

日本語文章の照応解決における談話的制約の有効性の検証

4R-4

藤澤伸二*

増山 繁*

内藤昭三**

fujisawa@smlab.tutkie.tut.ac.jp, masuyama@tutkie.tut.ac.jp naito@atom.ntt.jp

*豊橋技術科学大学知識情報工学系

**NTT基礎研究所

1 はじめに

照応とは、一般には、文章中の2つの言語形式が同一の対象を指示している場合に、その2つの言語形式の関係を意味する現象であり[2]、代名詞、指示詞などの照応語とその先行詞の組によって表されることが多い。日本語文章においては、冗長度を下げることが目的にしばしば用いられる。照応に関しては、照応の定性的評価、照応を用いた意味解析や談話解析、などの研究が行なわれている。本稿では、先に提唱した、照応に関する4つの談話的制約[1]が実際の科学技術解説文において照応語の先行詞を決定する上でどの程度有効であるかを検証するため、語の素性の一致制約に、さらにこれら4つの制約を加えれば、どの程度照応解析結果が改善されるかについて計算機実験を行ない、その結果について考察する。

2 照応に関する制約

文献[1]で提唱し、検証した照応・省略に関する8つの仮説(制約)のうち、照応に関する4つの制約は以下の通りである。なお、単文、複文、重文、最上位の節の定義は[1]に従う。

(仮説1) 照応は段落に跨ることがない。

(仮説2) 照応は交差することがない。

(仮説3) 複文および重文の先頭を除く節中の照応語の先行詞は、その節および、その節の上位の節には含まれない。

(仮説4) 単文中の照応語、および、重文の先頭節で、しかも埋め込み節でない節中の照応語に対する先行詞は、それより前の文(段落を跨ってもよい)にある。

3 照応解析実験

3.1 照応解析方法

動詞・形容詞辞書としてIPALの動詞辞書、形容詞

An Inspection on Effect of Discourse Constraints for Anaphora Resolution in Japanese Sentences

Shinji FUJISAWA*, Shigeru MASUYAMA*, Shozo NAITO**

*Toyoashi Univ. of Tech., **NTT Basic Research Lab.

辞書[3]を、また、名詞辞書として角川類語新辞典[4]を用い、日経サイエンスの記事5編の冒頭から7段落を調査対象として、照応解析実験を行なった。

照応解析方法としては、あらかじめ作成した照応語辞書を利用して照応語の認定を行ない、その照応語の文中での格に対する素性制約をIPAL辞書[3]から抽出し、その素性と同一の素性を持つ語を照応語の先行詞候補とする。照応語が複合名詞の修飾部となる場合(「その名前」など)が数多く出現したため、以下の2つの方法で、その照応語のとり得る素性の推定を行なった。

方法1 複合名詞に付されている格のとり得る素性

方法2 複合名詞の末尾語の持つ素性

その上で、これらの候補に4つの制約を適用する。

3.2 実験結果

結果を表1から表4に示す。表1、3は、すべての照応語の先行詞の候補のうち、それぞれの制約を満たすものの重複を含めた出現回数、および、照応語1語あたりの平均候補数を示したものである。但し、「制約なし」の場合は、必須格の持つ素性と同一の素性を持ち、照応語より前に出現する語の総数である。表1、3中の「未登録語」とは、今回使用した角川類語新辞典[4]に記載されていない語を先行詞とする照応である。また、表中の「名詞類照応」は、照応語の先行詞が名詞である照応、「述部類照応」は、「名詞類照応」「未登録語」を除く照応である。「述部類照応」には文や文章を先行詞とする照応、テキスト外照応が含まれる。それぞれの出現回数は、方法1で「名詞類照応」が31回、「述部類照応」が7回、「未登録語」が2回、方法2で「名詞類照応」が33回、「述部類照応」が6回、「未登録語」が3回であった。

表2、4は、「名詞類照応」の全出現回数(方法1で31回、方法2で33回)における照応語先行詞の候補の中に正しい先行詞が含まれている回数の割合(正解率)である。先行詞が正しいかどうかの判定は人手によって行なった。

