

国立国語研究所学術情報リポジトリ

Stochastic model of sentence structure in Japanese literature

メタデータ	言語: jpn 出版者: 公開日: 2019-02-15 キーワード (Ja): キーワード (En): 作成者: 田中, 章夫, TANAKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001780

句のエントロピーにもとづく構文合成

田 中 章 夫

0. はじめに

日本語の構文を、まったく機械的な手法で組み立てていったら、どのようなものになるか、この点についての考察が、この論文の第一の目的である。

きわめて単純に考えれば、センテンスの構造というものは、文頭の第一句が決まり、これによって、つぎの第二句が選ばれる、第二句が決まったことによって、第三句にどんな句がくるかが決まる、このようにして、文末のピリオドに至って成立するものだとも言えないことはない。

いま、かりにセンテンスの第一句が「○○が」という形で始まったとすれば、第二句に、どんな形の句が来るかには、第一句に「○○が」が来たということが、強い影響を及ぼすものと仮定する。その結果、第二句に「○○ヲ」という形の句が来たなら、これは、第三句にどんな形の句がくるかに影響をもつものと仮定する。センテンスは、このように、文頭の第一句から、つぎつぎに、その直後の句に影響を及ぼしつつ、文末に至って成立するものと仮定する。

文の成立を、以上のように考えると、ある特定の構文が成立する確率は、句から句への推移確率の積で求められることになる。

また、逆に、構文を文末から観察する立場に立つ場合には、文末句がある特定のタイプのものに決定したことによって、その直前にどんな句がくるか、それが決まったことによって、その前にはどんな句が来やすいかを、文頭に至るまでたどっていくことになる。その結果、ある特定のタイプの文末句が選ばれたときに、どんな構文が成立しやすいかは、さきの場合と同様に、句から句への推移確率の積で求められる。

これは、結局、センテンスというものを、句の確率過程にもとづいて成立するものと見なすわけであり、計算言語学の方面で使われる、単語単位あるいは

音節単位の推移確率にもとづいて日本語らしい文や英語らしいセンテンスを作り出していく手法を、フレーズについて試みたことにほかならない。計算言語学用語を使えば、以上のようにして求めた構文は「フレーズ単位の近以構文」といったようなことになろう。

こうした手法で作出した構文は、いうまでもなく、きわめて生じやすい構文と見なしうるものであるから、そこに現われる構文上の諸現象を考察してみることが、構文論の基礎的事実を明らかにするうえで、重要な課題である。

さらに、句から句への続きやすさ、すなわち推移確率を把握しておくことは、語順研究や呼応現象の考察、あるいは **Phrase Structure Grammar** における **Formation Rule** の設定などに、きわめて有効な手掛りを与えることになろう。

1. 調査の方法

この研究のデータは、志賀直哉の「城の崎にて」と山本有三の「ストウ夫人」の二つから得た。これらを選んだ理由は、前者は一般向の常体の文章であり、後者は少年少女向のデス・マス体の文章なので、あるいは、こうした面での違いが、調査結果に反映するのではないかと考えたためである。

具体的な調査方法のうえでは、まず対話部分の扱いである。言うまでもなく、対話文と地の文とでは、構文の型が、当然、異なってくる。また、対話部分には、構文の乱れなど、いわゆる不整構文が出やすい。こうしたことを考慮すると、はじめから対話文を扱っていくのは危険なので、今回の調査では、対話文は、すべて調査の対象からはずし、地の文の中に対話文が引用されている個所については、つぎのように処理した。

▷彼は「ほくも行く」と主張した。

→彼は／（引用部）と／主張した／

▷「きれいな花ね」などと、彼女は言った。

→（対話部）などと／彼女は／言った／

すなわち「（対話部）+ト」あるいは「（対話部）+ナドト」の類の句を認めただけである。

以上のような処理を施した結果、今回の調査において、調査対象となったセンテンスの数は、「城の崎にて」209文、「ストウ夫人」774文である。

つぎに、これまで「句」と呼んできたものの内容であるが、これは、橋本文法の「文節」の考え方にしたいが、センテンスを文節に区切ったうえで、つぎのような処理を施したものを言う。

文頭 → 雪の → 山が → たくさん → 見えた →。

① → (名詞+ノ) → (名詞+ガ) → (副詞) → (動詞+タ) →。

すなわち、この調査では、「雪の」とか「たくさん」などの、なまの文節そのものを扱うのではなく、下段に示したように、これらを、「名詞+ノ」とか「副詞」とかに置きかえたものを扱うわけである。

なお、上記の①は、センテンスの開始を表示する「文頭」の表示であり、文末の「。」はセンテンスの終結を表わす「文末ピリオド」である。

2. 句のエントロピー

エントロピーの概念は、普通、文字・記号等の連続確率について使われるが、ここでは、ある句が決定した場合、それに続く次の句に、どの程度のあいまいさがあるかという意味で使用される。したがって、ある特定な句のあとに、かならず、ある句が連続する場合、すなわち推移確率1の場合、そのエントロピーは最小である。たとえば、連体詞句のあとには常に名詞句が続くとすれば、これは推移確率1、エントロピー0ということになる。

こうした、「句のエントロピー」にもとづいて、構文を考察するということは、簡単にいえば、センテンスというものを、句から句への推移確率の、一連の系と見なすことにほかならない。したがって、句が、実際のセンテンスにおいて、どのように現われ、さらに、それがどのように結びついて使われているかを把握することが、最初の課題である。

