The University of Osaka
Institutional Knowledge Archive

Title	一対比較法による日本語名詞句階層の測定
Author(s)	麻, 子軒
Citation	現代日本語研究. 2018, 10, p. 66-80
Version Type	VoR
URL	https://doi.org/10.18910/70105
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

一対比較法による日本語名詞句階層の測定

An Attempt to Determine Noun-phrase Hierarchy
in Japanese by Using Paired Comparison

麻　子軒
MA Tzu-Hsuan

キーワード：名詞句階層，一対比較法，他動性，コーパス，計量的文法研究

要　旨

『現代日本語書き言葉均衡コーパス』を調査資料に，Silverstein（1976）の名詞句階層を「一対比較法（Paired Comparison)」で測定することを試み，さらにその分析結果をSilversteinのオリジナルの階層と比較することによって，一対比較法の有効性を検討した。結論として，無生物名詞に関してはSilverstein とほぼ同じ結果を得たが，有生物名詞に関して，特に人名詞同士の階層は明快に決められるものではない可能性があることを示した。また，分析の結果が安定するまではある程度のサンプル数が必要だが，それが足りていれば理論的に一対比較法で名詞句階層を量的に測定することは可能であることも分かった。

１．はじめに

本稿の目的は，Silverstein（1976）の「名詞句階層」という概念を日本語のコーパスで収集した用例で量的に測定し，その結果をSilversteinが提出したオリジナルの名詞句階層と比較することである。測定の方法として，「一対比較法（Paired Comparison)」という統計手法を用い，最後にその有効性についても検討する。調査資料には『現代日本語書き言葉均衡コーパス[1)](以下BCCWJ)』の「出版・書籍」サブコーパスを用いる。

名詞句階層は，他動性に関わるさまざまな文法現象を説明するために提案されたものである。これまで，このような文法の概念は内省によって研究される

ことが多く，計量的な方法でそれらを客観的に測定する試みはまだ少ない。本稿では，名詞句階層の測定を計量的文法研究の１つの事例として取り上げ，今後のこの領域における発展の足掛かりとしたい。

２．先行研究

名詞句階層という概念は，Silverstein（1976）によって提出されたもので，当初はオーストラリア原住民に見られる能格型格組織とそれ以外の対格型などの格組織の共存の仕方を説明するためのものであったが，後に文化人類学などの言語学以外の分野にも応用されている。その概略を図１に示す。左にある「１人称」が階層上最も高い位置を占め，右に行けば行くほどその階層性が徐々に低くなっていく。

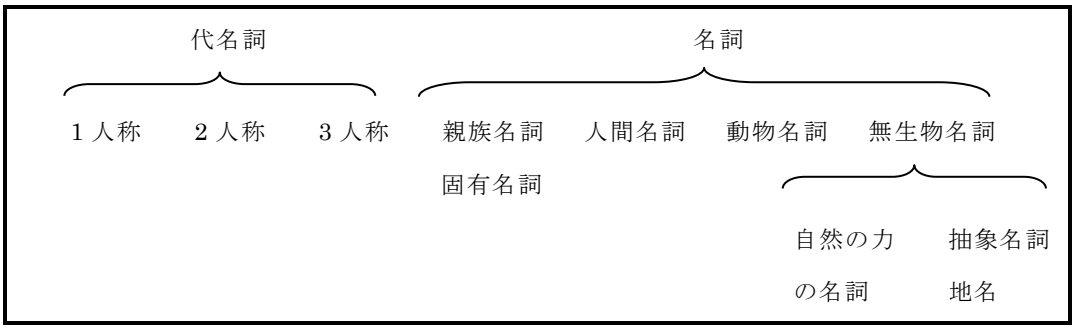

図１：Silverstein の名詞句階層

この階層が果たして何を表わしているのかについてはさまざまな説がある。以下，角田（1991）より引用する。

> Silverstein 自身は，この階層は動作者になりやすさの度合，動作の対象になりやすさの度合を表わすという。Dixon（1979）は，この階層は話し手にとっての重要さの程度を表わすという。Zubin（1979）は話し手の egocentrism 自己中心性を表わすと見ている。また，Wierzbicka（1981）は，話題になりやすさの度合，或いは，話し手にとっての身近さの度合を表わすと言っている。

