補として「机」と「本」がある。談話理論によれば、「本」が前文の焦点となっているので、照応詞「それ」を先行詞は「本」である。

(2) 太郎は次郎と飛行機を作った。
　　しかし、それは簡単に壊れた。

(2)では、照応詞「それ」に対する先行詞の候補として、「太郎」「次郎」「飛行機」である。意味理論によれば、選択制限情報から「壊れる」の主体は[物体]である。先行詞の候補で意味素性に[物体]を持つのは「飛行機」だけである。従って、照応詞「それ」の先行詞は「飛行機」となる。

コ・ソ系照応詞のなかでも、連体詞や形容動詞のように、後に名詞が続く場合、後続の名詞の情報を利用して先行詞を同定する手法もある[3]。

(3) 雨が降ってきたので、花子は傘をひろげた。
　　その傘は、ピンクの糸で刺繍がしてあった。

(4) コインに、鳩が刻まれている。
　　この鳥は平和の象徴である。

(3)では、照応詞「その」に続く名詞と同一の名詞「傘」が、先行詞である。(4)では照応詞「この」に続く名詞「鳥」の下位語にあたる名詞「鳩」が先行詞となる。

4．コーパスをもとにした同定法

既存の方法では、先行詞を同定できない場合があ

Determining the Antecedent of Demonstrative Adjectives *Kono* and *Sono*
Toshimasa Koga, Haodong Wu and Teiji Furugori
Department of Computer Science, The University of Electro-Communications

る。

(5) おそらく地域紛争は簡単には収まらず、今後も解決が最も難しい課題の一つとして国際社会に重くのしかかるだろう。その難しさの一因は「民族」という概念がはらむ両面性にある。

(5)では、照応詞「その」の先行詞の候補として「地域紛争」「解決」「課題」「国際社会」の４つの要素がある。談話理論を使うと「地域紛争」が先行詞となろう。ここでは、意味理論は適用不可能である。

本研究では、連体詞の照応詞を対象にし、その先行詞をEDRコーパスから得た統計情報量によって同定する手法を提案する。

5．アルゴリズム

連体詞系の照応詞の先行詞同定に使うアルゴリズムは以下のとおりである。

1．「連体詞＋名詞x」が出現したら、その個所から１文さかのぼり、その間に存在する名詞を取り出す。これを先行詞の候補とする。

2．先行詞の候補の類義語を『日本語語彙体系』[4]から抽出する。これらの類義語と先行詞の候補を合わせた集合yを先行詞候補群とする。

3．先行詞候補群と名詞xのコーパス内での連想度 I(x,y) を式(1)で求め、その中で一番連想度が高い値をその群の値とする。

4．一番高い値をもつ群の先行詞候補を先行詞として同定する。

$$I(x,y) = log_2 \frac{P(x,y)}{P(x)P(y)} \qquad (1)$$

(1)でのxとyは名詞であり、P(x)とP(y)はコーパス内での名詞xとyの出現確率である。またP(x,y)は、同一文内で名詞xとyの共起確率である。

6．実験例

５節に述べたアルゴリズムを使って、(5)での「その難しさ」の先行詞の同定結果を示す。

照応詞「その」の先行詞候補には、「紛争」「解決」「課題」「社会」の４単語がある。それぞれの類義語は「戦争、争い」「解消、決着」「問題」「世界、世間」である。これらと「難し(さ)」との連想度を計算すると、図１のようになる。この図から、先行詞として「解決」を得る。実験結果をもう一例示す。

(6) 自民党の前衆院議員が政治資金団体を経由して、総合建設会社から違法な献金を受けていたことが明らかになった。その額は四千万円と巨額だ。

(6)での照応詞「その」の先行詞は、前例と同じプロセスを経て「献金」と同定される。

	難し(さ)
紛争	0.000000
戦争	0.051167
争い	0.000000

	難し(さ)
解決	1.048108
解消	0.772575
決着	0.648647

	難し(さ)
課題	0.887529
問題	0.762590

	難し(さ)
社会	0.000000
世界	0.147406
世間	0.000000

図１「難し(さ)」との連想度

7．今後の課題

毎日新聞の社説30編を対象として、連体詞系の照応詞の先行詞同定実験を試みた結果では、70%の正答率を得ている。アルゴリズムに、連想度に重みを付けたり、名詞x以外の要素との連想度を計算するなどの実験をすれば、さらに精度が上がることが期待される。

参考文献

[1] Megumi Kameyama, "A Property-Sharing Constraingt In Centaring", 24th Annual Meeting of Association for Computational Linguistics, pp.200-206, 1986.

[2] 中岩 浩巳, 池原 悟：「語用論的 意味論的制約を用いた日本語ゼロ代名詞の文内照応解析」 自然言語処理 Vol.3 No4 Oct. 1996

[3] 村田 真樹, 長尾 眞：「用例や表層表現を用いた日本語文章中の指示詞、代名詞、ゼロ代名詞の指示対象の推定」自然言語処理 Vol.4 No.1 Jan. 1997

[4] 池原 悟 他『日本語語彙体系』, 岩波書店, 1997