

日本語文章における照応・省略現象の基本的検討

藤澤伸二*

増山繁*

内藤昭三**

* 豊橋技術科学大学知識情報工学系

*NTT 基礎研究所

本稿では、日本語文章において、文の冗長度を下げることが目的としてしばしば使用される照応や省略が、実際の文章でどのように用いられているかを調査した結果を報告する。今回は、科学雑誌の記事5編と天声人語70編を調査の対象として、どのような句が省略されているか、どのような指示詞、代名詞、代動詞が使われているか、これらの先行詞がどのくらい前(あるいは後)の文に現れるかを調査した。さらに、照応・省略に関する8つの仮説を立て、これらの成立についても検証を行なった。

A Basic Study on Ellipsis and Anaphora in Japanese Sentences

Shinji FUJISAWA*

Shigeru MASUYAMA*

Shozo NAITO**

*Dept. of Knowledge-based Info. Eng., Toyohashi Univ. of Tech.

*NTT Basic Research Laboratories

In this paper we investigate how ellipsis and anaphora are used in Japanese sentences. We study what kind of phrases are omitted, when anaphora is used, what kind of pronouns and proverbs are used, and where and how many times they appear against articles and essays. Then we pose eight hypotheses about ellipsis and anaphora and examine their validity.

1 はじめに

本稿では、日本語文章において、文の冗長度を下げることが目的としてしばしば使用される照応や省略的的を絞り、それらが実際の文章でどのように用いられているかを調査した結果を報告する。

照応・省略は、日本語においてしばしば行なわれるもので、言わなくても聞き手にとって自明の情報を省くことによって、文の冗長度を下げることが目的とする [1]。照応とは、代名詞、代動詞、指示詞などの照応語とその先行詞との組によって表される言語現象で、伝達される情報のかなりの部分が談話文脈から推定できる [7]。それに対し、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が、ここでは、省略を I P A L の辞書 [3][4] 記述中の格要素の省略と定義する。省略は、照応現象の一種(「ゼロ照応」とよぶ)とみなすことができる。

今回は、日本語文章において、省略や照応がどのように使われているか、また省略と照応にどのような違いがあるのかを知るために、統計調査を行なった。調査の対象とした文章は、科学雑誌(日経サイエンス)の記事5編と天声人語70編である。今回の調査は、どのような句が省略されているか、どのような指示詞、代名詞、代動詞が使われているか、これらの句や品詞がどのくらい前の文を指すのかについて行なった。さらに、照応・省略に関する8つの仮説の成立についても調査を行なった。

2 文の種類

文の種類は普通、単文、複文、重文に分けられる。以下に、それぞれの定義と例文を示す。この定義は、[5]に基づき、修正を加えたものである。

複文: 少なくとも一つ埋め込み構造を持つ文。

(例文 1) 雪の降る日は寒い。

重文: 最上位で複数の主述関係が対等に並列しているもの。

(例文 2) 花は咲き、鳥は歌う。

単文: 重文でなくかつ複文でない文。

(例文 3) 犬がいる。

この定義は、重文かつ複文も認めている。さらに、重文でない文(つまり、最上位での主述関係が一回だけのもの)を新たに単位文と定義する。上の例で、(例文 1)と(例文 3)が単位文である。また、複文の中の主述関係の認められる従属的部分や重文の中の主述関係の成り立っているそれぞれの部分を節という [5]。但し、重文を除いた文自身もひとつの節であると定義する。複文中の節を埋め込み文といい、次の(例文 4)のように何重にも入れ子状にできる。その時、入れ子状の外側(内側)を上位(下位)の埋め込み文(節)という。

(例文 4) 太郎が来たとき次郎がいったと三郎は思った。

3 省略・照応に関する仮説

今回検証した仮説は以下の通りである。

(仮説 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。

(仮説 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。

(例文 5) 日本では、「さきがけ」と「すいせい」のコンビがハレー彗星に送られた。そして、それまでに磁気圏内で磨いた技術をいかし、太陽風の観測に成功した。

2番目の文のそれぞれの節に、「日本は」が省略されている。これらの省略語の先行詞はどちらも最初の文の先頭にあり、確かにひとつめの先行詞と省略語の組とふたつの先行詞と省略語の組とは交差しない。従って、この例文は(仮説 2)を満たす。

