

「AはBがCだ」構文が出現する要因 —BCCWJを用いた検討—

傳研究室 20L1204W 吉田 真大

1. はじめに

日常生活において、頭の中に浮かんだイメージや考えをもとに発言をしたり、文を書いたりすることは非常に多い。発言や執筆は、いくつかの選択肢から選び取ることの幾度とない繰り返し、と言い換えることができるだろう。では、それぞれの選択はどのような秩序をもってなされているのだろうか。本研究では「AはBがCだ」構文について取り上げ、語やその並べ方を決定している方法について、その一端を検討した。

野田(1996)は、「～は～が…」構文を「～は」の部分に着目し、他の部分とどのような関係にあるかによって次に述べる6つに分類した(「この本は父が買ってくれた」構文、「象は鼻が長い」構文、「カキ料理は広島が本場だ」構文、「辞書は新しいのがいい」構文、「この問題は解くのが難しい」構文、「このにおいはガスが漏れてるよ」構文)。そして野田(1996)はこの分類、つまり文の構造と機能との関係性とを考察した。

また西山(1989)は、「象は鼻が長い」構文について、その解釈上での二つの可能性を示した。『象』がどうしたのかといえば『鼻が長い』という属性を有するのだ」という解釈と、『象の鼻』がどうしたのかといえば『長い』という属性を有するのだ」という二つの解の仕方を提示し、「象は鼻が長い」構文は単なる「象の鼻が長い」という文から象を主題化したものではないと考察した。

西山の行った、解釈が分かれうるという議論は、野田の分類した一つ「カキ料理は広島が本場だ」構文に対しても同様のことが言えるだろう。そしてこの二つの構文は共通点がある。「象は鼻が長い」構文であれば「象の鼻が長い」、「カキ料理は広島が本場だ」構文であれば「カキ料理の本場は広島だ」という、同じ格構造(語どうしの関係に注目した構造)を持つ文を助詞「の」を用いて表すことが可能である、という点だ。そしてこれら(「象の鼻が長い」、「カキ料理の本場は広島だ」)は一意に格構造を示すことができる文である。

ここで、助詞「は」が主題化でなく、かつ格構造を示しやすい他の形があるにもかかわらず、あえて「AはBがCだ」の形式を利用する場合はなぜか、どんな要因によるものか、という問いが生じた。そこで本研究では、現代日本語書き言葉均衡コーパスを用いて、「AはBがCだ」構文を含む文とそうでない文とを比較することでこれについて検討した。

2. 分析 1

2.1. 目的

分析 1 では、「AはBがCだ」構文が、同じ格構造の他の文と比べて解釈に幅があるのかどうか、について主観を除くことのできる方法で調べることを目的とした。

2.2. データ

データとして、現代日本語書き言葉均衡コーパス(Balanced Corpus of Contemporary Written Japanese: BCCWJ)の、係り受け構造データを含むコアデータ 1980 サンプルを対象とした。

2.3. 手続き

まず、コアデータの中から、助詞「は」、「が」をこの順で伴う名詞句（以下名詞句 A、名詞句 B）が、この順で一つの名詞句（以下名詞句 C とする）に係っている文を抽出した。また、このとき文として受動態であるものや否定形を含むものは除外した。その結果、67 文が抽出された。

次に、抽出した文から A、B、C に相当する部分を切り出し、「カキ料理は広島が本場だ」と「カキ料理の本場は広島だ」とのように、「A は B が C だ」構文の形式と、非「A は B が C だ」構文の形式の二つの形式に再構成した。

そして得られた 67 対、134 文を ChatGPT 3.5 に入力し、翻訳した。入力した日本語文の類型（野田（1996）に従った）と、出力の英文を類型したものを最終的な分析対象とした。

図 1. 例：データの加工

「AはBがCだ」構文
→ アワビ漁は「栽培漁業」が主流だ。
非「AはBがCだ」構文
→ 「栽培漁業」がアワビ漁の主流だ。

図 2. 例：出力の類型

アワビ漁は「栽培漁業」が主流だ。
Awabi fishing is mainly done through aquaculture.
→ Asubj_Cadv_Bcomp
※be動詞は補語などと統合、前置詞、助動詞は略記 (prep, aux)
略記：主語/subj, 補語/comp, 副詞 (句) adv

