

『日中言語対照研究論集』第21号 抜刷
(2019年5月16日 発行／白帝社)

日本語は無生名詞主語の自動詞文が 好まれる言語か？

—日本語文の主語となる有生・無生名詞の頻度比較—

玉岡 賀津雄
張 婧禕
牧岡 省吾

日本語は無生名詞主語の自動詞文が 好まれる言語か？

—日本語文の主語となる有生・無生名詞の頻度比較—

Is the Japanese language preferred to have an intransitive sentence
with an inanimate subject?:

A frequency comparison of animate and inanimate subject nouns

玉岡 賀津雄

TAMAOKA Katsuo

張 婧禕

ZHANG Jingyi

牧岡 省吾

MAKIOKA Syogo

Abstract : In Japanese, it is intuitively understood that inanimate nouns are more likely to be selected as the grammatical subject in comparison with European languages such as English and French (Hinds, 1986; Kunihiro, 1974a, 1974b). Furthermore, inanimate nouns are also likely to be used as the subject of intransitive verbs more frequently than transitive verbs (Ikegami, 1981, 2006; Nishimitsu, 2010; Teramura, 1976). To confirm these tendencies, the present study investigated the distribution of animate/inanimate subjects for 32 intransitive/transitive paired verbs (64 in total) using 18-years of the *Mainichi Shinbun Newspaper* corpus. The frequencies of animate and inanimate nouns (marked by the nominative case marker *-ga*) were analyzed separately for intransitive and transitive verbs. For the intransitive verbs, analyses of both token and type frequencies indicated that inanimate nouns were more likely to be selected as the subject

than the animate nouns. In contrast, for the transitive verbs, token frequencies showed no difference between animacy values. Type frequencies, however, revealed the same pattern such that inanimate nouns were more likely to be selected as the subject than animate nouns. Because token frequencies reflect actual linguistic characteristics, we argue that no difference exists among transitive verbs. To further investigate the difference in inanimate/animate subject frequencies between transitive and intransitive verbs, a two-way analysis variance of inanimate/animate subjects and transitive/intransitive verbs was conducted and revealed a significant interaction. In conclusion, the present study verifies the intuitive assumption that inanimate nouns are more likely to be selected as the subject for intransitive verbs.

キーワード： 主語 有生性 使用頻度 自動詞 他動詞

1 目的

英語やフランス語と比べて、日本語は無生物の主語を取る傾向があると言われている(国広 1974a, 1974b)。フランス語で2,000の状況についての2,000文の表現を集めた『*Le livre des deux mille phrases* [基本2千文]』(Frei 1953)がある。そのフランス語と英語訳を日本語訳に対応させると、日本語で「ボタンが取れちゃった。」という表現は、フランス語では、*J'ai perdu un bouton.*であり、英語でも*I've lost a button.*となる。フランス語も英語も同じSVO(主語・動詞・目的語)語順で他動詞を取って、一人称の「私」を主語で表現する。ところが、日本語では、無生物の「ボタン」が主語になり、SV(主語・動詞)語順の自動詞の「取れる」を使って、「ボタンが取れた。」と自動詞で表現するのが自然である。こうした例から、国広(1974a, 1974b)は、英語は「人間中心」であり、日本語は「状況中心」の言語であるとして、日本語では無生物の主語を取る文が多いと述べている。

Hinds(1986)も『*Situation vs. Person Focus* [日本語らしさと英語らしさ]』で、同様の指摘をしている。たとえば、日本語では「叫び声が出たぞ。」と「状況(situation)」に焦点(focus)をあてて言うが、英語では*I just heard shouting.*となり、「人間(person)」に焦点をあてて表現する。また、別の例では、日本語では、対象である無生物の「山」を主語にして「山が見える。」と表現する。しかし、英語では、「人間」に焦点をあてて「私(I)」を主語に取り、*I see a/the mountain.*とする。さらにまた、日本語では「腹

が減った。」と「腹」を主語にして述べる。「腹」は人間の体の一部であるが、それ自体では生きているとは言えないので、無生物と考えられる。しかし、英語では、やはり主語に「私 (I)」を立てて、*I am hungry.*とする。このように、Hinds (1986) も、英語は「人間」に、日本語は「状況」に焦点が置かれていると説明している。

以上の先行研究から、2つの仮説が立てられる。第1に、フランス語や英語などのヨーロッパ諸語と比べて、日本語では、無生名詞が主語になりやすい (Hinds 1986 ; 国広 1974a, 1974b) という仮説である。これは、動詞と共起する主語の無生名詞と有生名詞の頻度を比較することで検証することができる。第2に、日本語では他動詞よりも自動詞に無生名詞を主語として付けて表現する傾向がある (池上 1981, 2006 ; 西光 2010 ; 寺村 1976) という仮説である。

早津 (1989, 1995) は、740の基本的な動詞のうち自他対応のある動詞は220対の440語 (59.46%) であるとしている。このように、頻繁に使われると思われる基本的な動詞のうちの半数以上に自他対応があると考えられる。そこで、本研究では、1998年から2015年までの18年分の毎日新聞コーパス (形態素の総頻度合計は470,155,446) で主語と動詞の共起頻度が高い自他対応のある32対の動詞を選んで (合計64語の動詞、表1から表4を参照)、主語にどのくらいの無生または有生の名詞が選ばれているかを調べて、自動詞と他動詞を別々に頻度の差を分析し、上記の2つの仮説を検証することにした。

2 方法

2. 1 検索に使用したコーパスと検索方法

「CD-毎日新聞データ集」本社版 (日外アソシエーツ) の1998年から2015年までの18年間の新聞記事テキストデータを使用した。全国版の記事との重複を避けるため、大阪版記事を削除した。毎日新聞には、多くの新聞記者によって、政治、経済、芸能、スポーツ、IT、刑事事件、法律、クッキング、コンピュータ、ファッションなどの多様な記事が一般大衆に向けて書かれている。しかも、毎日新聞社の基準にしたがって標準的な日本語表現で書かれており、新聞記者や校閲者を複数の人の目を経て出版されている。そのため、自他対応動詞と共起する主語の無生名詞と有生名詞の頻度を比較するという意図に適していると考えられる。