(仮説 3) 照応は段落に跨ることが少ない。

(仮説 4) 照応は交差することが少ない。

(例文 6) 季節変動パターンがあればそれからバケツの放置時間を推定でき、

それによってすべてのデータを修正することができる。

最初の「それ」の先行詞は「季節変動パターン」であり、次の「それ」は「バケツの放置時間を推定できること」である。従って、この例文では照応の交差が見られない。

(仮説5) 最上位の節を除く各節において、省略の先行詞はその節またはその上位の節に含まれることが多い。

(例文7) 私は、お腹がすいたのでレストランに行った。

「お腹がすいた」の節の省略語は「私は」であり、その先行詞はその節の上位の節にある。

(仮説6) 単文、または重文の最上位の先頭節の埋め込み文以外の節中の省略の先行詞は、それより前の文中にある。

(例文8) 温度計が熱を持った近くに設置されたかを知ることはほとんど不可能である。このような場所に設置されておれば、温度計の読みは高くなってしまふ。

2番目の文は主格が省略されている。その省略語の先行詞は、最初の文の「温度計が」である。従って、この例文は(仮説6)を満たす。

(仮説7) 複文や重文の先頭を除く各節において、照応の先行詞は、その節やその上位の節には含まれない。

(例文9) 過去100年間の気象データによると地球は温暖化しているようだが、これだけで地球の温暖化を結論することはできない。

この文は重文であり、2番目の節の「これ」の先行詞は、最初の節である。従って、この例文は(仮説7)を満たす。

(仮説8) 単文、または重文の先頭の節において、その埋め込み文以外の照応の先行詞は、それより前の文(いくつかの文と段落を含む)にある。

(例文10) 機械観測史上初期の記録は、あまり役立たなかつた。これらの記録は西ヨーロッパしかカバーされていなかったからである。

2番目の文の「これら」の先行詞は、「機械観測史上初期」である。従って、この例文は(仮説8)を満たす。

これらの仮説は、書き手は「協調性の原則」に従う[2]、すなわち、できるだけ読み手にとってわかりやすい文章を書くとの仮定の下に立てたものである。

4 調査

4.1 調査方法

4.1.1 省略語の復元方法

今回の調査は、以下のように人間の視察によって行なつた。最初に、文の基本的構成要素である述語があるかどうかを見る。もしなければ、それ以前の文中や節中にある述語と文章中には出現していない関連する述語の中から最も適切なものを選び、それが省略されているとする。その根拠は、述語の省略には、並列表現による省略、「ダ」パターンによる省略、きまりきった表現による省略の3種類があるからである。述語が見つかったところで、その述語がとり得るそれぞれの格にそれ以前の文(省略語を含む文が重文である場合、その後の節も含む)の要素の中から最も適切なものをそれぞれ選ぶ。もし、適切なものがなければ、近くにある文中の要素や文章中には出現していない関連する語の中から最も適切なものを選ぶ。ここで、最も適切な選び方とは、選ばれた語を省略位置に代入しても文に意味上の矛盾が生じず、かつ、省略位置より近くにある構成要素をとることである。

4.1.2 照応の先行詞の決定方法

照応の先行詞の決定方法は、その照応語から予想される要素(句、節、文、段落)をその照応語の先行詞の候補とする。もし、同一のものを表す語が複数個ある場合、一番近くにあるものを候補として残す。そして、残った候補の中から最も適切なものを先行詞として選ぶ。

4.2 調査結果

調査結果を表1から表10に示す。但し、表5、6では、省略の全出現から「不特定(多数)者、筆者、読者」を除いたもの(出現数はサイエンスで308、天声人語で125)を対象とした。