角田自身は，Silverstein の立場に立って，名詞句階層を用いて日本語の受身文や無生物主語他動詞文の成立可否を説明している（角田 1991）。例えば，

「私が女を殺した」は言えるものの，「女は私に殺された」「女は私を殺した」が言いにくいのは，主語である名詞句が階層的に目的語のそれよりも上にあることが原則だからだ，という説明を行っている。しかし，そもそも名詞句階層という概念を無批判に日本語に適用していいものか,疑問に思われる。確かに，名詞句階層はさまざまな言語に見られる現象であるが，それがすべてに共通するものなのか，それとも一部にのみ共通するものなのかについては，これまで客観的な検証がなされていない。

ここで，これまでの先行研究の問題点をまとめると，①異なる言語が同じ名詞句階層の順序をもっていると最初から仮定すること，②名詞句階層を内省のみによって順序付けること，③各名詞句の間の順序のみを問題とし，互いの差の程度を考慮していないこと，以上の3点である。本稿では，一対比較法という統計手法を採用し，コーパスから収集した用例を量的に分析することで，日本語の名詞句階層の順序とその差の程度を測定することを試みる。その後，Silverstein が提出したオリジナルの名詞句階層と比較し，一対比較法という方法の有効性について検討する。一対比較法には,各項目間の順位だけでなく，差の程度も測定できるメリットがあるからである。

３．調査概要

調査資料はBCCWJの「出版・書籍」である。このサブコーパスを利用するのは，無生物主語他動詞文の名詞句階層に関する調査を行った熊（2009）が，新聞と文学作品とではその様相が異なると指摘しているからである。Silverstein が名詞句階層を設定した際，レジスターの違いについては言及していないが，熊の考察で示されたように，実際はその違いによって名詞句階層の順序が変わる可能性がある。もちろん,「出版・書籍」サブコーパスは日本語全体を代表するものではないが，書きことばの主要なレジスターであることから，まずはこれを調査すべきだと思われる。また，本稿の調査結果を踏まえ，今後は「出版・新聞」など，異なるサブコーパスとの比較もできる。

調査対象とするのは,「ＸガＹヲＺ」のような他動詞構文である（Ｘは主語，Ｙは目的語，Ｚは動詞を意味する)。自動詞・他動詞の定義は奥津（1967）に従う。そのため,「～を出る」「～を走る」などはヲ格が対象を表わすものではな

いため自動詞文とする。なお，受身文や慣用句，機能動詞を伴った文も対象外とする。受身文を除く理由は，その主語と目的語の配置が能動文と逆になっているからである。また，慣用句と機能動詞の構文を除外したのは，「ＹヲＺ」がすでに１つの陳述的意味をもっており，「主語＝Ｘ」と「目的語＝Ｙ」との両者を単純に比較できないからである。

以上の規定に従って収集した用例は 22755 例である。用例数が多いため，無作為に全体から一部を取り出して分析する。サンプル数に関しては，比較的安定した結果が得られるまで試行錯誤を繰り返した結果，最終的に 5000 例を調査対象とする[2)]。

４．研究方法

測定方法として，官能評価試験でよく用いられる「一対比較法」を用いる。天坂・長沢（2007）によると，一対比較法は任意の２つの対象を取り出して 1 対 1 で比較し，すべての比較結果を統合して評価を行う方法で，食べ物の美味しさやデザインの良さなどを比較する際に用いられる。通常，k 個の試料を比較しようとするとき,すべての試料を一度に順位付けることは困難であるため，代わりに k 個から２つずつ取り出して対にして比較し，全体の結果を総合的に評価するという方法である。例えば，A～D の４つの試料に対して順位を付けたい場合は，表１のように，A～D から２つを取り出す場合のすべての組み合わせについて被験者全員に選択させる。

表１：一対比較法の組み合わせ例

	A vs B	A vs C	A vs D	B vs C	B vs D	C vs D
被験者 1	B	A	A	C	B	C
被験者 2	B	C	A	C	B	D
被験者 3	A	A	A	B	D	C
…	…	…	…	…	…	…