このコーパスの形態素数 (重なり頻度 ; *type frequency*) は663,243、句読点などの記号を含む頻度総計 (延べ頻度 ; *token frequency*) は470,155,446、記号を除いた頻度総

計は398,406,147であった。検索対象となったのは17,090,360文であった。このコーパスに対してMeCab 0.996 (工藤・山本・松本 2004) による形態素解析を行い、テキストデータを形態素に分解した。ただし、MeCabで解析される形態素は、言語学で定義するところの意味上の最小単位とは異なり、固有名詞の「愛知時計電機」は、「愛知」「時計」「電機」で3語であるが、形態素1つとして数えられる。そのため、本研究が対象としたコーパスの総語数は、約4億語であると思われる。辞書はipadic 2.7.0 を使用した。さらに、日本語係り受け解析器のCaboCha 0.69 (工藤・松本 2002) を用いて、検索対象とするすべての文に対して係り受け解析を行った。

形態素の単位 (どこからどこまでを1つの名詞とみなすか) については、ipadic 2.7.0 の記載を基準とした。動詞の識別に際しては、ipadic 2.7.0に記載された動詞の基本形の表記が一致する場合に同一の動詞とみなした。「受賞者が決まりました」では「者」と「決まる」の共起関係が、「ハクチョウが両翼を大きく広げたように」では「ハクチョウ」と「広げる」の共起関係を記録した。検索対象である32対の自動詞と他動詞がそれぞれの名詞と共起する回数を積算して、共起頻度を算出した。

自動詞については、主語が明記されることが他動詞よりも多いので、主語との延べ共起頻度が150回 (文を基準としているので、文数と同じ) 以上を基準として32の動詞 (詳細は表1を参照) を選んだ。しかし、これらの32の自動詞に対応する他動詞では、最低の延べ共起頻度が2回で、最高が12,477回と大きな違いがみられた。それでも、平均延べ共起頻度は、1,439回 (詳細は表3を参照) であり、主語となる無生名詞と有生名詞の頻度を比較するには十分であると考え。これら以外の動詞を選ぶと、日本語では主語が省略されることが多く、主語との共起回数が少ないので、頻度分析の対象にはならない。本研究で使用している新聞コーパスは18年分の記事であり、これよりも大きいコーパスを使って目視での有生性判定をするのは労力的に難しい。

本コーパスでは、自動詞32語について311,791文、他動詞32語について46,048文で、合計357,839文を抽出した。最終的に、目視で主語の有生性を判断するため、研究対象としてはこのくらいの文数が限界であろう。なお、各語における文法情報 (受身、使役、可能、否定) は、動詞を基準として主語と共起する文を検索しているので、共起頻度の検索結果に含まれている。ただし、「売る」は他動詞で、「売れる」は自動詞であるが、「売れる」は「売る」の可能形でもあり、自他の頻度が重なるので比較の対象とはしていない。また、「教える」も「教わる」も他動詞で、自動詞との対応がないため本研究の

対象とはしていない。

2. 2 主語の判定

日本語では、基本的に主格を示す「ガ」が付いた名詞が主語になる。また、ハが付いた名詞も主語となることがある。しかし、ハは主題になる語に付くことが多い。たとえば、「マサヨとの口論は、君が悪いよ。」であれば、「口論は」の「は」は、主語ではなく、主題のハである。さらに、ハはさまざまな格に相当する名詞にも付く。「あのケーキはもう食べたよ。」であれば、「ケーキを食べた」と言い換えられるので、ハは対格のヲに相当すると考えられる。同様に、「南アルプスの北岳は去年登った。」であれば、「北岳に登った」と言い換えられるので、与格のニに相当すると考えられる。このように、ハが付加された名詞が主語であるかどうかを判定するのはきわめて難しい。そこで、本研究では、主語であることが確実に予想される典型的なガ格標識を持つ名詞を、調査対象の動詞ごとに抽出した。なお、1回だけ主語として出現する名詞は稀な語であるとして、2回以上出現した名詞のみを頻度として加算した。

本研究では、32対の自他対応動詞が、主語を担う名詞の意味表記において心的態度・意識・意志を持つかどうかで、主語の名詞が無生であるか有生であるかを判定した。たとえば、「生徒」「お子さま」「会長」などの人を表す名詞、「ペット」「稚魚」「牛」などの動物を表す名詞、および「李」「秀人」「イチロー」などの人の名前を表す固有名詞は、心的態度・意識・意志を持ち、それらには動きがあるので有生名詞とした。一方、「星」「給料」「画像」などの物事を表す名詞、「%」「一つ」や「割」などの数値や割合を表す名詞、「膝」「目元」「左肩」などの身体部分を表す名詞、「会社」「学校」「内閣」などの場所または組織を表す名詞、「タイ」「日本」「米国」などの国名を表す固有名詞などは、心的態度・意識・意志を持たず、それ自体では動くことができないので無生名詞と判定した。以上の手続きで、各動詞の主語となる無生名詞と有生名詞の「重なり頻度」と「延べ頻度」を目視で数えた。

3 分析結果

頻度については、同じ無生または有生名詞が複数回出現した場合、それらをすべて数えた延べ頻度が、実際の言語的な特性を示していると考えられる。一方、重なり頻度では、たとえば「イチロー」が新聞に何回出現しても1回と数える。しかし、実際の「イ

チロー」の出現頻度はそれよりはるかに多い。このように、重なり頻度は、実際に使われた頻度ではなく、使用された名詞の種類についての指標である。したがって、本研究では延べ頻度の結果に基づいて、主語になる無生名詞と有生名詞の出現傾向を考察する。主語の有生性については、自動詞と他動詞が異なる言語的特性を示す可能性がある。そこで、自動詞32語と他動詞32語について、別々に分析した。