そこで、まず、今回とりあげた二つの作品について、句の現われ方を調べてみると、「城の崎にて」における句の延べ数は1635、異り数は168となる。また「ストウ夫人」では延べ数8405、異り数は557であった。

この中から、出現頻度の高い句を挙げてみると、表1の通りになる。「城の

崎にて」では、最高頻度の句は「動詞＋タ」の形のものであり、以下「副詞文節」「名詞＋ハ」……の順であるが、「ストウ夫人」では、第一位の「名詞＋ノ」以下「名詞＋ニ」「名詞＋ハ」……と、名詞系統のタイプの句が、集中的に上位を占めている。この事実は、あるいは、両者の文体の相違が影響しているのかもしれない。

しかし、両作品全体について、句の現われ方を、品詞系統別に整理して示すと、表2のようになり、全体の比率では、両作品の間に、ほとんど違いがない。

これらの諸点についての分析や考察は、この論の主題ではないので、データの概観という意味にとどめ、つぎに、これらの句の結びつきについてみていくことにする。

表3が、句相互間の接続型態を示すものである。すなわち「城の崎にて」に現われる異り168種のタイプの句についてみると、もっとも頻繁に出てくる結びつきは、「動詞＋テ」に「動詞＋タ」の結びついた形である。以下、「(名詞＋ノ)——(名詞＋ニ)」「(名詞＋ハ)——(副詞文節)」の順になる。「ストウ夫人」の方は、「名詞＋ノ」に「名詞＋ニ」の結びついた形が、最も多く現われ、以下「(名詞＋ヲ)——(動詞＋テ)」「(名詞＋ニ)——(動詞＋テ)」などが、これに続いている。表3を見渡してみると、当然のこ

表1. 句タイプの出現率

表1.1

表1.2

城の崎にて			ストウ夫人		
句のタイプ	度数	比率%	句のタイプ	度数	比率%
動＋タ	132	8.07	名＋ノ	898	10.68
副	128	7.83	名＋ニ	552	6.57
名＋ハ	122	7.46	名＋ハ	549	6.53
名＋ノ	120	7.34	名＋ヲ	529	6.29
動＋テ	116	7.09	動＋テ	520	6.19
名＋ニ	109	6.67	副	476	5.66
動	102	6.24	名＋ガ	410	4.88
名＋ヲ	92	5.63	動	391	4.65
名＋ガ	84	5.14	連体	305	3.63
形	53	3.24	名	247	2.94
連体	41	2.51	動＋マシタ	230	2.74
名	38	2.32	形	219	2.61
形動	33	2.02	動＋タ	202	2.40
名＋へ	29	1.77	形動	162	1.93
接	27	1.65	名＋ト	151	1.80
名＋モ	25	1.53	名＋モ	128	1.52
			名＋デ	121	1.44
			動＋マセン	81	0.96
			動＋マス	79	0.94
			動＋ト	73	0.87
			名＋ニハ	71	0.84
			名＋カラ	71	0.84
			名＋デス	50	0.59
			動＋デ	45	0.54
			動＋レテ	43	0.51
			名＋デハ	42	0.50
			名＋デシタ	32	0.38
			名＋デモ	31	0.37
			動＋ヤウナ	31	0.37

表 2. 句タイプの品詞系統別分布

句のタイプ	城の崎にて		ストウ夫人	
	度数	比率(%)	度数	比率(%)
名詞系	767	46.91	4308	56.26
動詞系	537	32.84	2567	30.54
形容詞系	73	4.46	312	3.71
形容動詞系	37	2.26	170	2.02
副詞系	153	9.36	552	6.57
連体詞	41	2.51	320	3.81
接続詞	27	1.65	134	1.59
感動詞	0	0	29	0.35
引用句	0	0	13	0.15
計	1635		8405	

とではあるが、どちらの作品においても、きわめて常識的な結びつきが上位を占め、それらが、きわめて頻繁に現われていることが、よくわかる。単純に言えば、「城の崎にて」では168種のタイプの句があるのだから、その二乗通りの結びつきが作りうるわけであり、「ストウ夫人」では557の二乗通りの接続形が存在しうるわけであるが、実際には、きわめて限られた結びつきが、くり返

し出て来ているにすぎない。たとえば、「ストウ夫人」の「名詞+ノ」は、表1に示す通り、もっとも出現率の高い句であるが、頻度数898のうち、過半数

表 3. 句のタイプ間の推移確率

表 3.1 「城の崎にて」

順位	先行	後続	度数	推移確率	
				→	←
1	(動+テ)	— (動+ク)	38	32.76%	28.79%
2	(名+ノ)	— (名+ク)	26	21.67	23.85
3	(名+ハ)	— (副)	25	20.49	19.53
4	(動+テ)	— (動)	22	18.97	21.57
5	(名+ニ)	— (動+テ)	20	18.35	17.24
6	(名+ヲ)	— (動+ク)	16	17.39	12.12
7	(名+ヲ)	— (動+テ)	15	16.30	12.93
8	(動)	— (名+ガ)	13	12.75	15.48
8	(名+ノ)	— (名+ノ)	13	10.83	10.83
8	(名+ガ)	— (動+テ)	13	7.07	11.21
11	(名+ガ)	— (副)	12	6.57	9.38
11	(名+ノ)	— (名+ガ)	12	10.00	14.29
11	(副)	— (名+ニ)	12	9.38	11.01
11	(副)	— (動+テ)	12	9.38	10.34
15	(動)	— (名+ヲ)	11	10.78	11.96
15	(名+ノ)	— (名+ヲ)	11	9.17	11.96
15	(名+ニ)	— (動+ク)	11	10.09	8.33
15	(副)	— (名+ノ)	11	8.59	9.17
19	(名+ヲ)	— (動)	10	10.87	9.80
19	(連体)	— (名+ヲ)	10	24.39	10.87