一対比較法には，順位のみを付ける方法と，順位に加えてその差の程度も算出する方法との 2 種類があるが，本稿では順位だけでなく，差の程度も測定したいため，後者を選ぶ。これにはさらにサーストン法，シェッフェの原法など

の方法があるが，今回の調査では用例収集の制限上，サーストン法を用いる。その理由は，収集したデータで各組み合わせに対する試験回数が同じではないからである[3)]。

このように一対比較法では，２つの選択肢を複数の受験者に提示し，どちらがより程度性が高い（例えば美味しい）のかを答えさせるが，本稿では，２つの名詞を比較する際，どちらが名詞句階層において上位であるかを判断する手掛かりとして，実際の他動詞文でその名詞句が主語か目的語のどちらに立つかを見るという方法をとる。名詞句階層で相対的に程度性が強いものは目的語よりも主語になりやすいという角田の説明に従えば，順序として「主語の名詞句＞目的語の名詞句」と判断することができるからである。

表２：名詞の分類項目と定義

分類項目	定義	例
１人称	話し手，または書き手を表わす代名詞	私，僕，私たち
２人称[4)]	聞き手，または読み手を表わす代名詞	あなた，君，おまえ
３人称	話し手（書き手）と聞き手（読み手）以外の人を表わす代名詞	彼，彼女，彼ら
人間名詞	１人称，２人称，３人称，固有名詞，親族名詞以外の人名詞	人，女性，仙台市民
地名[5)]	他の場所から特定の地点や区域を区別するために，それのみに付けられた名称を表わす名詞	アメリカ，東京，名古屋駅
自然の力の名詞	自然の力によって起きた現象を表わす名詞	火山，地震，雷
固有名詞	他の人間から特定の人物を区別するために，それのみに付けられた名称を表わす人名詞	佐藤さん，金委員長，小林幹事長
抽象名詞	地名，自然の力の名詞，集合名詞以外の抽象概念を表わす名詞	怒り，責任，戦争
物名詞	無生物である具体物を表わす名詞	机，椅子，本
動物名詞	人間以外の有生物を表わす名詞	猫，馬，獅子
集合名詞[6)]	同じような性質をもった人からなる集合体を表す名詞	自民党，中国，委員会
親族名詞	親族や家庭内の関係を表す人名詞	父，母，妻，兄さん

例えば,「彼はタバコを取り出した」という文の場合は,「彼＝３人称」が主語で,「タバコ＝物名詞」が目的語であるため，主語が目的語よりも階層が上であるという仮定に従えば,「３人称＞物名詞」という結論が導かれる。これは，被験者に「彼」と「タバコ」との２つの選択肢を提示し，どちらが名詞句階層において上位かを選択させた場合，被験者が「物名詞」の「タバコ」よりも「３人称」の「彼」の方を選んだのと同じ意味をもつ。言い換えれば，１つの他動詞文の用例は，１回の試験として見なすことができる。

名詞の分類に関しては，Silverstein の結果と比較するため，原則的にそれと同じ分類にするが，その枠組みに入らないものがあるため,「集合名詞」「物名詞」という２つの分類を新たに立てることにする。本稿で用いた名詞の分類項目と定義を前ページの表２に示す。

以下では，本稿で用いたサーストンの一対比較法の手順を説明する。まず収集した他動詞文の用例の主語（X）と目的語（Y）を上の項目で分類し，すべての組み合わせの生の頻度で行列を作成する。その結果が表３である[7)]。

表３：生頻度行列

X＼Y	1人称	2人称	3人称	人間	地名	自然	固有	抽象	物	動物	集合	親族
1人称	4	2	6	24	1	0	10	150	71	2	6	3
2人称	7	1	0	3	0	0	0	20	12	0	2	0
3人称	2	2	4	7	3	0	4	47	22	1	1	0
人間	27	5	7	108	19	2	23	846	428	15	37	3
地名	0	0	0	0	1	0	1	2	0	0	0	0
自然	2	0	0	0	3	0	1	11	24	0	0	0
固有	18	1	4	41	27	0	73	433	412	6	11	4
抽象	11	6	5	35	7	0	27	482	51	1	15	2
物	6	1	3	15	1	3	6	180	184	1	0	0
動物	0	0	0	1	1	1	1	24	60	18	0	0
集合	2	0	1	24	19	0	8	497	79	1	51	0
親族	8	2	1	12	3	0	4	57	56	1	1	13