3. 1 自動詞32語についての分析

3. 1. 1 自動詞32語の主語となる名詞の無生と有生の延べ頻度の差

自動詞32語と共起するガ格を取って主語となる無生名詞と有生名詞の延べ頻度および無生名詞から有生名詞の延べ頻度を引いた数値を表1に示した。自動詞32語について、主語となる無生名詞の延べ頻度の平均は8,614回で、標準偏差は11,127回であった。有生名詞の延べ頻度の平均は1,130回で、標準偏差が2,083回であった。延べ頻度の分布をできるだけ正規分布に近づけるために自然対数に変換した。ただし、「閉まる」「冷える」「破れる」の有生名詞の頻度が0回であったため、そのまま対数変換したのでは計算ができない。こういう場合、頻度に1を足して $\log_e(x+1)$ の自然対数変換を行うことが多い。しかし、Yamamura (1999) は、離散分布を連続分布で近似するという考え方からすれば、1でなく0.5を足すほうがよいと報告している。そこで、 $\log_e(x+0.5)$ の自然対数に変換して分析した²⁾。たとえば、自動詞「起きる」の延べ頻度は23,201回であるが、自然対数に変換すると $\log_e(23,201+0.5)=10.05$ になる。

延べ頻度を自然対数に変換した値の無生名詞と有生名詞のピアソンの積率相関係数は高く ($N=32$, $r=.66$, $p<.001$)、この係数は有意であった。相関係数は、-1から+1の範囲の数値を取り、絶対値が1に近いほど相関が強くなる。0.7以上で非常に強い相関であるとされており、0.20以下では相関はほぼないと判断される。本研究の0.66という相関は、0.70に近く強い相関である。一方、延べ頻度を使った場合のピアソンの積率相関係数は低かった ($N=32$, $r=.26$, ns)。この0.26という相関は、非常に弱い相関であると判断される。自然対数に変換していない延べ頻度では、無生名詞と有生名詞が歪んだ分布になるので、相関係数が小さくなったのではないかと考えられる。この点からも延べ頻度を自然対数に変換して分析することに根拠があることが分かる。以下、自然対数を使った分析は、延べ頻度の分析よりも厳密な分析結果を導く数値であるとする。

自動詞の無生名詞と有生名詞の延べ頻度については、対応のあるサンプルの*t*検定で検討した。*t*検定は、2つの群の平均を比較して、両者に違いがあるかどうかを検定する統