2.4. 結果

訳文の類型は「A は B が C だ」構文が 33 種類、非「A は B が C だ」構文が 26 種類となった。また、67 対中 20 対において、訳文の類型が一致した。それらの日本語での類型の内訳は「かき料理は広島が本場だ」構文：10 件、「この本は父が買ってくれた」構文：4 件、「象の鼻が長い」構文：3 件、「辞書は新しいのがいい」構文：2 件、「この問題は解くのが難しい」構文：1 件となった。

2.5. 考察

訳文の類型を見ると、26 種類だった非「A は B が C だ」構文よりも、33 種類だった「A は B が C だ」構文の方がより多様な解釈が存在することを示唆している。

2.6. 問題点

この分析では再現性に関して大きな 2 つの問題点が見られた。まず、ChatGPT 3.5 を利用したが、その制御が確立していない点。そして、データの抽出や加工が恣意的、主観的であるという点である。

3. 分析 2

3.1. 目的

2.6 で述べられた問題点を踏まえて、「A は B が C だ」構文と非「A は B が C だ」構文の持つ特徴の違いについてより客観的な視点から調べるために分析 2 を行った。分析 2-1、2-2 とともに、「A は B が C だ」構文と非「A は B が C だ」構文とで比較を行なった。「象は鼻が長い」構文と「象の鼻は長い」構文、「カキは広島が本場だ」構文と「カキの本場は広島だ」構文との特徴を比較した。分析 2-1 では名詞句 A、B、C の文字数、分析 2-2 では名

詞句 A, B, C に含まれる格助詞「の」の数に着目して分析した。

3.2. データ

分析 1 と同じく、BCCWJ から抽出したデータを分析対象とした。その上で、分析 1 とは異なり、「A は B が C だ」構文と非「A は B が C だ」構文とのどちらも加工していないデータを利用した。また、野田 (1996) の分類があるように、「象は鼻が長い」構文と「カキ料理は広島が本場だ」構文では異なる特徴があると予測された。そこで、「A は B が C だ」構文を「象は鼻が長い」構文群と「カキは広島が本場だ」構文群とに、非「A は B が C だ」を「象の鼻は長い」構文群と「カキの本場は広島だ」構文群との計 4 群に割り振った (カキ料理の「料理」については省略した)。また、出現した順番に依存した名付けである A, B, C を廃し、文内での格構造にしたがってそれぞれの名詞句を「(象/鼻/長い/カキ/広島/本場) 相当名詞句」とした。以下ではカッコ内の語で呼称することとする。なお、この構文の割り振りと、文内の名詞句範囲についての注釈の付与については分析者の判断に基づいて実施した。

3.3. 分析 2-1

3.3.1. 手続き

各群の各名詞句について、それぞれが何文字からなる名詞句であるかを計算した (カギカッコ (「) や中黒 (・) などの記号も含めた)。その後、各群、各名詞句条件における平均文字数を算出し、その様相について分析した。

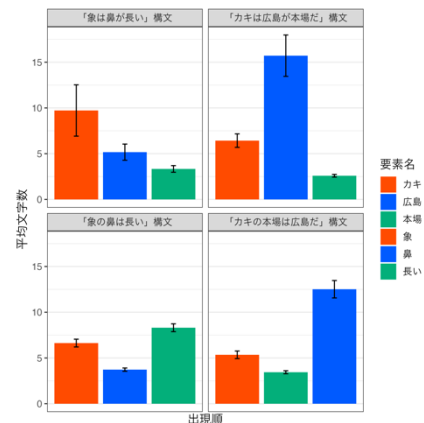
3.3.2. 結果

表 1 と図 3 によって示す。

表 1. 各名詞句の平均文字数

群	サンプル数	名詞句	平均文字数	標準誤差
「象は鼻が長い」構文	18	象	9.72	2.80
		鼻	5.17	0.88
		長い	3.33	0.36
「象の鼻は長い」構文	299	象	6.63	0.42
		鼻	3.72	0.18
		長い	8.31	0.43
「カキは広島が本場だ」構文	56	カキ	6.43	0.74
		広島	15.71	2.26
		本場	2.59	0.15
「カキの本場は広島だ」構文	167	カキ	5.35	0.42
		広島	12.51	0.95
		本場	3.44	0.16