表 3.2 「ストウ夫人」

順位	先行	後続	度数	推移確率	
				→	←
1	(名+ノ)	— (名+ニ)	143	15.92%	25.91%
2	(名+ヲ)	— (動+テ)	134	25.33	25.77
3	(名+ニ)	— (動+テ)	113	20.47	21.73
4	(名+ノ)	— (名+ヲ)	110	12.25	20.79
5	(名+ハ)	— (副)	104	18.94	21.85
6	(名+ノ)	— (名+ガ)	93	10.36	22.68
7	(名+ハ)	— (名+ノ)	89	16.21	9.91
8	(名+ヲ)	— (動)	87	16.45	22.25
9	(名+ノ)	— (名+ハ)	82	9.13	14.94
10	(名+ノ)	— (名+ノ)	81	9.02	9.02
10	(副)	— (名+ノ)	81	17.02	9.02
11	(連体)	— (名+ノ)	71	23.28	7.91
12	(動+テ)	— (動+マシク)	70	13.46	30.43
13	(動+テ)	— (動)	67	12.88	17.14
14	(動)	— (名+ガ)	59	15.09	14.39
15	(動)	— (名+ノ)	56	14.32	6.24
16	(動)	— (名+ハ)	53	13.55	9.65
17	(名+ニ)	— (動)	50	9.06	12.79

注) →は先行句からの確率を示す。
←は後続句からの確率を示す。

は、その直後に「名詞+ニ (143回)」「名詞+ヲ (110)」「名詞+ガ (93)」「名詞+ハ (82)」の、わずか四種のうちのいずれかをしたがえた形で使われている。また、「城の崎にて」の「動詞+タ」についてみると、出現度数132のうち、その先行句は、「動詞+タ (38回)」「名詞+ヲ (16)」「名詞+ニ (11)」「副詞文節 (6)」の四種で、すでに過半数に達してしまう。このようにして調べていくと、理論的には、さまざま接続形が想定しうるにしても、実際のセンテンスに現われる句の接続形態は、その種類が、きわめて限られているといえてよい。

つぎに、表3の推移確率を眺めてみると、たとえば、「城の崎にて」の「動詞+テ」と「動詞+タ」の結びつきのように、先行句からの推移確率と、後続句からの推移確率が、ともに高いもの、「(副詞文節)——(名詞+ノ)」のようにともに低いもの、あるいは、「ストウ夫人」の「(連体詞文節)——(名詞+ノ)」の結びつきのように、先行句からの推移確率が高く、後続句からのそれが低いタイプ、逆に「(動詞+テ)——(動詞+マシタ)」の場合のように先行句からの推移確率が低く、後続句からのそれが高いタイプの、以上四種のタイプを見出すことができる。言うまでもなく、両方の推移確率が、ともに高いということは、両者の結びつきが強いことを示し、ともに低いということは、結びつきが弱いことを物語る。また、先行句からの、あるいは後続句からの推移確率のみが高いということは、一方からの結びつきのみが強く、他方からの結びつきは必ずしも強くないことを表わす。

ここで重要なことは、この結びつきの強さ・弱さの様相が、表3でわかる通り、頻度とは、必ずしも一致しないという事実である。このことは、語順研究・呼応現象の考察をはじめ、構文論の諸研究において、語句の結びつきや関連性を究明する場合、度数のみにたよることの危険性を示唆している。

以上、二つの文章を資料として、実際のセンテンスに現われる、句の結合パターンおよび、その結合の様相について調べてきたが、このような事実を把握し、それを積み重ねていくことは、構文論を、より確実なものにしていくために、きわめて重要なことである。その方法の一つとして、推移確率による接近は、有効な手段ではないかと考えるのである。

3. 文頭の生起率

センテンスが、どのようなタイプの句で始まりやすいか、今回とりあげた二つの作品について、文頭句のタイプを示すと、表4のようになる。いずれの作品においても「名詞

＋ハ」の形が第1位を占め、「城の崎にて」では、総文数209文のうちの57文が、このタイプの句によって書き起されている。また「ストウ夫人」では全体の17.05%の文が「名詞＋ハ」によって起されている。この比率を、センテンスの「生起率」

表4 頻度の高い文頭句

表4.1「城の崎にて」

順位	文頭句	度数	生起率%
1	名詞＋ハ	57	27.27
2	名詞＋ノ	26	12.44
3	副詞	21	10.04
3	接続詞	21	10.04
5	連体詞	14	6.69
6	名詞＋ガ	10	4.78
7	名詞	6	2.87
8	形容詞	5	2.39
8	名詞＋デ	5	2.39
8	名詞＋ニ	5	2.39
8	名詞＋ヲ	5	2.39
12	形容動詞	4	1.91

表4.2「ストウ夫人」

順位	文頭句	度数	生起率%
1	名詞＋ハ	132	17.05
2	接続詞	123	15.89
3	名詞＋ノ	101	13.04
4	連体詞	78	10.07
5	副詞	73	9.43
6	名詞＋ガ	41	5.29
7	名詞	36	4.65
8	名詞＋ヲ	25	3.22
9	感動詞	20	2.58
10	名詞＋ニ	19	2.45
11	名詞＋ト	11	1.42
12	名詞＋モ	10	1.29