この表からは，例えば,「１人称」と「抽象名詞」の組み合わせ場合,「抽象名詞＞１人称（X＝主語＝抽象名詞，Y＝目的語＝１人称)」が 11 例のみであ

るのに対し,「１人称＞抽象名詞（Ｘ＝主語＝１人称，Ｙ＝目的語＝抽象名詞）」が 150 例あることが読み取れる。

次に，表３の同じ組み合わせの合計が１になるように，それぞれの欄の値を相対頻度に換算する。その結果が表４である。合計が１であるということは，これらの数値は各組み合わせにおいてどちらの名詞句が階層的に上位かを示す確率でもあることを意味する。

表４：相対頻度行列

X＼Y	1人称	2人称	3人称	人間	地名	自然	固有	抽象	物	動物	集合	親族
1人称	0.500	0.222	0.750	0.471	1.000	0.000	0.357	0.932	0.922	1.000	0.750	0.273
2人称	0.778	0.500	0.000	0.375	0.500	0.500	0.000	0.769	0.923	0.500	1.000	0.000
3人称	0.250	1.000	0.500	0.500	1.000	0.500	0.500	0.904	0.880	1.000	0.500	0.000
人間	0.529	0.625	0.500	0.500	1.000	1.000	0.359	0.960	0.966	0.938	0.607	0.200
地名	0.000	0.500	0.000	0.000	0.500	0.000	0.036	0.222	0.000	0.000	0.000	0.000
自然	1.000	0.500	0.500	0.000	1.000	0.500	1.000	1.000	0.889	0.000	0.500	0.500
固有	0.643	1.000	0.500	0.641	0.964	0.000	0.500	0.941	0.986	0.857	0.579	0.500
抽象	0.068	0.231	0.096	0.040	0.778	0.000	0.059	0.500	0.221	0.040	0.029	0.034
物	0.078	0.077	0.120	0.034	1.000	0.111	0.014	0.779	0.500	0.016	0.000	0.000
動物	0.000	0.500	0.000	0.063	1.000	1.000	0.143	0.960	0.984	0.500	0.000	0.000
集合	0.250	0.000	0.500	0.393	1.000	0.500	0.421	0.971	1.000	1.000	0.500	0.000
親族	0.727	1.000	1.000	0.800	1.000	0.500	0.500	0.966	1.000	1.000	1.000	0.500

例えば，先ほどの例では,「抽象名詞＞１人称」が 11/(11+150)=0.068,「１人称＞抽象名詞」が 150/(11+150)=0.932，両者の合計は１となる。なお，収集した用例の中で存在しない組み合わせ（欠損値）に関してはどちらが上位か判断できないため，仮にそれぞれの確率は等しいと想定し，0.5 として計算する。また，主対角線上の成分は，同じ名詞同士の組み合わせになるため，確率も 0.5 と考える。

続いて，表４の確率を，正規分布表に従ってＺスコアに換算する。その結果が表５になる[8)]。Ｚスコアはデータを標準化した指標で，この値を以て異なるデータの群を同列に扱うことが可能となる。そこで，各列（階層的にＹ＞Ｘのすべての場合）のＺスコアの平均を計算すると，Ｙに対するすべてのＸの名詞

句が与える平均的な影響力を求めることができる。例えば,「Ｙ＝１人称」に対するすべてのＸの名詞句の平均的な影響力は「Ｘ＝１人称＞Ｙ＝１人称」「Ｘ＝３人称＞Ｙ＝１人称」「Ｘ＝人間名詞＞Ｙ＝１人称」……「Ｘ＝親族＞Ｙ＝１人称」のＺスコアを合計し，その値を項目数で割ることで求められる。この例の場合は（0.000+0.765-0.674+0.074-4.350+4.350+0.366-1.488-1.419-4.350-0.674+0.605)/12=-0.566 である。これをすべての列に対して行い，最後に各列の平均に√2 をかければ最終的な評価値が求まる。検算として，各列の平均値を合計すると四捨五入の範囲内で０になるはずだが，表５のデータで試算するとちょうど０になるため，結果が正しく計算されたことになる。