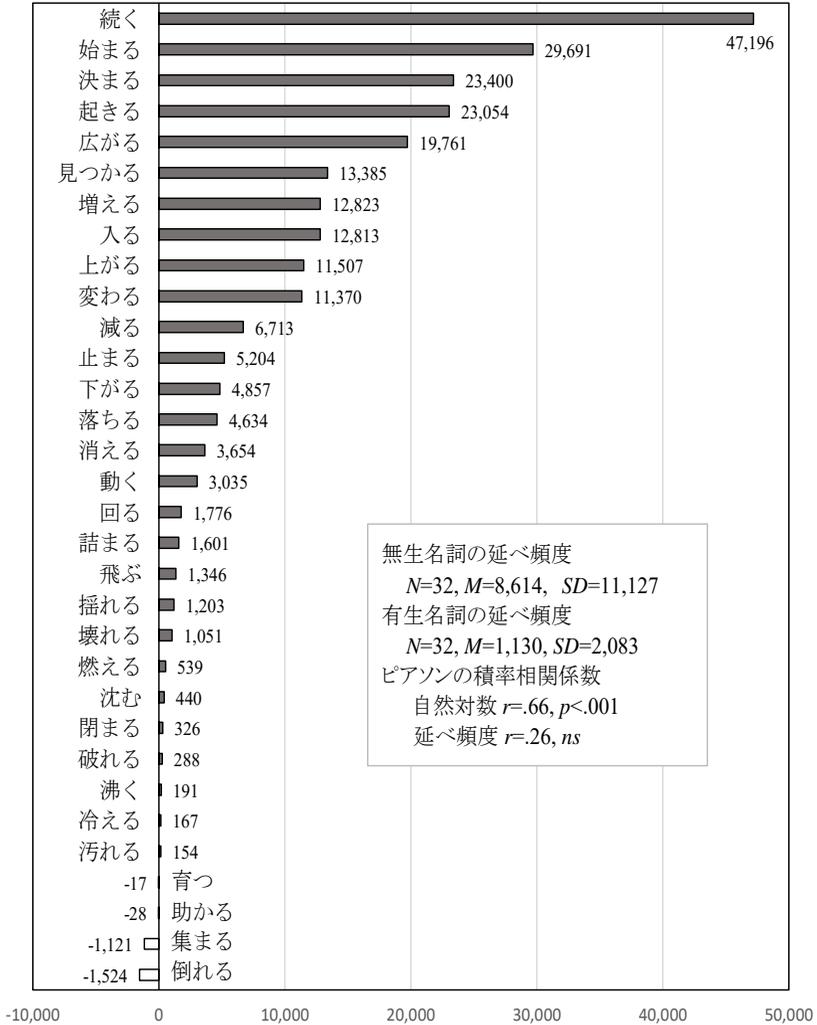
表 1 自動詞 32 語の主語になる無生または有生の名詞の延べ頻度および差

No.	自動詞	延べ頻度 token frequency			自然対数変換 $\log_e(x + 0.5)$		
		無生名詞	有生名詞	差	無生名詞	有生名詞	差
1	上がる	11,929	422	11,507	9.39	6.05	3.34
2	集まる	7,289	8,410	-1,121	8.89	9.04	-0.14
3	動く	3,769	734	3,035	8.23	6.60	1.64
4	起きる	23,201	147	23,054	10.05	4.99	5.06
5	落ちる	4,923	289	4,634	8.50	5.67	2.83
6	変わる	12,063	693	11,370	9.40	6.54	2.86
7	消える	4,211	557	3,654	8.35	6.32	2.02
8	決まる	25,771	2,371	23,400	10.16	7.77	2.39
9	壊れる	1,080	29	1,051	6.99	3.38	3.60
10	下がる	4,900	43	4,857	8.50	3.77	4.72
11	沈む	528	88	440	6.27	4.48	1.79
12	閉まる	326	0	326	5.79	-0.69	6.48
13	育つ	779	796	-17	6.66	6.68	-0.02
14	倒れる	793	2,317	-1,524	6.68	7.75	-1.07
15	助かる	72	100	-28	4.28	4.61	-0.33
16	続く	48,434	1,238	47,196	10.79	7.12	3.67
17	詰まる	1,613	12	1,601	7.39	2.53	4.86
18	飛ぶ	1,687	341	1,346	7.43	5.83	1.60
19	止まる	5,275	71	5,204	8.57	4.27	4.30
20	入る	18,287	5,474	12,813	9.81	8.61	1.21
21	始まる	29,768	77	29,691	10.30	4.35	5.95
22	冷える	167	0	167	5.12	-0.69	5.81
23	広がる	19,879	118	19,761	9.90	4.77	5.12
24	増える	20,468	7,645	12,823	9.93	8.94	0.98
25	減る	8,056	1,343	6,713	8.99	7.20	1.79
26	回る	3,068	1,292	1,776	8.03	7.16	0.86
27	見つかる	14,827	1,442	13,385	9.60	7.27	2.33
28	燃える	573	34	539	6.35	3.54	2.81
29	破れる	288	0	288	5.66	-0.69	6.36
30	揺れる	1,210	7	1,203	7.10	2.01	5.08
31	沸く	252	61	191	5.53	4.12	1.41
32	汚れる	154	0	154	5.04	-0.69	5.73
平均(M)		8,614	1,130	7,484	7.93	4.96	2.97
標準偏差(SD)		11,127	2,083	10,767	1.77	2.76	2.09
対応のあるサンプルの t 検定		延べ頻度: $t(31)=3.87, p<.001$			自然対数: $t(31)=7.93, p<.001$		

計方法である。2つの群に対応関係がある場合に「対応のあるサンプル」の t 検定が使われる。たとえば、表1の「上がる」であれば、自然対数で無生名詞は9.39、有生名詞は6.05であり、両者の差は3.34である。各自動詞について無生名詞と有生名詞が主語となるので、同じ動詞で主語の有生性に対応がある関係と仮定される。対応のあるサンプルの t 検定による分析の結果、無生名詞のほうが有生名詞よりも有意に頻度が高かった [$t(31)=7.93, p<.001$]。自然対数に変換していない延べ頻度についても、対応のあるサンプルの t 検定を行ったが、やはり無生名詞のほうが有生名詞よりも有意に頻度が高かった

[$t(31)=3.87, p<.001$]. 自然対数に変換した値でも、延べ頻度そのままの値でも、自動詞32語において、無生名詞のほうが有生名詞よりも主語になりやすいことを実証した。

また、無生名詞と有生名詞の延べ頻度の差を、図1のように棒グラフで描いた。図1か



無生名詞の延べ頻度－有生名詞の延べ頻度＝差

図1 自動詞32語の主語となる有生名詞と無生名詞の延べ頻度の差

ら分かるように、「続く」の主語の例では、「観測が・・・続く」や「対立が・・・続く」などの無生名詞（48,434回）のほうが「漁師が・・・続く」や「社長が・・・続く」などの有生名詞（1,238回）より47,196回も頻度が高かった。同様に、自動詞の「始まる」「決まる」「起きる」「広がる」では、「制度が・・・始まる」、「優勝が・・・決まる」、「事件が・・・起きる」、「動きが・・・広がる」のように無生名詞が主語になる場合が多い。逆の傾向になったのは、「育つ」「助かる」「集まる」「倒れる」の4つの自動詞だけで、「選手が・・・育つ」、「人が・・・助かる」、「仲間が・・・集まる」、「首相が・・・倒れる」などのような有生名詞の主語のほうが、「情熱が・・・育つ」、「家計が・・・助かる」、「物資が・・・集まる」、「棚が・・・倒れる」などのような無生名詞の主語よりも頻度が高いという結果であった。

ただし、例外的な自動詞の「育つ」「助かる」「集まる」「倒れる」の4つにおける無生名詞と有生名詞の頻度の差は、他の自動詞の頻度の差と比べて大きくない。図1からも、無生名詞が主語になることのほうが多いことが容易に捉えられる。

3. 1. 2 自動詞32語の主語となる名詞の無生と有生についての重なり頻度の差

他対応のある32対の自動詞についてガ格で主語となる無生名詞と有生名詞の重なり頻度および無生名詞から有生名詞の重なり頻度を引いた差を表2に示した。

重なり頻度も、自然対数 $\log_e(x + 0.5)$ に変換した (Yamamura 1999)。自動詞32語について、主語となる無生名詞の平均頻度は547回で、標準偏差は574回であった。有生名詞の平均頻度は80回で、標準偏差が104回であった。32語の自動詞のそれぞれについて無生名詞と有生名詞の頻度が存在するので、両名詞群は対応のあるサンプルであると考えられる。そこで、自然対数に変換した値を使って、対応のあるサンプルの t 検定で無生名詞と有生名詞の頻度の差を検討した。その結果、無生名詞のほうが有生名詞よりも有意に頻度が高かった [$t(31)=9.64, p<.001$]。変換していない重なり頻度についても、対応のあるサンプルの t 検定を行った。やはり、無生名詞のほうが有生名詞よりも有意に頻度が高かった [$t(31)=5.00, p<.001$]。

さらに、自動詞32語について、無生名詞と有生名詞の自然対数に変換した値でピアソンの積率相関係数を計算したが、非常に高く ($N=32, r=.71, p<.001$)、この係数は有意であった。変換していない重なり頻度をそのまま使った場合のピアソンの積率相関係数も、やはり高い相関を示した ($N=32, r=.58, p<.001$)。自動詞32語について、各動詞に付く

無生名詞と有生名詞の主語の頻度パターンは非常に類似していることが分かる。重なり頻度では、「続く」は有生名詞の使用頻度よりも無生名詞の2,084回だけ使用頻度が高かった。自動詞32語中で31語が同じ傾向を示した。逆の傾向になったのは、「倒れる」が55回、「助かる」が1回だけで、有生名詞のほうが無生名詞よりも頻度が高いという結果であった。これらの差も他の自動詞の頻度差と比べると小さい。