文頭総数(=総文数)：209 文頭総数(=総文数)：774と名づけると、生起率の上位のところには、文頭句としては、きわめて、ありふれたものが並んでくる。そのうえ、この二つの作品間に、それほど違いはない、というよりは、すくなくとも上位の七位ぐらいまでは、多少の入れ替わりはあっても、まったく一致している。

さらに、長尾真氏が、国立国語研究所の新聞語彙調査データについて調査した結果、新聞記事文の文頭文節に現われやすい助詞とした「ハ」「ノ」「ニ」「ガ」「ヲ」の類は、いずれも、今回調べた二つの作品でも、上位10位までの文頭句に姿をみせている。^(注1)

こうしたことを考え合わせると、あるいは、文章は異なっても、文頭句のタイプというものは、そう違わないのではないと思われる。

つぎに、文頭に来る句のタイプが決まった場合、その句の直後に、どんなタイプの句が位置しやすいか、これを、最も頻度の高い文頭句「名詞＋ハ」につ

いて示してみると表5のようになる。すなわち、「城の崎にて」では、「名詞＋ハ」122のうち、25

は、あとに「副詞文節」をしたがえている。「ストウ夫人」では、549の「名詞＋ハ」のうちの18.94%は、「副詞文節」に続いている。この比率は、言うまでもなく、「名詞＋ハ」から「副詞文節」への推移確率にほかな

表5 「名詞＋ハ」の後続句

表5.1 「城の崎にて」

順位	後続句	度数	推移確率%
1	副詞	25	20.49
2	名詞＋ノ	12	9.84
3	形容詞	9	7.38
4	名詞＋ニ	7	5.74
5	形容動詞	5	4.10
6	名詞＋デ	5	4.10
6	名詞＋ヲ	5	4.10
8	連体詞	4	3.28
8	名詞	4	3.28
10	名詞＋ヘ	3	2.46

「名詞＋ハ」の総数：122

表5.2 「ストウ夫人」

順位	後続句	度数	推移確率%
1	副詞	104	18.94
2	名詞＋ノ	89	16.21
3	名詞＋ニ	41	7.46
4	連体詞	33	6.01
5	名詞＋ヲ	30	5.46
6	形容詞	24	4.37
6	名詞	24	4.37
8	形容動詞	16	2.91
9	名詞＋ト	12	2.18
10	名詞＋モ	11	2.00
10	名詞＋ガ	11	2.00
12	動詞＋マセン	10	1.82

「名詞＋ハ」の句の総数：549

らない。それでは、「副詞文節」を得た場合、これに最も続きやすい句は、どんなタイプかということ、「城の崎にて」では、「名詞＋ニ」、「ストウ夫人」では「名詞＋ノ」のタイプということになる。このようにして、文頭句から、推移確率のもっとも高いタイプの句をたどっていくと、「城の崎にて」では、つぎのような構文が得られる(数値は推移確率)。

① $\xrightarrow{27.27}$ (名＋ハ) $\xrightarrow{20.49}$ (副) $\xrightarrow{9.38}$ (名＋ニ) $\xrightarrow{18.35}$ (動＋テ) $\xrightarrow{32.76}$ (動＋タ) $\xrightarrow{81.82}$ 。

また、「ストウ夫人」では、つぎのようになる。

① $\xrightarrow{17.05}$ (名＋ハ) $\xrightarrow{18.94}$ (副) $\xrightarrow{17.02}$ (名＋ノ) $\xrightarrow{15.92}$ (名＋ニ) $\xrightarrow{20.47}$ (動＋テ) $\xrightarrow{13.46}$ (動＋マシタ) $\xrightarrow{99.57}$ 。

すなわち、「城の崎にて」の場合は、「動詞＋タ」に至って、このタイプの句のあとにもっとも現われやすいものは、文末ピリオドということになり、この作業が終了するわけである。同様に、「ストウ夫人」の場合も、「動詞＋マシタ」の句を得ると、そのつぎには、99.57%の推移確率で文末ピリオドに到達することになる。

このようにして、今、作成した構文が、文頭から二つずつの句の推移確率を採っていった場合の、それぞれの作品における典型的な構文ということになる。

これら、二つの構文を比べてみると、文末が、常体か敬体かの相違だけで、ほぼ同一の構文と見なしうる。「名詞+ノ」のあるなしは、「(名詞+ノ)→(名詞)」までで一つの名詞句相当と見れば、構文の基本的な相違とはならない。

これが、日本語のセンテンス・パターンとして代表的なものかどうかは、もちろん、まだ断定できる段階ではないが、すくなくとも、二つずつの句の推移確率に基づいて、構文を作っていた場合には、きわめて成立しやすい構文の一つのタイプだとは言えるように思われる。

以上は、「名詞+ハ」のタイプの文頭句について、推移確率のもっとも高いところだけをたどってみたが、表5に示したように、「名詞+ハ」だけについても、推移確率の大小の違いはあるが、さまざまな句が続きうる。したがって、「名詞+ハ」だけを初期状態(出発点)としても、たいへん大きな樹型測度が描かれてくる。表6に、その一部を示すことにしよう。