今回調査を行なった科学雑誌と天声人語において、段落数と文数はそれぞれ、科学雑誌の記事5編で317、1087、天声人語70編で436、1426であった。1段落あたりの文数の平均は、科学雑誌で3.4、天声人語で3.2と大差ない。今回は、ほぼ同量の文章について調査したので、天声人語の方が1文あたりの文字数が少ないことになる。

語句	出現回数	
	サイエンス	天声人語
指示詞	420	281
これ	70	32
この	191	56
こんな	0	8
ここ	1	4
こう(副詞)	0	3
それ	58	56
その	85	79
そんな	0	15
そこ	15	19
そう(副詞)	0	4
あの	0	3
あれ	0	2
代名詞	15	27
わたし	1	2
彼	14	20
その他	0	5
代動詞	17	29
こう(動詞)	12	16
そう(動詞)	5	13

表1. 照応語の出現回数

出現位置	出現回数	
	サイエンス	天声人語
	426	307
1つ後	0	2
同一文中	121	82
1つ前	280	165
2つ前	19	27
3つ前	4	18
4つ前	2	8
その他	0	5

表2. 照応の先行詞の含まれる文
(文それ自身を指す場合も含む)

出現段落	出現回数	
	サイエンス	天声人語
	452	337
同一段落中	411	272
1つ前	41	62
2つ前	0	3

表3. 照応の先行詞の含まれる段落
(段落それ自身を指す場合も含む)

語句	出現回数	
	サイエンス	天声人語
名詞句	636	473
一は	407	406
一が	210	51
一に	10	9
一を	7	7
一と	2	0
修飾詞句	9	8
一の	7	8
一な	2	0
動詞句	18	38
並列	9	3
その他	9	35

表4. 省略語の出現回数

出現位置	出現回数	
	サイエンス	天声人語
文	330	348
同一文中	264	147
1つ前	62	121
2つ前	3	48
3つ前	1	12
その他	0	20

表 5. 省略の先行詞の含まれる文

出現段落	出現回数	
	サイエンス	天声人語
	330	348
同一段落中	328	295
1つ前	2	46
2つ前	0	7

表 6. 省略の先行詞の含まれる段落

	出現回数	
	サイエンス	天声人語
成立	18	14
不成立	4	8

表 7. 仮説 5 の成立状況

出現位置	出現回数	
	サイエンス	天声人語
同一文中	1	1
1つ前	29	78
2つ前	3	31
3つ前	1	11
その他	0	16

表 8. 仮説 6 の成立状況

	出現回数	
	サイエンス	天声人語
成立	416	322
不成立	36	15

表 9. 仮説 7 の成立状況

出現位置	出現回数	
	サイエンス	天声人語
1つ後	0	2
同一文中	33	15
1つ前	272	170
2つ前	18	28
3つ前	3	14
4つ前	2	6
その他の文	0	3
前の部分	26	30

表 10. 仮説 8 の成立状況

1. 照応に関して

- 照応語のほとんどは指示詞であった(科学雑誌で93%,天声人語で83%)。
- 科学雑誌では、「こ」で始まる指示詞が多かった(58%)のに対し、天声人語では、「そ」で始まる指示詞が多かった(51%)。
- どちらも「あ」で始まる指示詞は非常に少なかった(科学雑誌で5回のみ)。
- 科学雑誌で代名詞は「彼」「わたし」しか用いられなかったが、天声人語ではそれらに加え「だんな」や「氏」も使われていた。
- 先行詞は、照応語を含む文の直前の文にあることが多かった(科学雑誌で62%、天声人語で49%)。
- 天声人語には、後方照応も使用されていた(2回のみ)。

2. 省略に関して

- 省略句の大半は主格の後置詞句であった。(科学雑誌で93%, 天声人語で88%)。
- 先行詞は、科学雑誌では省略語を含む文に含まれていた(80%)のに対し、天声人語では、その文または一つ前の文に含まれていた(77%)。