表2 自動詞32語の主語になる無生・有生名詞の重なり頻度および差

#	自動詞	重なり頻度 type frequency			自然対数変換 $\log_e(x + 0.5)$		
		無生名詞	有生名詞	差	無生名詞	有生名詞	差
1	上がる	469	73	396	6.15	4.30	1.85
2	集まる	429	269	160	6.06	5.60	0.47
3	動く	472	105	367	6.16	4.66	1.50
4	起きる	657	40	617	6.49	3.70	2.79
5	落ちる	456	62	394	6.12	4.14	1.99
6	変わる	911	79	832	6.82	4.38	2.44
7	消える	597	66	531	6.39	4.20	2.20
8	決まる	1,157	80	1,077	7.05	4.39	2.67
9	壊れる	211	8	203	5.35	2.14	3.21
10	下がる	258	14	244	5.55	2.67	2.88
11	沈む	85	24	61	4.45	3.20	1.25
12	閉まる	32	0	32	3.48	-0.69	4.17
13	育つ	147	79	68	4.99	4.38	0.62
14	倒れる	122	177	-55	4.81	5.18	-0.37
15	助かる	13	14	-1	2.60	2.67	-0.07
16	続く	2,249	165	2,084	7.72	5.11	2.61
17	詰まる	203	5	198	5.32	1.70	3.61
18	飛ぶ	228	68	160	5.43	4.23	1.20
19	止まる	386	23	363	5.96	3.16	2.80
20	入る	1,930	503	1,427	7.57	6.22	1.34
21	始まる	1,539	20	1,519	7.34	3.02	4.32
22	冷える	33	0	33	3.51	-0.69	4.20
23	広がる	1,108	20	1,088	7.01	3.02	3.99
24	増える	1,437	247	1,190	7.27	5.51	1.76
25	減る	668	95	573	6.51	4.56	1.95
26	回る	333	183	150	5.81	5.21	0.60
27	見つかる	917	115	802	6.82	4.75	2.07
28	燃える	115	11	104	4.75	2.44	2.31
29	破れる	50	0	50	3.92	-0.69	4.62
30	揺れる	232	3	229	5.45	1.25	4.20
31	沸く	36	10	26	3.60	2.35	1.25
32	汚れる	35	0	35	3.57	-0.69	4.26
平均(M)		547	80	467	5.63	3.29	2.33
標準偏差(SD)		573	104	520	1.33	1.91	1.35
対応のあるサンプルのt検定		重なり頻度: $t(31)=5.00, p<.001$			自然対数: $t(31)=9.64, p<.001$		

3. 1. 3 自動詞における主語となる無生と有生名詞の頻度の比較結果

自動詞32語の延べ頻度および重なり頻度を自然対数に変換した値でも、延べ頻度および重なり頻度でも、対応のあるサンプルの検定で0.1パーセントという低い危険率で、無生名詞が有生名詞よりも主語になりやすいことが示された。個々の自動詞をみても、有生名詞が無生名詞より主語になることが多かったのは自動詞32語のなかで、延べ頻度では4つで、重なり頻度では2つだけであった。自動詞32語を検討した限り、無生名詞が有生名詞よりも自動詞の主語になりやすいという仮説を支持した。

3. 2 他動詞32語についての分析

3. 2. 1 他動詞32語の主語となる名詞の無生と有生についての延べ頻度の違いの検討

自動詞32対と対応のある他動詞32語についても、ガ格を取って無生と有生の主語となる名詞の延べ頻度および無生名詞から有生名詞の延べ頻度を引いた数値を算出して、表3に示した。

他動詞「冷やす」について、無生名詞が主語になる頻度が0になっているので、自動詞の場合と同様に、 $\log(x + 0.5)$ の自然対数に変換して分析した。他動詞32語については、「舞台が・・・消す」「政権が・・・決める」「ゴミが・・・汚す」などの無生名詞が主語となる場合の平均頻度は797回で、標準偏差は1,338回であった。一方、「魚が・・・消す」「若者が・・・決める」「子供が・・・汚す」などの有生名詞が主語となる場合の平均頻度は642回で、標準偏差が1,067回であった。

32語の他動詞の自然対数に変換した値を使って、無生名詞と有生名詞の頻度の差を対応のあるサンプルの検定で検討した。その結果、主語となる無生名詞と有生名詞の頻度に有意な違いはみられなかった [$t(31)=1.44, ns$]。変換していない延べ頻度についても対応のあるサンプルの検定を行った。その結果、主語となる無生名詞と有生名詞の頻度には有意な違いはみられなかった [$t(31)=1.58, ns$]。したがって、他動詞については、日本語では無生名詞が主語になるという傾向は他動詞では観察されなかった。無生名詞と有生名詞の延べ頻度の差を、図2のように棒グラフで描いた。

他動詞32語の自然対数に変換した延べ頻度の無生名詞と有生名詞のピアソンの積率相関係数は、きわめて高く ($N=32, r=.92, p<.001$)、この係数は有意であった。変換していない延べ頻度をそのまま使った場合のピアソンの積率相関係数も、やはりきわめて高い相関を示した ($N=32, r=.88, p<.001$)。他動詞32語について、無生名詞と有生名詞