方法 1

	名詞類照応	述部類照応	未登録語
制約なし	2425(78.2)	659(94.1)	216(108.0)
制約 1	361(11.6)	93(13.3)	9(4.5)
制約 2	504(16.3)	130(18.6)	25(12.5)
制約 3	2419(78.0)	659(94.1)	216(108.0)
制約 4	2425(78.2)	659(94.1)	216(108.0)
全制約	233(7.5)	43(6.1)	5(2.5)

表 1. 先行詞の候補の総数と 1 語あたりの平均先行詞数 (語)

制約なし	61.3
制約 1	35.5
制約 2	54.8
制約 3	61.3
制約 4	61.3
全制約	35.5

表 2. 名詞類照応の先行詞選択の正解率 (%)

方法 2

	名詞類照応	述部類照応	未登録語
制約なし	2433(73.7)	573(95.5)	219(73.0)
制約 1	367(11.1)	73(12.2)	10(3.3)
制約 2	499(15.1)	122(20.3)	26(8.7)
制約 3	2422(73.4)	573(95.5)	212(70.7)
制約 4	2433(73.3)	573(95.5)	219(73.0)
全制約	229(6.9)	55(9.2)	6(2.0)

表 3. 先行詞の候補の総数と 1 語あたりの平均先行詞数 (語)

制約なし	69.7
制約 1	45.5
制約 2	57.6
制約 3	69.7
制約 4	69.7
全制約	42.4

表 4. 名詞類照応の先行詞選択の正解率 (%)

4 考察

1. 制約 1、2 はすべての文を対象とし、しかも、照応語の先行詞の探索範囲を限定している制約なので、先行詞の候補をかなり削減できる (表 1、表 3 参照)。
2. 制約 3、4 が適用可能な文の出現は、実際の文章であまり多くなく、また、これらの制約により削

減できる先行詞の探索範囲は現在処理中の文のみなので、先行詞の候補数は制約不使用時とあまり変わらない (表 1、表 3 参照)。

3. 素性一致制約のみを用いて照応解決を行なった場合でも、正解率が 100% にならない理由は、動詞・形容詞の文型選択誤り、IPAL 動詞・形容詞辞書にその用言に対する素性、あるいは、文型そのものが未登録であることなどである (表 2、表 4 参照)。
4. 複合名詞中の照応語のとり素性を 2 つの方法で推定し、その素性を用いて照応解析を行なった結果を比較すると、照応語 1 語あたりの先行詞の候補数に大差はないが、正解率 (先行詞の候補の中に正しい先行詞が含まれる割合) は方法 2 の方が高い。この理由は、1 度文章中に出現した語を限定指示するために照応語を含む複合名詞が使われる場合が多いからである (表 1 ~ 表 4 参照)。
5. 複合名詞中の照応語のとり素性を推定する 2 つの方法によって、それぞれの照応の出現回数が異なる。この理由としては、推定法によって照応語の素性が異なり、それを基にして行なう文型選択回数が異なる文型を選択したためである。

5 おわりに

本稿では、照応に関する 4 つの談話的制約を用いて照応解決実験を行なうことで、これらの談話的制約が照応解決に有効であることを検証した。

今後の課題は、照応解決の精度をさらに改善するために、話題 (照応語や省略語の先行詞となり得る語) を併用すること、新しい制約を探すこと、などである。

謝辞

「角川類語新辞典」[4] を計算機可読の形で提供して頂き、その使用許可を頂いた (株) 角川書店に深謝する。

参考文献

- [1] 藤澤、増山、内藤：日本語文章における照応・省略現象の基本的検討、情報処理学会論文誌、vol.34, No.9, pp.1909-1918 (1993)
- [2] 国立国語研究所、日本語の指示詞 (1981)
- [3] 情報処理振興事業協会技術センター：計算機用日本語基本動詞辞書 I P A L 解説編、辞書編 (Basic Verbs, 1987)、(Basic Adjectives, 1990)
- [4] 大野、浜西：類語新辞典、角川書店 (1981)