「名詞+ハ」をめぐる構文として、すぐ思い浮べられるのは、「象は鼻が長い」式の構文であるが、表6に示した範囲には現われてこない。あるいは、テ

表6. 「名詞+ハ」で始まる樹型の一部

表6. 1 「城の崎にて」

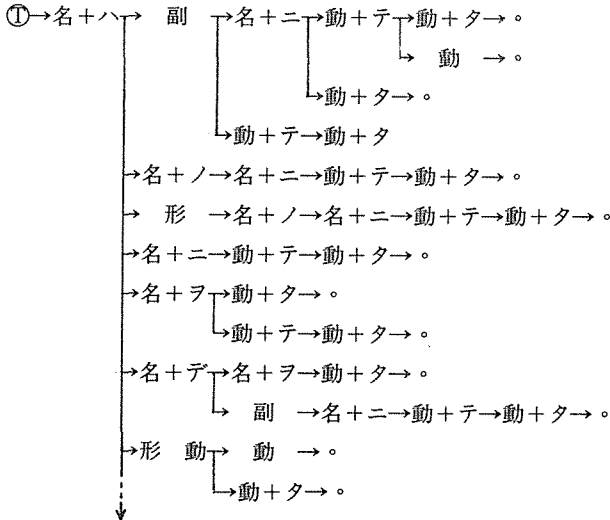
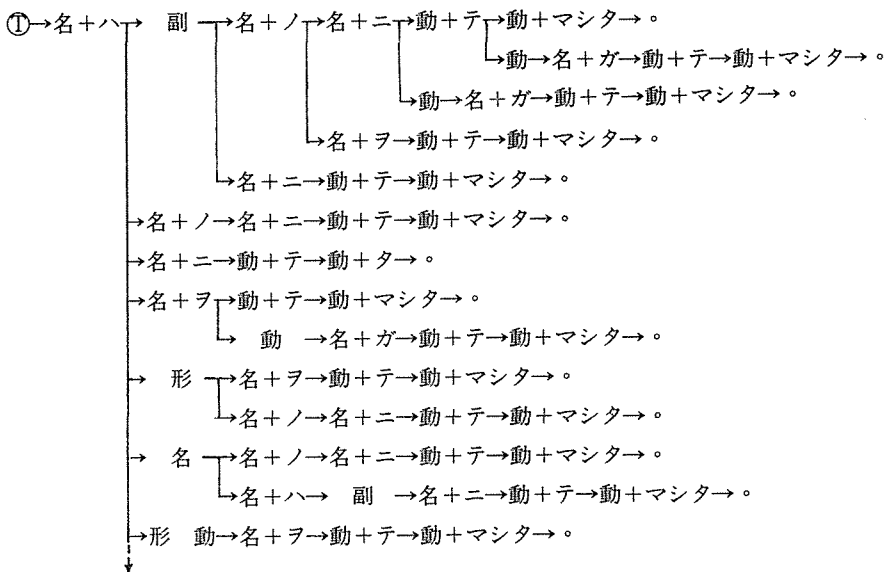


表 6. 2 「ストウ夫人」



一タの性格によるのかもしれないが、この形の構文は、今回のようなやり方で確率的にたどっていった場合には、あまり成立しやすいものではないように思われる。これについては、また、第6節でとりあげることにする。

なお、さきに「生起率」としたものは、言うまでもなく、文頭表示①に、それぞれの文頭句が続く推移確率にほかならない。

(注1) 長尾真「構文の意味の解析の試み」計量国語学・64集

4. 文末の終束率

前節で調べた文頭句の場合と同じように、それぞれの作品のセンテンスの文末句を調べてみると、表7のようになる。この表を見てみると、どちらの作品においても、最も頻度の高い文末句は、過去の「タ」を伴う動詞句である。「城の崎にて」では総文数209文のうち108文が「動詞+タ」のタイプの句によって閉じられている。「ストウ夫人」でも総文数の約3割が「動詞+マシタ」の句で結ばれている。この比率を、センテンスの「終束率」と呼べば、終束率の高

表7. 頻度の高い文末句

表7.1 「城の崎にて」

順位	文末句	度数	終束率%
1	動詞 + タ	108	51.67
2	動詞	29	13.87
3	形容詞 + タ	7	3.34
3	動詞 + ナカッタ	7	3.34
5	動詞 + ナイ	6	2.87
6	名詞 + ダ	5	2.39
6	名詞 + タッタ	5	2.39
8	動詞 + レタ	4	1.91
8	形容詞	4	1.91
10	形容動詞 + タ	3	1.43

表7.2 「ストウ夫人」

順位	文末句	度数	終束率%
1	動詞 + マシタ	229	29.58
2	動詞 + マセン	80	10.33
3	動詞 + マス	79	10.20
4	名詞 + デス	50	6.45
5	名詞 + デシタ	32	4.13
6	動詞 + マセンデシタ	25	3.22
7	動詞 + タノデス	24	3.10
8	名詞	18	2.32
9	名詞 + ダ	8	1.03
9	名詞 + デセウ	8	1.03
9	動詞	8	1.03
12	動詞 + ノデス	7	0.90

いところで目立つのは、過去形の文末句と否定形の文末句である。

表7に現われたものだけでも、過去形の文末句の総終束率は、「城の崎にて」では64.11%にのぼり、「ストウ夫人」でも40.05%に達する。また否定形の文末句のそれは、「城の崎にて」6.22%、「ストウ夫人」13.57%となる。

