

# 日本語話者の文章に関する直感の数量化から何が学べるか ——個々の直感の壁を越えて——

福 田 伸一郎\*

## 1. はじめに

多くの言語の統語研究がそうであるように、日本語の統語研究も日本語話者の日本語の文章の認容度に関する直感を基本的なデータベースとして発達してきた。これらの直感的判断は多くの場合少数の話者によって提供されたものであり、研究者自身が直感的判断の提供者であることが一般的だ。言語学者の直感的判断は、日本語の統語研究の発展において重要な役割を果たした数々の観察を生み出し、日本語の統語研究の発展に多大な貢献をしてきた。しかし1990年代から、数多くの話者から文章の認容度に関する判断を引き出す実験方法が統語研究に導入され始め、多数の話者が実験対象として比較的簡単に集められる言語、例えば英語などの研究において、実験によって引き出された文章の認容度の判断をもとにした研究がより一般的になり始めている。本稿では、日本語からの実際の例をもとに、実験によって文章の認容度を調べることの利点について述べる。

## 2. ケーススタディ：好き

先に述べたように、日本語母国語話者の直感的判断は、日本語の統語研究の発展において重要な役割を果たした数々の観察を生み出し、日本語の統語研究の発展に多大な貢献をしてきた。多くの場合そのような直感的な判断は、一部分を除いて

は全く同じ二つの文章の認容度の違いを示すことによって提示される。例えば下記の(1)と(2)の2つの文は、目的語である「寿司」に着く格助詞が(1)では「を」であるのに対し、(2)では、「が」となっている違い以外は、まったく同じである。しかしたいの日本語話者が(1)を自然な文であると判断するのにに対し、(2)を日本語の文章としては認められない、非文と判断する。本稿では慣例に沿い、非文と判断される文には「\*」を文頭に置くこととする。

(1) 太郎が寿司を食べた。

(2) \*太郎が寿司が食べた。

(1)と(2)の唯一の違いが目的語の格助詞であることから、(2)が非文であると言う事実は、(1)のような文章では目的語の格助詞が「が」であってはならないことを示していると言える。

(1)と(2)は比較的単純な例であるが、基本的には同じ形で、様々な異なった種類の、時には非常に複雑な構造を伴った日本語の文章に関する数多くの観察が生み出され、日本語の統語研究は目覚ましい発展を遂げてきた。しかし、(1)と(2)のような「ミニマル・ペア」、または最小対立の関係にある文章の容認度を比較するという形でのデータの獲得には制限があることも事実だ。具体的な例として、三原・平岩(2006)からの下記の例文を見てみよう。

(3) 太郎はコーヒーが好きだ。

(4) \*太郎はコーヒーが好きだ。

(3)と(4)の唯一の違いは、(3)では目的語である「コーヒー」が「が」格でマークされているのに

\*ハワイ大学マノア校

対し、(4)では同じ目的語が「を」でマークされていることである。従って(3)と(4)は最小対立の関係にある。三原・平岩 (2006) は(3)と(4)に似通った最小対立の関係にある文章をいくつかあげ、「好き」を含む特定の述語の場合、目的語は「を」格ではなく、「が」格でマークされなければならないことを指摘している (180)。同様に Iwasaki (2013) でも、「好き」は目的語を「が」格でマークする述語の一つとして分類されている (105-106)。しなしながら、Martin (1975) や柴谷 (1978) が述べているように、「好き」や「嫌い」の目的語が「を」格でもマークされうるといふ観察は以前から存在する。下記は柴谷 (1978) からの例文である (229-230)。

(5) ア。「あなた、もしかしたら先生がお好きなんじゃない？」

イ。「あなたは先生をお好きなの？」

(6) ア。「村井さんは子供がおきらいらしいですね。」

イ。「あなたはうちの子をおきらいでしたものね。」

(7) ア。僕は飛行機が大好きである。

イ。それなのに、何故事更に飛行機を好きだと宣伝するのかという…。

(5)から(7)のア、イの例文は、同じ作家によって書かれた、同じ書物から抜き出されたものである。これをもって柴谷は、「好き」や「嫌い」の目的語の格助詞の「か」と「を」との交換が、同じ話者の文法の中で起こることを示している。