の頻度の各動詞の頻度パターンは、非常に類似していることが分かる。

表3 他動詞32語の主語になる無生名詞と有生名詞の延べ頻度および差

#	他動詞	延べ頻度 token frequency			自然対数変換 $\log_e(x + 0.5)$		
		無生名詞	有生名詞	差	無生名詞	有生名詞	差
1	上げる	1,260	1,198	62	7.14	7.09	0.05
2	集める	1,627	919	708	7.39	6.82	0.57
3	動かす	259	195	64	5.56	5.28	0.28
4	起こす	1,376	1,900	-524	7.23	7.55	-0.32
5	落とす	665	574	91	6.50	6.35	0.15
6	変える	993	409	584	6.90	6.01	0.89
7	消す	317	283	34	5.76	5.65	0.11
8	決める	6,967	5,510	1,457	8.85	8.61	0.23
9	壊す	44	75	-31	3.80	4.32	-0.53
10	下げる	257	143	114	5.55	4.97	0.58
11	沈める	17	10	7	2.86	2.35	0.51
12	閉める	36	12	24	3.60	2.53	1.07
13	育てる	169	374	-205	5.13	5.93	-0.79
14	倒す	47	53	-6	3.86	3.98	-0.12
15	助ける	64	193	-129	4.17	5.27	-1.10
16	続ける	2,067	1,365	702	7.63	7.22	0.41
17	詰める	206	162	44	5.33	5.09	0.24
18	飛ばす	29	79	-50	3.38	4.38	-0.99
19	止める	1,280	569	711	7.16	6.34	0.81
20	入れる	769	988	-219	6.65	6.90	-0.25
21	始める	3,038	2,027	1,011	8.02	7.61	0.40
22	冷やす	31	0	31	3.45	-0.69	4.14
23	広げる	581	284	297	6.37	5.65	0.71
24	増やす	438	79	359	6.08	4.38	1.71
25	減らす	359	51	308	5.88	3.94	1.94
26	回す	48	45	3	3.88	3.82	0.06
27	見つける	333	2,082	-1,749	5.81	7.64	-1.83
28	燃やす	14	39	-25	2.67	3.68	-1.00
29	破る	2,157	899	1,258	7.68	6.80	0.87
30	揺らす	49	9	40	3.90	2.25	1.65
31	沸かす	2	2	0	0.92	0.92	0.00
32	汚す	4	17	-13	1.50	2.86	-1.36
	平均(M)	797	642	155	5.33	5.05	0.29
	標準偏差(SD)	1,338	1,067	547	1.94	2.09	1.10
	対応のあるサンプルのt検定	延べ頻度: $t(31)=1.58, ns$			自然対数: $t(31)=1.44, ns$		

また、図2から分かるように、「見つける」は、無生名詞の主語のほうが有生名詞の主語よりも1,749回だけ頻度が高かった。一方、「決める」は、逆に有生名詞の主語のほうが無生名詞の主語よりも1,457回だけ頻度が高い。図2の分布からも、無生名詞が主語に

なることが多い他動詞から、有生名詞が主語になることが多い他動詞まで均等に広がっており、他動詞32語の主語となる無生名詞と有生名詞の頻度は、個々の他動詞で違いが大きく、全体としてみると、違いがないことが視覚的にも見てとれる。

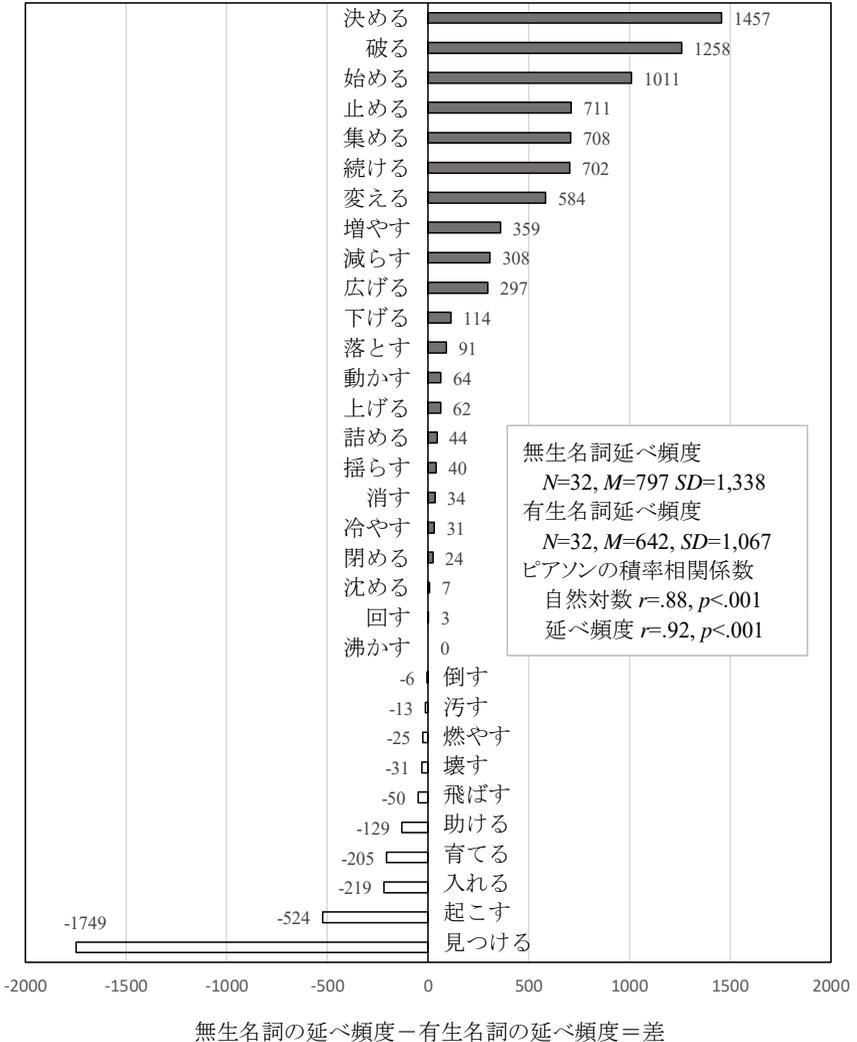


図2 他動詞32語の主語となる有生名詞と無生名詞の延べ頻度の差

3. 2. 2 他動詞32語の主語となる名詞の無生と有生についての重なり頻度の差

他動詞32語について、主語となる無生名詞と有生名詞の重なり頻度および無生名詞から有生名詞の重なり頻度を引いた差を、表4に示した。「沸かす」の無生名詞が0回であるため、重なり頻度も自然対数 $\log_e(x+0.5)$ に変換した(Yamamura 1999)。念のために、重なり頻度の値についても、同じ分析を行った。他動詞32語について、主語となる無生