文末句のタイプとしては、理論的には、さまざまなものが、想定しうが、実際の文章では、かなり大きな比重で、過去形の句ついで否定形の句にかたよるようである。

つぎに、表7について、この二つの作品のセンテンスの文末句を比べてみると、当然のことながら、敬体の文章と常体の文章との相違が、きわめてはっきりと出てきている。特に「城の崎にて」では、ハダカの動詞文節が、第2位を占めているのに対して、「ストウ夫人」では、このタイプの文末句は、わずかに1%台に落ちこんでしまっている点などは、その顕著な例といえよう。これは、言うまでもなく、「ストウ夫人」においては「動詞+マス」がカバーしてしまった結果である。このようにして、終束率上位のところで、両作品の文末句を対応させてみると、「城の崎にて」の「名詞+ダ」と「ストウ夫人」の「名詞+デス」あるいは「動詞+ナイ/動詞+マセン」「動詞+ナカッタ/動詞+マ

センデシタ」「名詞+グッタ/名詞+デシタ」などのセットが作りうる。こうしたところに、両者の文末句の特徴が現われているとみることができる。

それでは、前節で文頭句の生起率を起点として試みたのと同じ方法で、ここでは、文末句を起点として、構文作成を試みることにする。まず、「城の崎にて」の場合、文末ピリオドの直前に来やすい句、すなわち終束率最高の句のタイプは、表7に示す通り「動詞+タ」である。「動詞+タ」を得たとき、その直前に位置しやすい句は、表8の通りである。この表から、「動詞+タ」があ

表 8. 終束率最高の句の先行句

表 8.1 「城の崎にて」の「動詞+タ」 表 8.2 「ストウ夫人」の「動詞+マシタ」

順位	先行句	度数	推移確率	順位	先行句	度数	推移確率
1	動詞+テ	38	28.79%	1	動 詞 + テ	70	13.04%
2	名詞+ヲ	16	12.12	2	名 詞 + ヲ	37	16.09
3	名詞+ニ	11	8.33	3	名 詞 + ニ	23	10.00
4	副 詞	6	4.55	4	名 詞 + ガ	17	7.39
5	名詞+ガ	5	3.79	5	副 詞	10	4.35
5	名詞+デ	5	3.79	6	動詞+ヤウニ	7	3.04
5	名詞+ノ	5	3.79	6	動詞+レテ	7	3.04
8	動詞+デ	4	3.03	8	名 詞 + デ	6	2.61
8	形容動詞	4	3.03				

「動詞+タ」の総数：132

「動詞+マシタ」の総数：230

ったとき、その直前には「名詞+テ」が最も来やすいことがわかる。こうして、文頭表示①を得るまで、この作業をくり返していくと、つぎのような構文が作り出されてくる（数値は推移確率）。

。 $\xrightarrow{51.67}$ (動+タ) $\xrightarrow{27.54}$ (動+テ) $\xrightarrow{17.24}$ (名+ニ) $\xrightarrow{23.85}$ (名+ノ) $\xrightarrow{21.31}$ ①

同様に「ストウ夫人」で試みると、つぎようになる。

。 $\xrightarrow{29.58}$ (動+マシタ) $\xrightarrow{30.43}$ (動+テ) $\xrightarrow{25.77}$ (名+ヲ) $\xrightarrow{20.79}$ (名+ノ) $\xrightarrow{11.25}$ ①

これらが、文末ピリオドから、二つずつ、推移確率が最高になる句をたどっていった場合に成立する、それぞれの作品の典型的な構文である。この二つを比べてみると、目的格が「ニ格」か「ヲ格」かの違いだけで、その他は、完全に一致している。どちらの構文も、文頭に「名詞+ノ」が来て、そのあとに目的格が続き、「動詞+テ」を経て、過去形の動詞文節で閉じられる点で、構造

上、全く同一の構文と見なしうる。

この結果を、今回対象とした作品特有のものともみるか、もっと一般性のあるものと見なしうるかには論議の余地もあろうが、すくなくとも、偶然の一致とってしまうわけにはいかない。まったく機械的な操作によって、別々の作品から、構造上、きわめて類似した構文が得られた点は注目すべきである。

前節表6において「名詞+ハ」の文頭句について試みたのと同じ方法で、最高終束率の文末句について、樹型の一部を描いてみると、表9のようにになる。この表で興味深い点は、「ストウ夫人」の表9-2に、「ニ格→ヲ格→動詞」の語順は、現われているにもかかわらず、「ヲ格→ニ格→動詞」のオーダーが、すくなくとも表8の範囲に見えてこない点である。今回のような方法で、推移確率をたどっていった場合、「ヲ格→ニ格」の順序は、「ニ格→ヲ格」ほどには出てこないように見える。この点については、また第6節において、とりあげてみることにしよう。