それではなぜ、三原・平岩 (2006) では(3)と(4)の間に顕著な認容度の違いがあるとされているのに対し、柴谷 (1978) があげるように、「好き」や「嫌い」の目的語の格助詞が「が」でも「を」でもよい場合が存在するのであろうか。一つ考えられる理由は、目的語の指示対象の違いである。三原・平岩 (2006) の(3)と(4)では、目的語は無生物である「コーヒー」だ。しかし柴谷の例の多くは、人物が目的語となっている。実際、柴谷 (1978) の

あげる例文の中で無生物を目的語としてもつ例文は、(7)ともう一つ、下記の(8)だけである。

(8) 昔風の職人気質を信吾は好きで、仕事をさせてみた。

これらの観察から、(9)の仮説が立てられる。

(9) 「好き」や「嫌い」の目的語の格助詞は、目的語の指示対象の有生性によって左右される。目的語が生物を指示する場合、格助詞は「が」でも「を」でもよいのに対し、目的語が無生物を指示する場合は、「を」格の認容度は低い。

もし(9)が正しければ、「好き」や「嫌い」が述語である文章の容認度を比較する場合、(10)の二つの要素を考慮する必要があることになる。

(10) ア. 目的語の有生性 (生物⇔無生物)

イ. 目的語の格助詞 (が格⇔を格)

従って、「好き」や「嫌い」が述語である文章の容認度を考慮するにおいては、(3)と(4)のような最小対立の関係にある二つ文章を比較するのでは十分ではなく、(10)の二つの要素の二つの可能な値を掛け合わせた4つのパターンの文章を比較し、二つの要素の相互作用の有無を検証する必要がある。

しかしながら、二つの文法的要素の相互作用の有無を、話者の文章の容認度に関する直感から検証することは非常に難しい。なぜなら、表①の二つの要素間の相互作用の有無を知るには、AとBの違いとCとDの違いを同時に評価し、これらの二つの違いの間の差を比較する必要があるからだ。

ではどうしたら二つの文法的要素の間に相互関係があるかどうかを検証することができるだろうか。一つの方法は、自然に起こった「好き」や「嫌い」を述語とする文を分析し、二つの要素間の相

【表①】

	が格	を格
生物	A. 太郎は花子が好きだ。	B. 太郎は花子を好きだ。
無生物	C. 太郎はコーヒーが好きだ。	D. 太郎はコーヒーを好きだ。

相互作用が認められるかどうかを調べる方法（コーパスサーチ）だろう。もう一つの考えられる方法は、二つの要素を体系的に操作した文章を作り、その容認度の高さを複数の話者を対象にした実験を行って調べる方法だ。本稿ではこれらの二つの方法を試験的に実施した結果を報告し、実験によって文章の容認度を調べることの利点を論じる。

### 3. コーパスサーチ

コーパスサーチによって事象の性質を検証する場合に鍵となるのは、対象となる事象の頻度である。Martin (1975) によると、1964年に国立国語研究所が出版した90の雑誌を調査対象とした研究では、「好き」や「大好き」を述語とする文章で、目的語が「を」格をとっている例は一件しか確認されなかったという。さらに目的語が「が」格でマークされていた例は17件、そして「の」格であった場合は2件と、見つかった例文の数は合計20にしかない(200)。言うまでもなく、たった20件の例文からは学べることは非常に限られている。

本研究でも似たような結果に終わることを避けるには、「好き」の目的語がどのような環境で「を」格をとる可能性が高いかを考慮し、そのような環境をコーパスの中で探す必要がある。この点において、柴谷(1978)では下記の興味深い考察が見られる。柴谷は、(11)のように「好き」が名詞を修飾する文章の述語であった場合、名詞句「子供の」の解釈が曖昧であると指摘している。

(11) 子供の好きなオバさんがやって来た。

(11)は「オバさんが子供が好き」とも、「子供がオバさんが好き」とも解釈できる。前者の場合「子供」は目的語であり、後者の場合は主語である。同じ曖昧さは「子供」の格助詞が「が」である場合にもみられる。