3. 仮説の成立に関して

- 文に跨る省略のうち段落を越えるものは、科学雑誌で66回中2回(3%)、天声人語で201回中53回(26%)使われていた。従って、(仮説1)は科学雑誌で成立しているが、天声人語では必ずしも成立しているとはいえない。
- 省略の交差現象は現れていなかった。従って、(仮説2)は成立している。
- 文に跨る照応で段落を越えるものの割合は、科学雑誌で331回中41回(12%)、天声人語で255回中65回(25%)であった。従って、(仮説3)は科学雑誌でほぼ成立しているが、天声人語では必ずしも成立しているとはいえない。
- 照応の交差現象は、科学雑誌で2回、天声人語で1回現れていた。従って、(仮説4)は成立している。
- トップレベルの節を除く各節において、省略の先行詞はその節またはその上位の節に含まれることが多かった(科学雑誌で81%, 天声人語で64%)。従って、(仮説5)は成立している。
- 単文、または重文の最初の節において、その埋め込み文以外の省略の先行詞は、ほとんどそれより前の文中にあった(科学雑誌で97%, 天声人語で99%)。従って、(仮説6)は成立している。
- 複文や重文の先頭を除く各節において、照応の先行詞がその節やその上位の節には含まれない割合は、科

学雑誌で92%、天声人語で96%であった。従って、(仮説7)は成立している。

- 単文、または重文の最初の節において、その埋め込み文以外の照応の先行詞がそれより前の文(いくつかの文と段落を含む)にある割合は、科学雑誌で91%、天声人語で94%であった。従って、(仮説8)はほぼ成立している。

5 考察

今回の調査結果より明らかになった省略と照応の使われ方の違いは、以下の通りである。

1. 省略では先行詞が同一文中に現れる場合が最も多いのに対し、照応では直前の文中に現れる場合が最も多い。
2. 省略句は、ほとんど主格であるが、照応語の先行詞は必ずしも主格とは限らない。
3. 照応は省略ほど非交差の制約は厳しくない。
4. 複文や重文の先頭を除く節において照応や省略が使われる場合、照応の先行詞は照応語の含まれる文より前に、省略の先行詞は省略語の含まれる文にある。
5. 単文、または重文の最初の節に現れる省略の割合は、そこに現れる照応の割合に比べかなり小さい。(不特定多数者を除く省略678回中そこに現れる省略は171回で25%、また、全照応765回中そこに現れる照応は622回で87%である。)

1から3の理由のひとつとしては、省略では先行詞に関する情報が表層的には全く失われているのに対し、照応では照応語の種類により、ある程度先行詞に関する情報を持ち得るという相違点が考えられる。

また、4と5は省略が主として1つの文中で最上位を除く節間をつなぐのに使われ、照応が文と文の間、または、重文の節間をつなぐのに使われることを示している。この理由は、同一文(最上位の節)中に先行詞がある時に照応が使われると、表現がくどく、冗長

になるからだと思われる。

また、天声人語は科学雑誌に比べ、段落に跨る照応や省略が多い。この理由としては、天声人語では1つの記事は1つの主題だけを扱うことが多いため、その主題については段落間に跨る照応や省略の使用が多くなるということが考えられる。

ここで、(仮説6)と(仮説7)の反例(例文11と例文12)を紹介し、それが今回の調査対象の文章中にはほとんど出現しなかった理由を考察する。

(例文11) 読書に熱中していたので、太郎は呼ばれたことに気付かなかった。

(例文12) 自分に勇気があることは、太郎自身が知っている。

このような文が調査対象とした文章にほとんど出現しなかった理由として、これらは人物の描写なので、文学作品によく使われるが、科学雑誌や天声人語にはあまり用いられないからだと思われる。

参考文献

- [1] 久野：談話の文法、大修館(1978)
- [2] H.P.Grice: Logic and Conversation、Readings in Natural Language Processing, Morgan Kaufmann、p41-58、1975
- [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 (Basic Verbs) 解説編、辞書編、1987
- [4] 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形容詞辞書 I P A L (Basic Adjectives) 解説編、辞書編(上、下)、1990
- [5] 国語学会：国語学大辞典 1991
- [6] 日経サイエンス：日経サイエンス社
- [7] 山村、大西、杉江: 日本語文の照応問題 解決システム、電子情報通信学会論文誌、Vol73-D-2、No6、pp887-896、1990