なお、さきに「終束率」としたものは、言うまでもなく、文末ピリオドから文末句への推移確率に当たるものにほかならない。

表9. 最高終束率の句で始まる樹型の一部

表9.1 「城の崎にて」

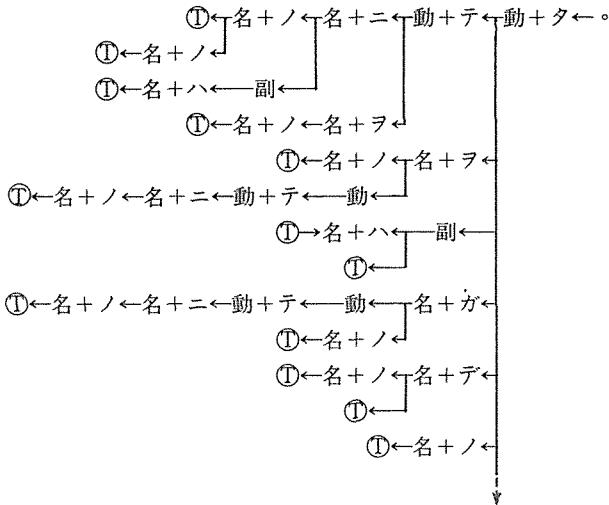
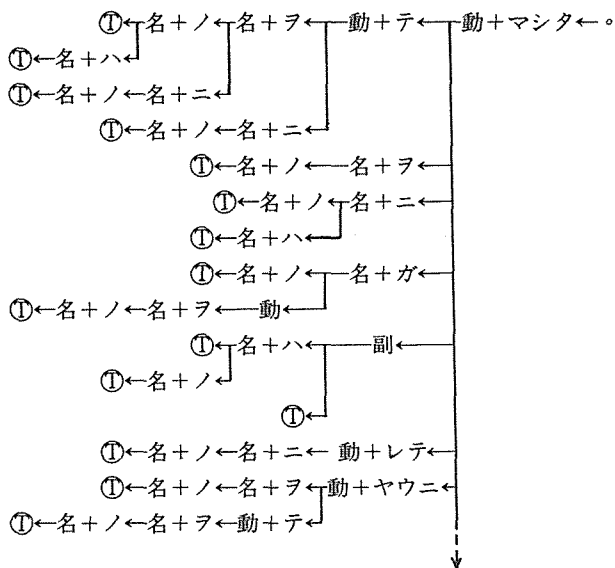


表9.2 「ストウ夫人」



5. 構文の確率モデル

第3節・第4節において作成を試みた構文は、いずれも、推移確率の最も高い句の連続だけをたどっていったものであるが、一つ一つのタイプの句には、推移確率の大小の違いはあっても、隣接しうる多くの句が存在する。これを全部書きこんで、樹型測度を作っていたら、文頭表示①あるいは文末ピリオドを初期状態とした膨大な樹型が描かれてくる。とても全部は示せないで、今回とりあげた二つの作品について、文頭からの樹型の一部を示してみると、表10のようになる。「城の崎にて」から得た樹型モデル(表10-1)と、「ストウ夫人」のモデル(表10-2)とを比べてみると、文頭句の同一のものについては、ここでも、きわめて似た構文になっている。したがって、文頭句が決まった場合、それが導いてくる構文タイプには、文章による違いが、あまりないのではないかという推測が成り立ってくる。

つぎに、文末ピリオドからの樹型の一部を示してみると、表11の通りになる。当然のことではあるが、文末を起点とした場合には、敬体と常体との違いが強

く影響して、同一の文末句を得ることは、ほとんどない。そのため、構文のタイプも、両作品の間に、かなりの違いが見られる。この点に、この二つの作品の文体の相違が反映しているということもできよう。いずれにしても、文末からのモデルにおいては、敬体と常体の違いがあるため、共通性がとらえられないので、これらのモデルが、どの程度一般的なものと見なしうるかは論じられない。しかし、きわめて常識的な構文タイプが得られている点を考えると、すくなくとも、これらの構文タイプは、文末句が決定した場合には、かなり成立しやすいものではないかと思われる。