表4 他動詞 32 語の主語になる無生・有生名詞の重なり頻度および差

#	他動詞	重なり頻度 type frequency			自然対数変換 $\log_e(x+0.5)$		
		無生名詞	有生名詞	差	無生名詞	有生名詞	差
1	上げる	254	173	81	5.54	5.16	0.38
2	集める	349	114	235	5.86	4.74	1.12
3	動かす	67	40	27	4.21	3.70	0.51
4	起こす	240	158	82	5.48	5.07	0.42
5	落とす	139	101	38	4.94	4.62	0.32
6	変える	235	70	165	5.46	4.26	1.21
7	消す	96	52	44	4.57	3.96	0.61
8	決める	908	811	97	6.81	6.70	0.11
9	壊す	17	22	-5	2.86	3.11	-0.25
10	下げる	56	40	16	4.03	3.70	0.33
11	沈める	3	5	-2	1.25	1.70	-0.45
12	閉める	14	4	10	2.67	1.50	1.17
13	育てる	45	60	-15	3.82	4.10	-0.28
14	倒す	15	19	-4	2.74	2.97	-0.23
15	助ける	19	41	-22	2.97	3.73	-0.76
16	続ける	300	139	161	5.71	4.94	0.77
17	詰める	32	42	-10	3.48	3.75	-0.27
18	飛ばす	9	21	-12	2.25	3.07	-0.82
19	止める	77	74	3	4.35	4.31	0.04
20	入れる	142	124	18	4.96	4.82	0.14
21	始める	401	177	224	6.00	5.18	0.82
22	冷やす	10	1	9	2.35	0.41	1.95
23	広げる	132	51	81	4.89	3.94	0.95
24	増やす	85	22	63	4.45	3.11	1.34
25	減らす	78	9	69	4.36	2.25	2.11
26	回す	15	13	2	2.74	2.60	0.14
27	見つける	47	128	-81	3.86	4.86	-1.00
28	燃やす	5	12	-7	1.70	2.53	-0.82
29	破る	343	140	203	5.84	4.95	0.89
30	揺らす	7	3	4	2.01	1.25	0.76
31	沸かす	1	1	0	0.41	0.41	0.00
32	汚す	2	5	-3	0.92	1.70	-0.79
平均(M)		129	84	46	3.86	3.53	0.33
標準偏差(SD)		182	142	76	1.62	1.48	0.78
対応のあるサンプルのt検定		重なり頻度: $t(31)=3.39, p<.01$			自然対数: $t(31)=2.32, p<.05$		

名詞の重なり頻度の平均は129回で、標準偏差は182回であった。有生名詞の平均頻度は84回で、標準偏差が142回であった。

他動詞32語のそれぞれについての無生名詞と有生名詞の頻度について、無生と有生の名詞群に対応があると考えられるので、自然対数に変換した値を使って、対応のあるサンプルの t 検定で無生名詞と有生名詞の頻度を検討した。その結果、無生名詞のほうが、有生名詞よりも有意に頻度が高かった [$t(31)=2.32, p<.05$]。変換していない重なり頻度についても、対応のあるサンプルの t 検定を行った。やはり、無生名詞のほうが有生名詞よりも有意に頻度が高かった [$t(31)=3.39, p<.01$]。

さらに、他動詞32語について、自然対数について無生名詞と有生名詞のピアソンの積率相関係数を算出した。相関は非常に高く ($N=32, r=.88, p<.001$)、この係数は有意であった。変換していない重なり頻度をそのまま使った場合のピアソンの積率相関も、やはり高い相関を示した ($N=32, r=.92, p<.001$)。これは、自動詞32語の自然対数の相関と同様に、他動詞32語についても、各動詞の主語として付く無生名詞と有生名詞の頻度パターンが非常に類似していることを示している。

3. 2. 3 他動詞32語の主語となる名詞の無生と有生についての重なり頻度の差

他動詞32語の延べ頻度については、自然対数に変換した値でも、延べ頻度の値でも、対応のあるサンプルの t 検定で、主語になる無生名詞と有生名詞の頻度に違いはなかった。一方、重なり頻度については、対応のあるサンプルの t 検定では、無生名詞のほうが有生名詞より主語になる頻度が高かった。しかし、延べ頻度が、実際の言語的特性を反映していると考えられるので、重なり頻度ではなく、延べ頻度の結果から、無生名詞と有生名詞で他動詞32語の主語になる頻度には違いがないと判断する。

3. 3 自他对応のある32語の動詞の無生・有生主語の延べ頻度比較

最後に、無生名詞と有生名詞の主語としての共起頻度が自動詞と他動詞で異なることを実証するために、2 (自他動詞) \times 2 (無生・有生名詞) の2つの変数の交互作用が有意であるかどうかを確認することにした。自動詞と他動詞は、活用形が異なることが多く、また自他对応のない動詞群も多い。たとえば、使用頻度の高い36対の自他对応動詞のうち活用形が一致したのはわずかに8対 (22.22%) で、28対 (77.78%) については、お互いに違った活用形である (玉岡・張・牧岡, 2018)。Tsujiura (2014) が Jacobsen

(1992)の研究を基に指摘したように、日本語の自他対応動詞の形態素変化については、明確な活用形の規則性がみられない。そのため、基本的には異なる動詞であると考えられるので、自動詞と他動詞については、たとえ自他対応があっても、独立したサンプルであるとした(反復なし)。一方、主語の有生・無生名詞は、同じ動詞について存在しているので、対応があるとした(反復あり)。そして、2(動詞の種類:自動詞・他動詞で反復なし)×2(主語:有生名詞・無生名詞で反復あり)の分散分析を行った。

延べ頻度を自然対数に変換した値を使った分析の結果は、自動詞・他動詞の違い $[F(1, 62)=6.03, p<.05]$ および無生・有生名詞の違い $[F(1, 62)=59.05, p<.001]$ の主効果が有意であった。また、両変数の交互作用も有意であった $[F(1, 62)=40.20, p<.001]$ 。延べ頻度の値についても同じ分析を行った。その結果も、自動詞・他動詞の違い $[F(1, 62)=14.65, p<.001]$ と無生・有生名詞の違い $[F(1, 62)=15.56, p<.001]$ の主効果が有意であり、両変数の交互作用も有意であった $[F(1, 31)=14.33, p<.001]$ 。この分析により、主語になる無生名詞と有生名詞の頻度パターンが、自動詞と他動詞で異なることおよび無生名詞と有生名詞の頻度の有意な違いがあり、それが自動詞と他動詞で異なるパターンであることが示された。自動詞と他動詞の無生名詞と有生名詞の違いについては、これまで両動詞を別々に分析した通りである。