表５： Ｚスコア行列

X＼Y	1人称	2人称	3人称	人間	地名	自然	固有	抽象	物	動物	集合	親族
1人称	0.000	-0.765	0.674	-0.074	4.350	-4.350	-0.366	1.488	1.419	4.350	0.674	-0.605
2人称	0.765	0.000	-4.350	-0.319	0.000	0.000	-4.350	0.736	1.426	0.000	4.350	-4.350
3人称	-0.674	4.350	0.000	0.000	4.350	0.000	0.000	1.304	1.175	4.350	0.000	-4.350
人間	0.074	0.319	0.000	0.000	4.350	4.350	-0.360	1.754	1.827	1.534	0.270	-0.842
地名	-4.350	0.000	-4.350	-4.350	0.000	-4.350	-1.803	-0.765	-4.350	-4.350	-4.350	-4.350
自然	4.350	0.000	0.000	-4.350	4.350	0.000	4.350	4.350	1.221	-4.350	0.000	0.000
固有	0.366	4.350	0.000	0.360	1.803	-4.350	0.000	1.566	2.187	1.068	0.199	0.000
抽象	-1.488	-0.736	-1.304	-1.754	0.765	-4.350	-1.566	0.000	-0.770	-1.751	-1.891	-1.826
物	-1.419	-1.426	-1.175	-1.827	4.350	-1.221	-2.187	0.770	0.000	-2.135	-4.350	-4.350
動物	-4.350	0.000	-4.350	-1.534	4.350	4.350	-1.068	1.751	2.135	0.000	-4.350	-4.350
集合	-0.674	-4.350	0.000	-0.270	4.350	0.000	-0.199	1.891	4.350	4.350	0.000	-4.350
親族	0.605	4.350	4.350	0.842	4.350	0.000	0.000	1.826	4.350	4.350	4.350	0.000
平均	-0.566	0.508	-0.875	-1.106	3.114	-0.827	-0.629	1.389	1.248	0.618	-0.425	-2.448
√2平均	**-0.801**	**0.718**	**-1.238**	**-1.565**	**4.404**	**-1.169**	**-0.890**	**1.965**	**1.764**	**0.874**	**-0.601**	**-3.462**

なお，日科技連官能検査委員会編（1973：475）に従って，当てはまりの検定をしたところ，表５のデータは 99%水準で有意差ありという結果が出た。ただし，今回のサンプル数が 4061 例と非常に多く，必然的に有意差が出ることになるため，検定に積極的な意味はないと考えられる。

５．考察

表５で最終的に得た評価値を数直線上にプロットすると，各名詞句の順序とその差の程度を視覚的に観察することができる。図２は，その結果である。

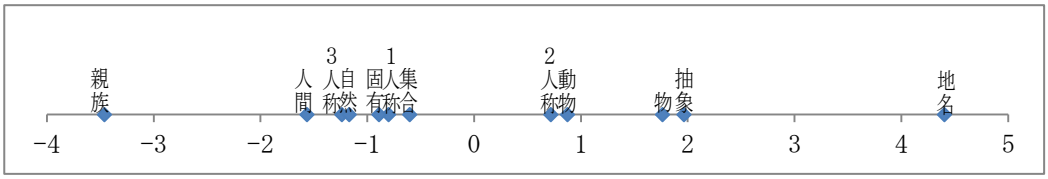

図２：日本語の名詞句階層

大まかに，左側に有生物名詞，右側に無生物名詞が位置している。Silverstein の名詞句階層と比べると，無生物名詞の順序に関しては，「自然の力の名詞＞抽象名詞＞地名」と，同じ結果になっている。もっとも，「自然の力の名詞」は有生物名詞の方に寄っているが，無生物名詞の順序自体には影響しない。そして，本稿で新たに立てた「物名詞」は，「抽象名詞」より少し階層の高いところにあることも観察された。一方で，有生物名詞に関しては少し複雑である。まず，代名詞については「３人称＞１人称＞２人称」というように，オリジナルの階層と違って「３人称」が最も上位に来ている。次に，「人間名詞」と「固有名詞」の順序も逆になっている。最も大きな相違点は「親族名詞」で，図２ではそれが数直線の左端に位置し，階層的に一番上に立っている。なお，新たに立てた「集合名詞」は「１人称」より少し下にプロットされていることも分かる。

