

新聞社説記事における照応現象

5F-4

柴田昌宏⁺ 田中 理⁺ 福本淳一⁺⁺⁺(株)沖テクノシステムズラボラトリ⁺⁺沖電気工業(株)総合システム研究所

1. はじめに

自然言語処理において、1文を超える処理を行うためには文章中の各文間の関係を扱うことが重要であり、これまでも多くの研究がなされている[1]。文間の関係には、照応関係、一貫性などがあるが、本稿では、このなかでも照応関係に注目し、新聞社説記事を用いて照応関係の調査を行った。

照応とは、一般に先行のコンテキスト内にある要素を後続する文の中で別の形式をもって代用する言語現象であり[2]、文内のもの、文間にまたがるものが存在する。本稿では、文間にまたがる照応について調査を行い、日本語文章において照応語や先行詞にどのようなものがあり、また、それらの関係としてどのようなものが成立するかの調査を行った。

2. 照応について

照応現象とは、コンテキスト内のある要素を、その前後にある文の中において、別の形式(同じ形式でもかまわない)で表現する現象のことを言う。あるいは、名詞句などの語句が、その前後のコンテキスト内で言及されているモノ・コトを指示参照する現象であるとも言える[2]。この場合の名詞句などの語句を照応語、照応語が指示参照するコンテキスト内の要素を先行詞、その先行詞と照応語の関係を照応関係と呼ぶ。以下では、照応語、先行詞、照応関係およびその位置関係についての分類を示す。(以下の例では□は照応語、○は先行詞を指すものとする。)

(1) 照応語の分類

a. 照応語が指示詞を含む場合

「これ/それ/あれ」または「この/その/あの」+名詞などの形式をとる。

b. 照応語が名詞句からなる場合

「名詞」または「修飾句+名詞」などの形式をとる。

c. 照応語がゼロ代名詞である場合

コンテキスト内には現れていないが、文脈や知識などにより存在が想定できるもの。

(例) 社会党は北朝鮮首脳と意思疎通をはかれる関係を維持してきた。□このパイプを生かして、今回も懸案の…。

(昭和62年9月30日付朝日新聞朝刊より)

(2) 先行詞の分類

Anaphora in Newspaper Editorials

Masahiro SHIBATA⁺, Osamu TANAKA⁺, Jun-ichi FUKUMOTO⁺⁺⁺Oki Technosystems Laboratory, Inc.⁺⁺Systems Laboratory, Oki Electric Industry Co., Ltd.

a. 先行詞が語、句である場合

b. 先行詞が節、文、あるいはその内容である場合

c. コンテキスト内に明示されてなく、背景・知識などにより推定される場合

(例) 都心から30-35キロ圏に一戸建てを買いとすると…普通のサラリーマンに買える家は、もうないというわけだ。

(昭和62年10月1日付朝日新聞朝刊より)

(3) 照応関係の分類

a. 同一名詞による照応

先行詞と照応語が同じ名詞であるもの。

b. 上位-下位関係

一方が他方の上位概念をもつもの。

c. 部分-全体関係

一方が他方の構成要素の一部を表現しているもの。

d. 限定照応

指示詞や代名詞により限定を受けた照応語によるもの。

e. ゼロ照応

(4) 照応の位置関係の分類

a. 前方照応

前方のコンテキスト内ですでに言及されているモノ・コトを参照する照応のことをいう。

b. 後方照応

前方照応とは逆の照応である。

(例) 当時の勧告は□述べている。「精神病院はごく一部を除いて…」と。

(昭和62年10月5日付朝日新聞朝刊より)

c. 照応の距離

先行詞と照応語の間にはコンテキスト内における距離が存在する。この距離の測定基準としては、文数、文節数、命題数、指示対象の候補数などが考えられる。

3. 調査内容

朝日新聞の社説記事(昭和62年8月~10月)を用い、2節の分類を基に、以下の観点から照応現象の調査を行った。図1に調査結果の記入フォーマットを示す。

(1) 照応語

a. 人称代名詞 「彼」など

b. ゼロ代名詞 (明らかに省略されていると判断できる主格のものについてのみ調査する。)

c. 指示代名詞 「これ/それ」

d. 指示連用詞 「こう/そう」+用言

e. 指示連体詞 「この/その」+体言

f. サ変動詞の語幹

g. 名詞, 名詞句

h. 代名詞的表現 「两国」、「同氏」など

照応語がある述語の格要素である場合は、その述語、格要素についても調査した。

(2) 先行詞

照応語によって指示されている語、句、文、パラグラフを選んだ。また、照応語が内容を指示参照しているときは、その内容を含む文、パラグラフを選び、背景・知識などにより推定されるものは対象外とした。