図 3. 各名詞句の平均文字数



3.4. 分析 2-2

3.4.1. 手続き

各群の各名詞句について、それぞれに助詞「の」が何回出現するかを計算した。その後、各群、各名詞句条件における平均出現頻度を算出し、その様相について分析した。

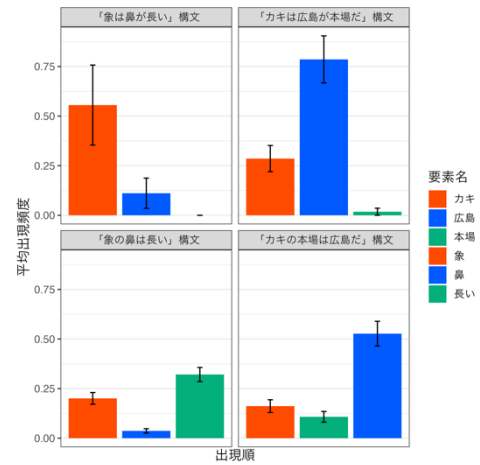
3.4.2. 結果

表 2 と図 4 によって示す。

表 2. 各名詞句における助詞「の」の出現頻度

群	サンプル数	名詞句	平均出現頻度	標準誤差
「象は鼻が長い」構文	18	象	0.56	0.20
		鼻	0.11	0.88
		長い	0.00	0.00
「象の鼻は長い」構文	299	象	0.20	0.03
		鼻	0.04	0.01
		長い	0.32	0.04
「カキは広島が本場だ」構文	56	カキ	0.28	0.66
		広島	0.79	0.19
		本場	0.02	0.02
「カキの本場は広島だ」構文	167	カキ	0.16	0.03
		広島	0.53	0.06
		本場	0.10	0.03

図 4. 各名詞句における助詞「の」の出現頻度



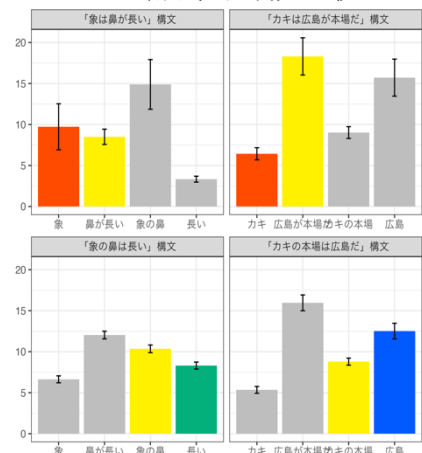
3.5. 考察

まず分析 2 全体について述べる。図 3, 図 4 を見ると, 2 枚の図のあいだで, 各群内の名詞区間の大小関係が一致していることを読み取ることができる。ここから, 名詞句の文字数と助詞「の」の出現頻度の間には相関があることが示唆されているといえるだろう。

まず「象は鼻が長い」構文群と「象の鼻は長い」構文群とを比較する(図 3 左側)と, 棒グラフの概形が大きく異なっている。したがって, 「象は鼻が長い」構文と「象の鼻が長い」構文とを分ける要因の一つとして名詞句の文字数が関わっていることが示唆されているといえる。次に「カキは広島が本場だ」構文と「カキの本場は広島だ」構文とを比較する(図 3 右側)。文中に出現する順で並べたこの図では概形が異なるように見えるが, 名詞句の役割(「カキ」, 「広島」, 「本場」)で比較すると大小関係は一致し, 名詞句の文字数はこの二つの群を分ける要因としては弱いものであると考えられる。

そして, これらの結果, 特に分析 2-1 から, 名詞句の文字数の中でも「は」の前後を比べたときに文字数が近づくような構文を選択しているのではないかというものが考えられた。そこで図 5 に, 各群での助詞「は」の前後での文字数の比較を示した。色がついているのが当該構文でのもので, 灰色となっているのが, 構文を上下で交換した場合のものである。左上以外の 3 群において, 色がついているものどうしの差の方が灰色どうしよりも小さいことがわかる。したがって,

図 5. 各群における, 「は」前後での名詞句の文字数の比較



4. 総合考察

分析 1 から, 共通した格構造を持っていても「A は B が C だ」構文の方が解釈に幅があることが示唆された。分析 2 から, それぞれの名詞句の文字数などの要因が構文の選択に関わっていることが示唆された。しかし, 「カキは広島が本場だ」構文群のみが最後の助詞「は」の前後での文字数の比較による記述にあてはまらず, 他の要因などのさらなる分析が求められる。