差の程度に関しては，「親族名詞」と「地名」はそれぞれ数直線の両端にあり，ほかの名詞句と非常に離れているのが見て取れる。それに対して，「人間名詞」「３人称」「自然の力の名詞」「固有名詞」「１人称」「集合名詞」などは互いに距離が近く，階層間の差がそれほど大きくないため，その順序が明快に決められない結果となっている。

すでに述べたように，一対比較法はすべての組み合わせの相対頻度をもとに，各項目の平均的な影響力を求める方法であるため，最終評価値で得た順位は，各組み合わせの実際の頻度と矛盾することがある。例えば，「１人称」と「２人称」の組み合わせでは，生の頻度では「１人称＞２人称」が２例，「２人称＞１

人称」が７例で，「２人称」が「１人称」より上位に来るはずだが，最終評価値では逆の結果になっている。なぜこのようなことが起こるかというと，最終評価値は「１人称」と「２人称」との両者の関係だけでなく，ほかの名詞句との相対的関係も計算に入れたからである。

もし矛盾となっているところが多くあれば，その順位の信頼性も疑わしいため，詳しく検討する必要があると思われる。そこで，組み合わせに矛盾が生じた箇所に×印を付け，表６に示してみた。一番下の欄に，各名詞句に矛盾が何箇所あるかを集計した数値を記している。

表６：矛盾箇所を示した行列

X＼Y	1人称	2人称	3人称	人間	地名	自然	固有	抽象	物	動物	集合	親族
1人称		×	×									
2人称	×				×	×				×	×	
3人称	×			×		×	×				×	
人間			×				×					
地名		×										
自然		×	×							×	×	×
固有			×	×								×
抽象												
物												
動物		×				×						
集合		×	×			×						
親族						×	×					
×計	2	5	5	2	1	5	3	0	0	2	3	2

表６に示したように，「２人称」「３人称」「自然の力の名詞」には５つ，そして「固有名詞」「集合名詞」には３つ，「１人称」「人間名詞」「動物名詞」「親族名詞」には２つの矛盾箇所がある。比較的正しく測定されたのは，図２で最も右端にプロットされた，矛盾箇所が０か１の「地名」「抽象名詞」「物名詞」という無生物名詞のグループである。

数字だけを見れば，矛盾箇所は多いように感じられるが，実際に矛盾が生じた組み合わせの生の頻度を確認すると，「固有名詞」と「人間名詞」の 41 対 23

以外，他はすべて頻度が 1 桁のものばかりである。その中には，収集したデータ内で存在しない組み合わせ，つまり用例数が 0 のところも半分以上ある。用例数が少なすぎると相対頻度の数値が不安定になり，連動して最終評価値にも影響を与える[9)]。言い換えると，矛盾が起きた主な原因は，一部の組み合わせにおいて用例数が少ないことによると考えられる。もちろん，ある程度用例が集まったにも関わらず順位が論理的に整合しないパターンもあるが，それに該当するのは「固有名詞」と「人間名詞」の組み合わせの 1 箇所のみである。以下，その理由について詳しく検討する。

まず，実際に「固有名詞」と「人間名詞」の組み合わせの用例を見ると，以下のようなものがある。(1)（2）は「固有名詞＞人間名詞」，(3)（4）は「人間名詞＞固有名詞」の例である。

(1) よどんだ空気を怒声がつらぬき，雪混じりの風がふきこんだかのように，座は冷たい静寂に支配された。口角にあふれた酒をぬぐって，**翟譲が男をにらみつける**。（小前亮『李世民』）