照応語と同様に、述語、格要素についても調査した。

(3) 照応関係

- 同一名詞による照応
- 上位-下位関係
- 部分-全体関係
- 指示詞、代名詞による照応
- ゼロ照応

(4) 照応の位置

照応の距離を測定する。距離は、それぞれの語が含まれている文間の距離を文数で表した。前方照応のとき正值、後方照応のとき負値をとる。同一名詞による照応などのように、先行詞の候補が複数ある場合は、距離が最短になるものを先行詞とする。

PP-SS-		PP-SS-
VV...V		VV...V
CCC : WW...W	R	CCC : WW...W
(FF...F)		(FF...F)
MM...M		

P: パラグラフ番号
S: 文番号
C: 格

V: 述語
F: 付帯語
W: 先行詞および照応語

R: 照応関係
M: コメント

図1 照応現象記入フォーマット

4. 調査結果

以上の調査方法に基づき、31編の社説記事について調査を行った(抽出データ数は1572)。

社説記事においては近い距離の照応が多く、距離が2以下の場合が半数を占めた(約52%)。また表1より、同一名詞による照応が全体の73%弱、指示詞

表1 照応関係の種類別割合

照応関係の種類	%
同一名詞による照応	72.7
部分-全体の関係	2.2
上位-下位の関係	4.0
指示詞等による照応	17.8
ゼロ照応	3.3
計	100.0

等による照応が18%弱であり、同一名詞による照応が目立った。

そのほか照応語・先行詞について、次のようなものがあった。

- ・敬称による照応 土井委員長 ← 委員長
- ・敬称の変化 土井委員長 ← 土井氏
- ・固有名詞の一部省略による照応 海上保安白書 ← 白書

5. 考察

以上の調査結果から、ある照応語に対する先行詞を探すためには、次のことが考えられる。

- (1) 指示詞は、語、句、文など様々なものを指し得るが、近いものを指示参照する機会が多い(約90%)。従って、指示詞の先行詞の探索は近いものについて深く行うのが望ましいと考えられる。
- (2) 名詞句は、半数程度しか近いものを指しておらず、近いものを重点的に探す必要はないと考えられる。また、名詞に対してはシソーラス的な知識や、「アメリカ訪問 ← 訪米」のような照応から、言い換えや、省略形などの知識も扱う必要があると考えられる。

先行詞は、照応語から距離的に近いコンテキスト内に重点をおいて探索すると仮定すると、先行詞の候補が複数あるときは、より距離が短いほうを選択することになる。しかし、例えば、あることがら「P」について2文A、Bで説明をし、文Aの説明文Cが2文A、B間にある場合、文Bの照応語「P」の先行詞は、文Cよりも文Aにあるとしたほうが自然である。こうするためには、文章の構造をとらえたうえで照応を扱う必要がある。

6. おわりに

今回の調査から、新聞社説記事における照応現象の傾向を捕えることができた。また、派生したいくつかの問題点を示し、考察を行った。

今後は、さらに詳細な調査を行い、照応関係の抽出方法について考察する予定である。また、現在PSI上に支援ツールを開発しており、照応現象データの電子化も並行して進めている。

本研究はICOT受託研究の一環として行われたものであり、研究の機会を与えて下さったICOT第2研究室内田室長、吉岡室長代理に感謝致します。

【参考文献】

- [1]大特集：自然言語理解，情報処理，vol.30, No.10 (1989)。
- [2]上野田 鶴子 他：日本語と外国語との照応現象に関する対照研究，国立国語研究所報告79，研究報告集5，pp.199-275，秀英出版 (1984)。
- [3]久野 嘩：談話の文法，大修館 (1978)。