(12) 子供が好きなオバさんがやって来た。

これは、名詞を修飾する文章内の名詞句が「が」格を持つ場合、下記の(13)の例文のように、「の」

格と交換することが可能であるためである。

(13) ア. これは太郎が食べたケーキです。

イ. これは太郎の食べたケーキです。

これに対し「子供」の格助詞が「を」であった場合、「子供」は目的語としてしか解釈されず、したがって(14)には一つの解釈しかない。

(14) ?子供を好きなオバさんがやって来た。

(14)の「?」は、柴谷(1978)がコメントするように、「好き」や「嫌い」の目的語の格助詞が「が」である場合と「を」である場合を比較した場合、前者の方が「スッキリ」しているとか、「より自然」であると感じられることが多いという観察を反映したものである(232)。しかしながら、(14)の重要な点は、「好き」の目的語を「を」格でマークすることによって、曖昧な文章がさけられるという点である。従って、「好き」が名詞を修飾している文章の述語である場合には、目的語を「を」格でマークすることに利点があるといえ、名詞を修飾する文章の中では、「好き」の目的語が「を」格を取る可能性が高いのではという仮説が立てられる。

上記の仮説をもとに、本研究では、国立国語研究所、情報通信研究機構、そして東京工業大学が共同開発した750万語を含む話し言葉のデータベース、「日本語話し言葉コーパス」を利用して、「好き」が名詞を修飾する文章内の述語である文例を集め、その中から「好き」の前に位置する名詞句が目的語として解釈できる例を抽出した。表②はコーパスサーチの結果をまとめたものである。

検索の条件にあう例は、合計83件確認された。その内、目的語と分析できる名詞句が生物を指示している例は14件と全体の17%に留まり、さらに目的語と分析できる名詞句が「を」格でマークさ

【表②】

	を格	が格	の格	他	計
生物	2	9	2	1	14
無生物	5	46	15	3	69
計	7	55	17	4	83

れている例は7件と、全体の8パーセントであった。従って、数字の上では1964年の研究よりは多いものの、依然統計的に分析するには少なすぎる結果となった。さらに予想外な点としては、目的語が「を」格をとっている7件の内、5件で名詞句が無生物を指示していることだ。

- (15) ア. その理由ってというのがまー早実を好きな理由でした。  
 イ. お祭りを好きな方が何かあの辺っこ揃っててお祭りも盛んだし  
 ウ. 要は音楽を好きな人を一人でも多く世の中に作っていこうということがえー根底にあります。  
 エ. 確かに彼女は気が強い部分てあるんですが、まそれを好きな人もかなりいますし  
 オ. そういうところでやっぱり魅力があってやっぱりボクシングを好きな人っていうのもやっぱり多いんだなっていう風に僕自身は思います。

確認された例の数が非常に少ないので、この結果から何か結論づけることはできないが、少なくとも(9)の仮説を支持する結果でないことは確かだ。

#### 4. 実験

文章の容認度を実験で調べることの最大の利点は、実験文を作成するに段階で、研究対象である事象において重要な役割を担うと考えられる文法的要素を特定し、それを体系的に操作することができることだ。そしてそのようにして作成した実験文の容認度の判断を数多くの話者から集めることので、特定した要素が複数である場合、その要素間に相互関係があるかどうかを統計的に調べることが出来る。先に述べたように、個人の話者が複数の要素が関わる複数の文章の容認度を比較することは非常に難しい。従って複数の文法的要素が関わりと考慮される文章の容認度を調べるには、

実験が最も適していると言える。

本稿では、「好き」や「嫌い」を述語とする文章において、目的語の有生性とその格助詞の間に相互関係があるという仮説を追及しているの、実験文ではこの二つの要素を体系的に操作した。(16)は実験文の基本的な構造を示したものである。

- (16) 名詞句<sub>1</sub>は {名詞句<sub>2</sub>が/を/の/に} {好き/大好き/嫌い/大嫌い} な名詞句<sub>3</sub>を動詞

名詞句<sub>2</sub>の有生性を操作し、格助詞を「が」、「を」、「の」そして「に」の4種類に変化させた。本研究の焦点である「が」と「を」以外に「の」格と「に」格を実験文に加えたのは、下記の理由からだ。まず先に述べたように、「が」格をとる名詞句が名詞を修飾する文章内にある場合、「の」との交換が可能である。従って、「好き」や「嫌い」が名詞を修飾する文章の述語であった場合、その目的語は「が」、「を」、「の」の三つの格でマークされる可能性がある。ということで、事象の全体像を掴むために、今回「の」格も条件に加えた。これに対し、一般的に「に」格は「好き」、「大好き」、「嫌い」、「大嫌い」の目的語をマークできないと言われている。従って、目的語が「に」格をもつ実験文は、他の条件がどうであれ、非文と評価されると予想される。よって「に」格を条件に加えることで、非文に対する話者の評価の基準が得られる。また実験文を作成する際には、「好き」と「嫌い」に加え「大好き」と「大嫌い」を使った。そして4つの述語のそれぞれにつき、語彙を変化させて2パターンの実験文を作ったので、実験文の合計数は64となった(有生性(2)×格助詞(4)×述語(4)×パターン(2)=64)。(17)は、名詞句<sub>2</sub>が生物で述語が「大好き」である文の例だ。