(2) このやり取りの一部始終を見ていた**八五郎が禅僧を捕まえて**，一体全体なにが起きたのかと尋ねると，禅僧は『和尚の禅思想は…（中略）…』と答え逃げていった。（金子守『ゲーム理論と蒟蒻問答』）

(3) かなり古くからやっている店らしい。『まちかど』と書かれた看板の文字が，ところどころ読めなくなっている。「いらっしゃい」女装が似合いそうな**店主が葛原を迎えた**。（谷川涼太郎『京都・尾道れんが坂の殺人』）

(4) そのとき，鋭い視線を感じて，ふり返った。鏡ごしに，いつのまに入ってきたのか胸の薄い若い**女が綾をみつめている**。鏡の中で目があった。鏡ごしって，ほんとうにいやな視線，と綾はおもう。（辻原登『発熱』）

どれも自然な文のように見えるが，仮に連体修飾節抜きの単文の状態で見れば，(1)（2）の「翟譲が男をにらみつける」「八五郎が禅僧を捕まえる」に比べ，(3)（4）の「店主が葛原を迎えた」「女が綾をみつめている」の方が若干許容度が低いと思われる。つまり，(3)（4）で「人間名詞＞固有名詞」になったのは，連体修飾節の影響が 1 つの要因として考えられよう。もし単文という同じ

条件で両者を比較するならば，階層的には生の頻度で示されたように，「固有名詞＞人間名詞」の結果になるはずである。

では，最終的評価値と実際の頻度の結果が矛盾した原因はどこにあるかというと，どうやら「自然の力の名詞」と関係があるようだ。表３と表４を見ると，「固有名詞」と「人間名詞」とを直接比べた場合，生の頻度は 41 対 23（相対頻度 0.641）と，「固有名詞」の方が上だが，「自然の力の名詞」も入れて比較すると，「人間名詞＞自然の力の名詞」が 2 対 0（相対頻度 1），「自然の力の名詞＞固有名詞」が 1 対 0（相対頻度 1）で，順位としては「人間名詞＞自然の力の名詞＞固有名詞」というように，「人間名詞」の方が逆に上になっている。つまり，最終評価値で「人間名詞＞固有名詞」になったのは，「自然の力の名詞」による影響力が間接的に介在しているからだと考えられる。上に示したように，「人間名詞＞自然の力の名詞」も「自然の力の名詞＞固有名詞」も相対頻度が 1 で，Ｚスコアに換算すると非常に影響力の大きい数値になるが，それらの生の頻度を見ると，どちらも用例数が極めて少なく，相対頻度の値が信頼できるとは言い難い。換言すると，この組み合わせに起きた矛盾も実は用例数が少ないことに起因する可能性が高い。

以上をまとめると，今回の調査では，「自然の力の名詞」以外の無生物名詞は比較的正確に階層の測定ができたが，有生物名詞，特に人名詞に関してはまだ改善する余地があるようだ。その原因は，一部の組み合わせにおいて用例数が少ないため，相対頻度の値が不安定になることによるものだと思われる。しかし，この問題点はサンプル数を増やせば解決できるものであるため，理論的に一対比較法は名詞句階層を測定するために有効な方法だと考えられよう。

６．まとめと今後の課題

本稿では，BCCWJ を調査資料に，日本語の名詞句階層を一対比較法という統計手法で測定することを試みた。そして，その結果を Silverstein が提出したオリジナルの階層と比較することも行った。結論として，無生物名詞に関しては Silverstein とほぼ同じ結果が出たが，有生物名詞に関して，特に人名詞同士の階層は明快に決められるものではない可能性があることが示された。また，一対比較法の有効性に関して，分析の結果が安定するまではある程度のサンプ

ル数が必要だが，それが足りていれば理論的に一対比較法で名詞句階層を量的に測定することは可能であることも分かった。今回は，一部の組み合わせで用例数が少ないこともあるため，解析の結果がすべて正確とは言えないが，少なくとも無生物名詞の階層に関しては正しく測定できたと思われる。今後は用例数の問題を解決できれば，より信頼性の高い測定結果が得られよう。

