

日本語文章における省略補完の精度向上に関する考察

3M-6

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

文献[1]において提唱した談話的制約のうち、省略に関する制約を使用し、省略補完を試みた[2]が、必ずしも十分な結果が得られなかった。そこで、取り立て助詞が付加された時の格助詞の語形変化と、文型の選択誤りを減ずるような文型選択関数についての再検討を行なうことで、省略補完の精度の向上を計った。

2 省略補完の精度向上について

2.1 助詞変化の精密化

格助詞に取り立て助詞が付加されることによって生ずる語形変化を、文献[5]に従い精密化した。ここでいう格助詞、取り立て助詞とは、文献[5]と同様な助詞の分類を行なっている文献[6]に基づき以下のものを指している。

• 格助詞

格助詞は、体言または体言に準ずるものに付いて、それが文の成分としてどういう役割をするか、また他の語に対してどういう関係に立つかを示す助詞[6]。

(例) 「が」、「を」、「に」、「で」、「と」、「へ」など

• 取り立て助詞

多かれ少なかれ話し手の主観に基づく何らかの種類の強調を与える助詞[6]。

(例) 「は」、「も」、「こそ」、「なら」、「でも」、「さえ」など

2.2 文型選択関数

IPAL辞書にはひとつの用言に対して、一般に複数の文型が記載されている。文献[2]においては、動詞(形容詞)のとりそれぞれの格形式とそれらのとり得る素性の両方が一致した場合とそうでない場合の2つの

場合のみを考慮(表1参照)していたが、この文型選択関数では約14%の選択誤りが生ずる[2]。そこで、いずれか一方のみが一致する場合も考慮に入れた2つの文型選択関数(表1参照)を作成した。

			格形式	
			一致	不一致
方法1	素性	一致	+5	-1
		不一致	-1	-1
方法2	素性	一致	+25	-1
		不一致	+5	-5
方法3	素性	一致	+100	-5
		不一致	-1	-25

表1. 文型選択関数

3 計算機実験

3.1 実験概要

動詞・形容詞辞書としてIPAL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段落と「朝日新聞」の天声人語5編を対象として、語の素性一致制約のみを用いて省略語の補完を行なった。但し、省略語の先行詞の候補の中に正しい先行詞が含まれているか否かの判断は、人間の視察により行なった。

3.2 実験結果

実験結果を表2から表5に示す。表中に使用されている、「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省略語、「省略語(b)」とは、「わたし(筆者、話者)は」などを指す省略語、「省略語(a)」とは、「省略語(b)」、「未登録語」を除く省略語である。

3.2.1 助詞変化の精密化

この実験では、方法1(表1)の文型選択関数を用いた。

A Consideration on Improvement of Ellipsis Supplement in Japanese Sentences

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

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

		サイエンス	天声人語
変更前	省略語 (a)	4211 (75.2)	3328 (48.2)
	未登録語	568 (114)	62 (62.0)
	省略語 (b)	3969 (58.4)	1643 (42.1)
変更後	省略語 (a)	5303 (79.1)	3280 (49.0)
	未登録語	590 (118)	31 (31.0)
	省略語 (b)	3643 (61.7)	1535 (39.4)

表 2. 変更前後の先行詞の候補の総数と省略語 1 語あたりの先行詞候補数 (語)

(※括弧の中が、省略語 1 語あたりの先行詞候補数)

	サイエンス	天声人語
変更前	85.7 (48/56)	85.5 (59/69)
変更後	86.6 (58/67)	82.1 (55/67)

表 3. 変更前後の正解率 (%)

(※括弧の中は正解数 / 出現回数)

3.2.2 文型選択関数

この実験は、助詞変化の精密化を組み込んで行なった。

		サイエンス	天声人語
方法 1	省略語 (a)	5303 (79.1)	3280 (49.0)
	未登録語	590 (118)	31 (31.0)
	省略語 (b)	3643 (61.7)	1535 (39.4)
方法 2	省略語 (a)	5097 (77.2)	3108 (50.1)
	未登録語	579 (116)	32 (32.0)
	省略語 (b)	3109 (54.5)	1782 (45.7)
方法 3	省略語 (a)	5332 (80.8)	3576 (54.2)
	未登録語	586 (117)	31 (31.0)
	省略語 (b)	3712 (62.9)	1735 (44.5)

表 4. 方法別の先行詞の候補の総数と省略語 1 語あたりの先行詞候補数 (語)

	サイエンス	天声人語
方法 1	86.6 (58/67)	82.1 (55/67)
方法 2	92.4 (61/66)	87.1 (54/62)
方法 3	89.4 (59/66)	84.8 (56/66)

表 5. 方法別の正解率 (%)

4 考察

4.1 助詞変化の精密化

助詞変化の精密化に伴い、適用可能な助詞変化の場合の数が多くなったため、次の 2 つの結果が生じた。

1. 正しい格助詞が適用されない場合が減少した
2. 正しい格助詞が選択される確率が減少した

検証の結果、サイエンスでは 1 の傾向が、天声人語では 2 の傾向が強かったため、サイエンスでは正解率の上昇、天声人語は正解率の減少を招いた。

4.2 文型選択関数

文型選択関数を変更した結果、文型選択誤りを 1 割程度 (方法 2) に抑えることができた。これにより、省略格の認定の精度向上、さらに、省略補完の精度向上が期待できる。

5 おわりに

省略補完の精度を向上させるために、助詞変化の精密化と文型選択関数の変更を行なった結果、省略語 1 語あたりの先行詞候補数を増やすことなく、正解率を向上させることができた。今後は、これらの改善を行なった後、省略に関する談話的制約と話題を併用し、実際の文章を対象として省略補完を行なう予定である。

謝辞

「角川類語新辞典」[3] 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] 藤澤、増山、内藤：省略補完における談話的制約の有効性の検証、情報処理学会自然言語研究会研究報告、NL.96-6(1993)
- [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)
- [5] 鈴木重幸：日本語文法・形態論、むぎ書房 (1972)
- [6] 宮田幸一：格助詞と取り立て助詞、月刊言語、Vol.9, No.12, pp.68-77(1980)