- (17) そのデパートのイベントは韓国のアイドル {が/を/の/に} 大好きな女性をターゲットにしていた。

64の実験文は、同じパターンからの実験文が違うリストに分配されるように4分割され、各リストの16の実験文は32文の他の実験の実験文と混

ぜられて、文章の順番は関連した実験条件の文章が二つ続かないように調整した。実験はインターネット上で実験を行えるウェブサイト (<http://spellout.net/ibexfarm>) を利用して行い、30名の東京に住む大学生が参加した。文章の容認度の評価は7段階（1が「全く不自然」で7が「全く自然」）で行うように指示した。参加者から集められた評価は、個々の話者の7段階評価の使い方の違いを考慮し、分析前に標準化（z-score transformation）した。データの分析は線形混合モデル（linear mixed effects model）を用い、名詞句<sub>2</sub>の有生性、格助詞（「が」格か「を」格か）、そして述語（好き、嫌い、大好き、大嫌い）を固定効果（fixed effects）として、そして被験者と実験文を変量効果（random effects）として分析した。

分析はR（R Core Team 2015）を使って行った。

分析の結果4つの述語間に有意な違いが見られたので、下記では各述語別に結果を見ることとする。図1は「好き」の実験文の容認度の平均を示す。平均が正数の場合は、全体の平均（0.0）よりも高く評価されたことを示し、負数の場合は全体の平均よりも低かったことを示す。誤差範囲は95%の信頼区間を示す。

「好き」の実験分では、格助詞は有意であったが（ $p < 0.01$ ）有生性は有意ではなかった（ $p = 0.12$ ）。格助詞と有生性の相互関係は有意に近いものであった（ $p = 0.06$ ）。

図2は「大好き」の実験文の容認度の判断の平均を示す。こちらも格助詞は有意であったが（ $p < 0.01$ ）有生性は有意ではなく（ $p = 0.32$ ）、格助詞と有生性の相互関係は有意であった（ $p < 0.01$ ）。

図3は述語が「嫌い」の実験文の評価の平均を示し、図4は「大嫌い」の実験文の結果を示す。

「嫌い」と「大嫌い」の実験文では格助詞だけが有意な違いを示し（嫌い： $p < 0.01$ ；大嫌い： $p < 0.01$ ）、有生性および格助詞と有生性の相互関係はともに有意ではなかった。

実験の結果は、「好き」と「大好き」を述語と

図 1

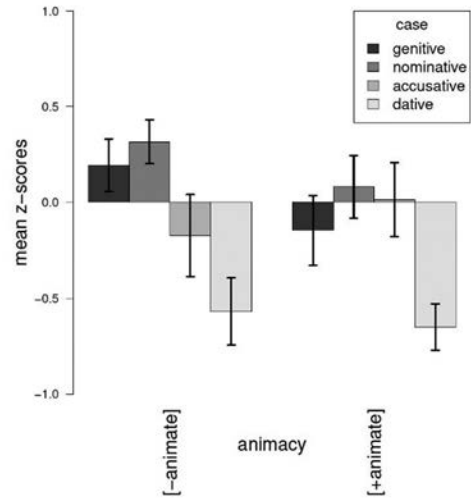
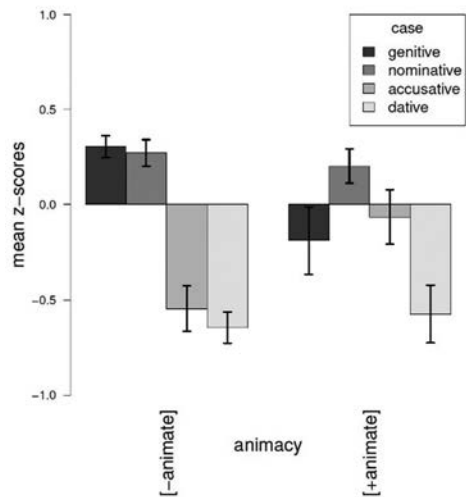