今後の課題として，統計処理の結果と合わせて，実際の用例を分析することも必要な作業として挙げられる。なお，麻（2017）では，日中両言語の名詞句階層は相対的に程度性の強さが異なる可能性のあることを指摘したが，本稿の方法を用いれば，このような異なる言語の名詞句階層を客観的に比較することも期待される。

注

1）本稿で用いた『現代日本語書き言葉均衡コーパス』は，大阪大学大学院文学研究科日本語学講座が国立国語研究所と交わした利用許諾契約に基づき使用したものである。

2）実際の操作は 1000 例ずつサンプル数を増やしていくという方法で行ったが，5000 例ではじめて比較的安定した結果が得られた。なお，後にも述べるように，同じ階層の名詞句同士の組み合わせは行列では主対角線上に位置するため，実際の計算に用いたのは 4061 例である。

3）本来の一対比較法はすべての組み合わせに対する実験回数を人為的にコントロールできるが，本稿の場合はコーパスから用例を無作為に抽出する方法をとるため，次の用例がどの組み合わせになるかは予見することができない。また，組み合わせによって欠損値が発生する場合もある。

4）日本語では，人称代名詞の使用は待遇表現と深く関わっており，指し示される人物との社会的関係や場面によって，人称代名詞の代わりに「先生」「社長」のような社会的身分を表わす語を用いることがある。「３人称」に関しても同じである。

5）「中国が中日関係を改善する」であれば「集合名詞」，「総理大臣が中国を訪れた」であれば「地名」というように，同じ形の名詞であっても文脈によって分類が異なることがある。

６）本稿での「集合名詞」は，英語のように，加算か不可算を規定する「集合名詞」とは異なる意味で用いている。

７）以下，図表に示す際，「人間名詞」は「人間」，「自然の力の名詞」は「自然」，「固有名詞」は「固有」，「抽象名詞」は「抽象」，「物名詞」は「物」，「動物名詞」は「動物」，「集合名詞」は「集合」，「親族名詞」は「親族」と，略記することにする。

８）一対比較法には，相対行列の値が１か０のときに，Ｚスコアが計算できない問題点がある。それを解決するために，本稿では山本（1967）の提案に従い，それぞれ 4.35 と-4.35 で代入することにする。山本の検証結果では，この値は一般的なやり方であるギューリクセン法で計算した数値と有意差が見られないことが分かっている。

９）相対頻度は生の頻度を全体の用例数で割ったもので計算されるからである。

参考文献

奥津敬一郎（1967）「自動化・他動化および両極化転形―自・他動詞の対応―」『国語学』70:46-66

角田太作（1991）『世界の言語と日本語』くろしお出版

天坂格郎・長沢伸也（2007）『官能評価の基礎と応用―自動車における感性のエンジニアリングのために―』日本規格協会

長沢伸也編（2011）『Excel でできる統計的官能評価法―順位法，一対比較法，多変量解析からコンジョイント分析まで―』日科技連出版社

日科技連官能検査委員会編（1973）『新版　官能検査ハンドブック』日科技連出版社

麻子軒（2017）「連語論的アプローチによる無生物主語他動詞文の日中対照―対格名詞が事名詞である場合―」『阪大日本語研究』29:43-70

山本直成（1967）「一対比較法における不完備行列の計数処理」『大阪市立大学家政学部紀要』15:47-52

熊鶯（2009）『鍵がドアをあけた』笠間書院

Dixon, R. M. W. (1979) Ergativity, *Language*, 55, 59-138.

Silverstein, M. (1976) Hierarchy of features and ergativity. In Dixon,

Robert. M. W. (ed.) *Grammatical Categories in Australian Languages* (pp. 112-171). Canberra: Australian Institute of Aboriginal Studies.

Wierzbicka, A. (1979) Ethno-syntax and the philosophy of grammar. *Studies in Language*, 3(3), 313-383.

Zubin, D. (1979) Discourse function of morphology: the focus system in German. In T. Givon (ed.) *Discourse and syntax* (pp.469-504). New York: Academic Press.

使用データ

『現代日本語書き言葉均衡コーパス』DVD 版 Version 1.1

（文学研究科外国人招へい研究員）