適切な照応詞の選択についての分析

志自岐 真奈　　指導教員：山村 毅

1　はじめに

文章では，同じ言葉の繰り返しを避けるために指示・人代名詞に置き換えたり，省略したりする照応現象と呼ばれるものがある．特に機械翻訳システムや日本語文生成システムにおいて，正しい文脈にするためには適切な照応詞の選択が重要になる．これに関する研究として，橋本ら [1] が分析結果を元に照応表現選択モデルを構築し，モデルの有効性を示した．適切な照応表現を選択可能であることは明らかであるが，このモデルは複雑であるため，汎化の面から，私はシンプルなモデルの方が良いのではないかと考えた．しかし，「適切な照応詞の選択」に関する研究は少ない．そこで，まず更なる分析が必要であると考え，本研究では照応表現の調査を行う．

2　照応表現の調査

京都大学テキストコーパスを用いて，以下の調査を行う．

2.1　照応詞と先行詞の距離の関係

2.1.1　「省略は段落間を跨らない」という仮説の新聞記事における検証

藤澤ら [2] によると，科学雑誌は約 99.4% の割合で，天声人語は約 84.8% の割合で省略と同一の段落にその先行詞が含まれていることが分かった．しかし，各記事には書き方の特徴があるため，この仮説の成立性も変化するのではないかと考えた．そこで，新聞記事ではどの程度立証されるのか，調査を行った．

調査した結果，「省略は段落間を跨らない」という割合は約 74.5%(3230/4334) であった．藤澤ら [2] の研究と比べて低い結果であり，必ずしもこの仮説が成立するわけではないといえる．特に科学雑誌との間に大きな差が生じたのは，科学雑誌は文量が少なく，1 つの段落で 1 つの事柄について述べる傾向があるため，段落間を跨る省略が少なかったと考えられる．

2.1.2　各照応表現における照応詞と先行詞の距離の関係

照応表現のタイプと，照応詞と先行詞の距離に関係があるのではないかと考え，調査を行った．ここで，距離とは，照応詞候補と先行詞の文単位での距離のことである [3]．表 1 に調査結果の一部を示す．この表でいうところの格とは照応詞の深層格である．

表 1　照応詞と先行詞の平均距離

	ガ格	ニ格	ヲ格
出現した名詞句の繰り返し	3.38	3.78	4.09
指示代名詞・人代名詞	1.27	1.17	0.83
連体詞 + 名詞	2.14	2.75	2.12
その他 (言い換えなど)	3.13	4.38	2.61
省略 (ゼロ照応)	1.36	1.79	1.25

距離が遠い時は出現した名詞句を繰り返したり，言い換えたりすることが多く，反対に距離が近い時は省略や代名詞を使うことが多いと分かった．これは省略や代名詞は情報量が少ないため，距離の遠い名詞句を指しにくいからだと考えられる．

2.2　接続詞の前後における二つの節の主語の照応表現のタイプの違い

南 [4] は隣接する節間の関係を隣接する節間に出現している接続詞に基づいて 3 つに分類している．これらのうち，「〜ナガラ (継続)」のような接続詞 (A 類) は，接続詞の前後における 2 つの節の主語が一致しやすく，反対に「〜ノデ」のような接続詞 (B 類) は，一致しにくいという傾向がある．そこで，この傾向と照応表現のタイプに関係があるのかという観点から，A,B 類の接続詞の後節の主語の照応表現のタイプについて，調査を行った．下の例を用いて調査方法を説明する．

私は仕事が終わり，家に帰ってきた．
そして，ご飯を食べながらテレビを見た．

この例の場合は，接続詞が「〜ナガラ」なので A 類である．また，「見た」の主語は 1 文目の「私」であるため，「見た」の照応表現のタイプは省略になる．表 2 に調査結果の一部を示す．

表 2　接続詞の後節の主語における各照応表現の個数・割合

	A 類	B 類
出現した名詞句の繰り返し	2(2.47%)	9(7.83%)
指示代名詞・人代名詞	1(1.23%)	4(3.48%)
省略 (ゼロ照応)	71(87.65%)	80(69.57%)
外界ゼロ照応	6(7.41%)	10(8.70%)
不特定 (推論による照応など)	1(1.23%)	10(8.70%)
合計	81(100%)	115(100%)

B 類と比較して，A 類の接続詞の後節の主語は省略が使われる割合が高いと分かった．これは接続詞の前後における 2 つの節の主語が一致している場合に，特に高い傾向を示した．

3　まとめ

本研究では，コンピュータに実用性の高い「適切な照応詞の選択」をするシステムを実装するために，7222 個の照応表現の調査，分析を行った．そして，照応表現のタイプと，照応詞と先行詞の距離に関係があることや，A 類の接続詞の後節の主語は高い割合で省略が使われることが分かった．今後の課題としては，この調査結果を数学的に分析し，他の視点からの更なる分析も進めることで，実用性の高いよりシンプルな照応表現選択モデルの構築をしていくことが挙げられる．

参考文献

[1] 橋本ちさ恵，乾健太郎，白井清昭，徳永健伸，田中穂積：日本語文生成における照応表現の選択，情報処理学会研究報告，NL143-5，pp33〜40，2001.

[2] 藤澤伸二，増山繁，内藤昭三：日本語文章における照応・省略現象の基本的検討, NL86-6，1991.

[3] 奥村学:文脈解析 -述語項構造・照応・談話構造の解析-，株式会社コロナ社，2017.

[4] 南不二男：現代日本語の構造, 大修館，1974.