4 総合考察

本研究では、2つの仮説を立てた。第1に、日本語では、有生名詞よりも無生名詞が主語になりやすいという仮説(Hinds 1986; 国広 1974a, 1974b)である(仮説1)。第2に、日本語では、他動詞よりも自動詞で無生名詞が主語になりやすいと言われている(池上 1981, 2006; 西光 2010; 寺村 1976)。そのため、仮説1は、自動詞のみにみられる傾向ではないかという仮説である(仮説2)。これらの2つの仮説を検討するために、本研究では、毎日新聞18年分の大規模コーパスで、主語との延べ共起頻度が150回以上を基準として自動詞32語(合計311,791文)を選んだ。他動詞としては、これらの32の自動詞に対応する他動詞32語(合計46,048文)を選んだ。そして、主語になる無生名詞か有生名詞の頻度を調べて、自動詞と他動詞を別々に検討した。

仮説1の自動詞では無生名詞が有生名詞よりも主語になりやすいことを検証するために、延べ頻度および重なり頻度の両方について、自然対数に変換した値と頻度をそのまま使った値で対応のあるサンプルの検定を行ったが、0.1パーセントという低い危険率

(表1と表2の*t*検定の結果を参照)で無生名詞が有生名詞に比べて出現頻度が高いことが実証された。つまり、自動詞においては、仮説1の日本語では無生名詞が主語になりやすいという仮説 (Hinds 1986 ; 国広 1974a, 1974b) を支持した。

さらに、仮説2の無生名詞が有生名詞よりも主語になりやすいのは、自動詞のみの傾向であるかどうかを検証するために、他動詞32語についても、自動詞32語と同じ分析を行った。対応のあるサンプルの*t*検定で分析した結果、自然対数に変換した延べ頻度でも延べ頻度そのものの値でも、主語になる無生名詞と有生名詞の頻度に違いはなかった(表3を参照)。一方、重なり頻度については、同じ対応のあるサンプルの*t*検定を行った(表4を参照)が、無生名詞のほうが有生名詞より主語になる頻度が高かった。延べ頻度のほうが実際の言語的特性を反映していると考えられるので、結論として、他動詞では無生名詞と有生名詞で主語になる頻度には違いがないと考えてよいであろう。

最後に、32語の動詞の無生・有生主語の延べ頻度が自動詞と他動詞で異なった頻度パターンであるかどうかを、両方の頻度データを一緒にした動詞の種類(自動詞・他動詞)と主語(無生名詞・有生名詞)の二元配置の分散分析で確認した。分析の結果、動詞の自動詞・他動詞と主語の無生・有生名詞の交互作用が有意であった。これにより、自動詞と他動詞の主語になる無生名詞と有生名詞の頻度パターンが異なっていることをさらに裏づけた。また、自動詞32語について無生名詞と有生名詞の頻度の差を図1に、他動詞32語については図2に描いた。無生名詞が有生名詞より主語になりやすいのは自動詞のみの傾向であることが、視覚的にも確認できる。

以上のように、本研究は、過去の直感的な議論に対して、日本語では無生名詞が有生名詞よりも主語になりやすく、その傾向は自動詞に限定されることを、大規模コーパス研究で実証した。

注

- 1) 「くろしお言語大学塾」に収録されたゼミ『『する』言語と『なる』言語を考え直す』(講師：西光義弘)の第3回講義「3.人間中心と状況中心—国広哲弥氏の研究から—」(http://www.gengoj.com/seminar/view.php?seminar_list_id=2 (2019年4月12日アクセス))を参照した。
- 2) 表1の自然対数に変換した値の無生名詞と有生名詞の差は、少数13位で計算しているのので、表1の少数第2位までの値で計算すると0.1のズレが生じている。たとえば、「集まる」であ

れば、 $8.89-9.0 = -0.15$ になるが、小数13位の計算では、 -0.14 となる。同様に、「育つ」「倒れる」の値も同様である。

参考文献

- Frei, Henri (1953) *Le livre des deux mille phrases*. Genève: Librairie Droz.
- Hinds, John (1986) *Situation vs. person focus* [日本語らしさと英語らしさ]. くろしお出版.
- Jacobsen, Wesley (1992) *The Transitive Structure of Events in Japanese*. Tokyo: Kurocio.
- Tsujimura, Natsuko (2014) *An introduction to Japanese linguistics*. West Sussex, UK: John Wiley & Sons.
- Yamamura, Kohji (1999) Transformation using $(x+0.5)$ to stabilize the variance of populations. *Researches on Population Ecology*, 41: 229-234.
- 早津恵美子 (1989) 「有対他動詞と無対他動詞の違いについて」『言語研究』95: 231-256.
- 早津恵美子 (1995) 「有対他動詞と無対他動詞の違いについて」須賀一好・早津恵美子 (編) 『動詞の自他』(pp. 179-197), 東京: ひつじ書房.
- 池上嘉彦 (1981) 『「する」と「なる」の言語学—言語と文化のタイポロジーへの試論—』大修館書店.
- 池上嘉彦 (2006) 『英語の感覚・日本語の感覚—〈ことばの意味〉のしくみ—』NHK出版.
- 工藤拓・松本裕治 (2002) 「チャンキングの段階適用による係り受け解析」『情報処理学会論文誌』43(6): 1834-1842.
- 工藤拓・山本薫・松本裕治 (2004) 「Conditional Random Fields を用いた日本語形態素析」『情報処理学会研究報告 自然言語処理 (NL161)』47: 89-96.
- 国広哲弥 (1974a) 「人間中心と状況中心—日英語表現構造の比較—」『英語青年』2: 688-690.
- 国広哲弥 (1974b) 「日英語表現体系の比較」『言語生活』3: 46-52.
- 寺村秀夫 (1976) 『「ナル」表現と「スル」表現—日英「態」表現の比較—』国語シリーズ別冊4 『日本語と日本語教育—文字・表現編』49-68, 国立国語研究所.
- 西光義弘 (2010) 「他動性は連続体か?」西光義弘, ブラシャント・パルデシ (編) 『(シリーズ言語対照 〈外から見る日本語〉 第4巻) 自動詞・他動詞の対照』(pp. 211-234), くろしお出版.
- 玉岡賀津雄・張婧禕・牧岡吾吾 (2018) 「日本語自他対応動詞36対の使用頻度の比較」『計量国語学』31: 443-460.