

日本語文の省略補完に関する談話的制約の有効性の検証

8B-5

藤澤伸二\*

増山 繁\*

内藤昭三\*\*

fujisawa@toki.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

省略は、聞き手にとって自明の情報を省く言語現象である[2]が、ここでは、省略をIPALの辞書[3]記述中の必須格要素の省略と定義する。省略に関する研究としては、省略の定性的評価をしているもの[5]や、省略を用いて意味解析や談話解析を行なったという報告[7][8]、さらに、省略語の補完(復元)に関するもの[9][10][11]がある。省略語の補完(復元)に関する研究は、数文以内の談話を対象としたものがほとんどであり、複数段落よりなる文章を対象にしたものはあまりない。本稿では、先に提唱した[1]、省略に関する4つの制約が実際の科学技術解説文において省略語の復元を行なう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、どの程度省略語補完の結果が改善されるかを実際に計算機実験により調べ、その結果について考察する。

2 省略に関する制約

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、省略に関する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。

- (制約1) 省略は段落間を跨らない。すなわち、省略の先行詞は、省略語と同一段落中にある。
- (制約2) 省略は交差しない。すなわち、省略と先行詞の組の間には、交差が生じない。
- (制約3) 最上位の節を除く各節において、省略の先行詞はその上位の節に含まれる。
- (制約4) 単文中の省略、および、重文の先頭節で、しかも埋め込み節でない節中の省略に対する先行詞は、それより前の文中にある。

3 省略語の補完方法

動詞・形容詞辞書としてIPALの動詞辞書、形容詞辞書[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編のそれぞれに対し、最初の

7段落を調査対象として、省略語の補完を行なった。

補完方法としては、IPALの辞書[3]に記載されている必須格で現在処理している文にはないものを省略語とし、省略位置より前に出現した語の中からその必須格の持つ素性と同一の素性を持つ語を省略語の先行詞の候補とする。その上で、これらの候補に4つの制約を適用する。

4 結果

結果を表1から表3に示す。表1は、すべての省略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、省略語より前に出現する語の総数である。表1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(省略)語である。また、「話者等を指す省略語」とは、「わたし(筆者、話者)は」や、「あなたがた(読み手、聴者)は」を指す省略語であり、その先行詞は、ほとんど文章に現れない。それぞれの出現回数は、「省略語」が56回、「未登録語」が5回、「話者等を指す省略語」が68回であった。以下、「省略語」とは、「話者等を指す省略語」、「未登録語」を除く省略語を指す。

表2は、「省略語」の全出現回数(56回)における省略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。

表3は、制約不使用時の候補数に対する制約使用時の候補数の割合を表したものである。制約を用いることによる候補数削減効果を示している。

	省略語	未登録語	話者等を指す省略語
制約なし	4218	568	3990
制約1	683	51	637
制約2	238	23	251
制約3	2895	436	3731
制約4	4211	568	3979
全制約	207	18	156

表1. 先行詞の候補の総数(語)

制約なし	8 6
制約 1	8 0
制約 2	4 3
制約 3	7 1
制約 4	8 4
全制約	4 5

表 2. 省略語の先行詞選択の正解率 (%)

	省略語	未登録語	話者等を指す省略語
制約 1	1 6	9	1 6
制約 2	6	4	6
制約 3	6 9	7 7	9 4
制約 4	1 0 0	1 0 0	1 0 0
全制約	5	3	4

表 3. 制約不使用時に対する候補数の割合 (%)

## 5 考察

以上の結果から次のことが分かった。

1. 制約 1、2 はすべての文を対象とし、しかも、省略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。
2. 制約 3、4 が適用可能な複文、重文の出現は、実際の文章であまり多くないので、先行詞候補の削減効果は小さい。制約 4 の適用範囲は現在処理中の文章のみなので、先行詞候補の削減効果は特に小さく、先行詞の候補数は、制約不使用時とほとんど変わらない。
3. 制約 2 を用いて省略語の先行詞の候補を絞り込んだ場合、残された候補の中に、省略語の正しい先行詞が含まれる比率が低下する。その理由は以下の通りである。制約 2 を単独で使用すると、他の制約を併用した場合に比べ、先行詞の探索範囲が広くなり、候補として多くの先行詞以外の語を選んでしまう。ところが、現在の制約の使い方では、それらの語はすべて省略の非交差性により、後に出現する省略語に対する先行詞の候補からは除外されてしまうので、候補の中に正しい先行詞の含まれる比率が低下することになる。
4. 制約 2 は、他の制約と組み合わせて使用の方が効果的である。つまり、他の制約で、ある程度省略語の先行詞の探索範囲を限定することで、省略の非交差性によって候補から正しい先行詞が除外される可能性を減少させることができる。
5. 全制約を使用するより、制約 2 以外の制約を単独に、あるいは、制約 2 を含む制約をいくつか組み合わせて使用の方が効果的である。この理由としては、すべての制約を用いたために、正しい先行詞が除外される可能性が高まるためであると考えられる。制約 2 の単独使用を除外する理由としては、制約 2 を単独で使用した正解率が全制約を使用した正解率より低いからである。

## 6 今後の課題

今回は調査対象となる文章を各テキストの冒頭 7 段落に限定したため、(話者等を除く)省略の出現頻度は比較的低かった。また、全制約の場合を除き、複数の制約を組み合わせる場合を実験していない。今後、対象とする文章の段落数を増した場合や、複数の制約を組み合わせた場合について、検討する必要がある。今後の課題は、省略語補完の精度をさらに改善するために新しい制約を探ること、科学技術解説文以外の文章や会話文についても制約の有効性を検証すること、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、情報処理学会自然言語研究会研究報告、NL.86-6(1991)
- [2] 久野 暉：談話の文法、大修館(1978)
- [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(1985)
- [5] 石崎、井佐原：日本語文の複雑さの定性的・定量的特徴抽出、情報処理学会自然言語研究会研究報告、NL.67-6 (1988)
- [6] 内藤、島津、野村：談話処理の意味表現について、情報処理学会自然言語研究会研究報告、NL.58-4 (1986)
- [7] 竹本、田口、中川：質疑応答システムにおける日本語文の理解と生成、情報処理学会自然言語研究会研究報告、NL.74-13 (1989)
- [8] 平井、北林：省略語と事象間関係から見た日本語文の意味解析と文脈解析の枠組について、電子情報通信学会技術研究報告 86-8(1986)
- [9] 吉本 啓：談話処理における日本語ゼロ代名詞の扱いについて、情報処理学会自然言語研究会研究報告、NL.56-4 (1986)
- [10] 三吉、安川、平川、向井、田中、横井：状況意味論に基づく談話理解システム DUALS、情報処理学会自然言語研究会研究報告、NL.50-7 (1985)
- [11] 柏岡、高野、平井、北林：対話参加者の知識状態を用いた省略語の補充、情報処理学会論文誌、Vol.33, No.10, pp.1203-1210 (1992)