図 2



する文章に関しては、(9)の仮説を支持するものとなった。特に「大好き」の場合、目的語の有生性と格助詞との間に有意な相互関係が見られ、「大好き」を述語とする文章の容認性が二つの要素の相互関係によって予想できるという結果となった。

図 3

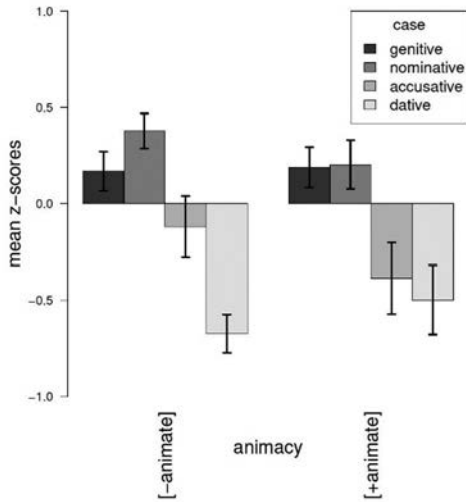
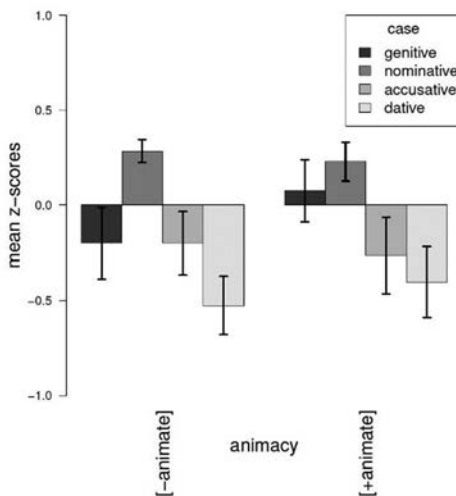


図 4



## 5. 終わりに

本稿では、「好き」や「嫌い」を述語とする文章をテストケースとして、実験的に文章の容認度を調べることの利点を論じた。先行研究に見られる例文を検証すると、「好き」や「嫌い」を述語とする文章の目的語の格助詞は、目的語の指示対象の有生性によって左右されるのではないかという仮説が立てられる。この仮説を検証するには、「好き

や「嫌い」を述語とする文章の容認度が、目的語の有生性と格助詞によっていかに変化するかを検証する必要がある。しかしこれは従来の最小対立の関係にある文章の容認度の比較という形で検証するのは非常に難しいし、事象の頻度の低さから、コーパスサーチでの検証も難しい。しかしながら、実験的に「好き」や「嫌い」を述語とする文の容認度の判断を複数の話者から集めることによって、上記の問題を乗り越えることが可能となる。具体的には、実験を行うことによって「好き」や「嫌い」の目的語が「を」格を取るケースが非常に稀であるという問題を乗り越え、また目的語の有生性と格助詞の選択が相互関係にあるのではないかという複雑な仮説も、統計を用いて検証することが可能となった。そして実験結果は、(9)の仮説を支持するだけでなく、「(大)好き」と「(大)嫌い」の間に、目的語の格助詞の選択における明らかな違いがあるという新しい観察も提示した。

言うまでもなく、統語学者にとって最も重要なのは、「好き」や「嫌い」を述語とする文章の構造の分析において上記の発見がどのような意味を持つのかを追及することである。しかし文章の構造の分析に取り組む前の段階として、まずは有益で信頼できるデータが必要なのは間違いない。今回取り上げた「好き」や「嫌い」を述語とする文章は、その様なデータを獲得するのに、実験が有益な手段であることを示している。

## 参考文献

- Iwasaki, Shoichi. (2013) Japanese. Amsterdam: John Benjamins.  
 Martin, Samuel. E. (1975) A Reference Grammar of Japanese. Tokyo: Charles E. Tuttle Company.  
 三原健一・平岩健 (2006) 『新日本語の統語構造—ミニマリストプログラムとその応用』 松伯社  
 R Core Team. (2015) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.  
 柴谷正良 (1978) 『日本語の分析：生成文法の方法』 大修館書店