表10. 文頭からの樹型の一部

表10.1 「城の崎にて」

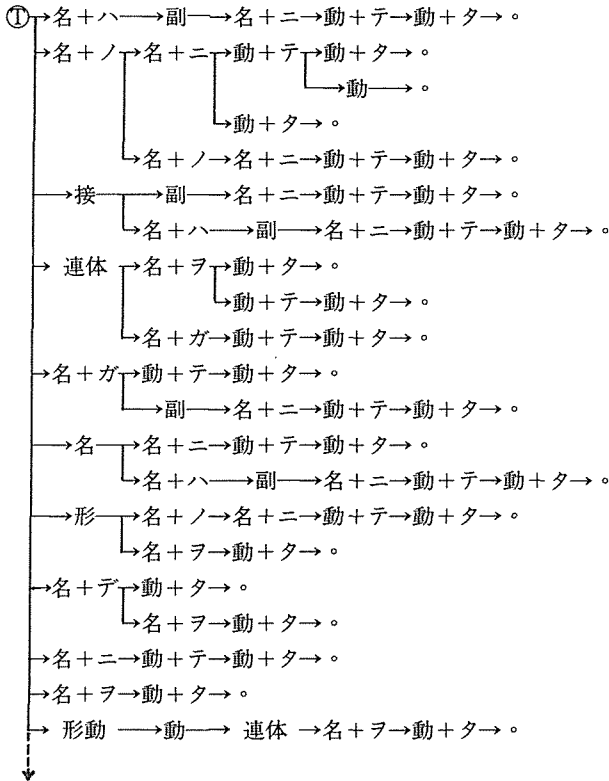


表10.2 「ストウ夫人」

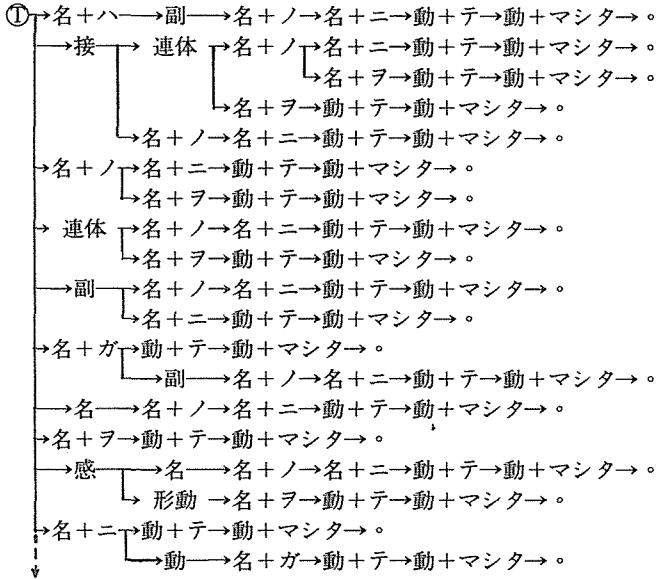
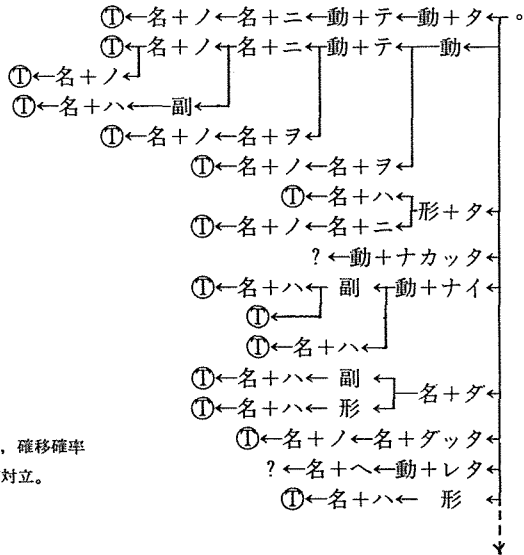


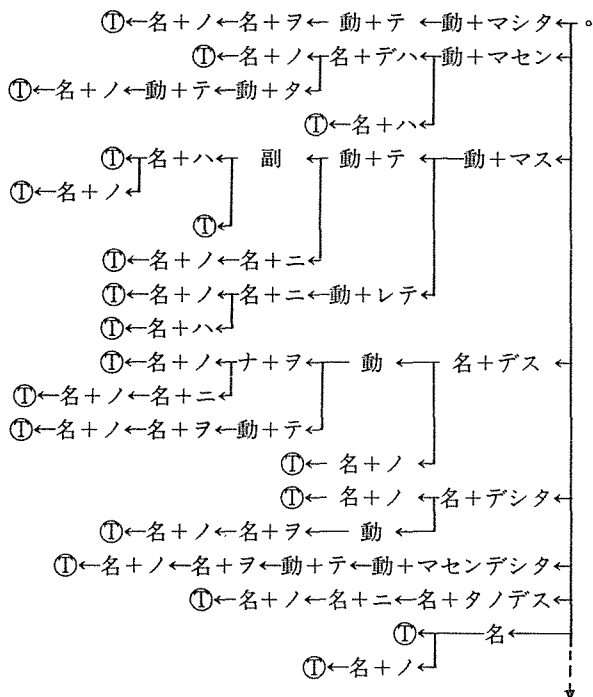
表11. 文末からの樹型の一部

表11.1 「城の崎にて」



?の個所は、確移確率
 同率の句が対立。

表11.2 「ストウ夫人」



はじめにも述べたように、今回、作成を試みた構文タイプは、「フレーズ単位の近似構文」とでも呼ぶべきものであり、計算言語学の方面の概念を借用すれば「二次近似の近似構文」ということになる。したがって、もし、これらの構文の品詞名の個所に、原文によく出てくる実際の単語を挿入すれば、一応「城の崎にて」らしい文、あるいは、「ストウ夫人」に出てきそうなセンテンスになるはずである。いま「城の崎にて」の語彙調査の結果を使^(注²)って、表10-1、および、表11-1に掲げた構文に、具体的な単語をあてはめてみると、つぎのような文が得られる（数字は、語彙調査における品詞別の頻度順位）。

{ ① → (名+ハ) → (副) → (名+ニ) → (動+テ) → (動+タ) → 。
 自分は さう いもりに して るた。
 <1位> <1> <3> <1> <2>

{ ① → (接) → (副) → (名+ニ) → (動+テ) → (動+タ) → 。
 { しかし もう 石に な って し ま っ た 。
 <1> <2> <7> <3> <12>

{ ① ← (名+ノ) ← (名+ニ) ← (動+テ) ← (動) ← 。
 { ね ず み の 前 に 行 っ て み る 。
 <5> <7> <8> <7>

{ ① ← (名+ハ) ← (副) ← (動+ナイ) ← 。
 { 蜂 は ま っ た く 動 か な い 。
 <4> <4> <13>

これらは、センテンス・ジェネレートの一つの手法として、確率的に、一旦構文を合成した上で、意味要素をあてはめていくやり方を示したものである。従来、しばしば試みられているような、直接、単語や音節の連続確率をとって文を作っていく方法よりは、自然なセンテンスを得やすいのではないかと思う。しかし、センテンスのバラエティーを求めるのならば、今回試みたような、二文節連続の推移確率では、やはり限度がある。コンピュータでも使って、三文節連続のエントロピーを計算し、三次近似の構文を作って進めれば、もっと自然な、そして多様なセンテンスが作りうるはずである。

また、こうしたところにコンピュータを使って、句のエントロピーを、さまざまな形で把握しておくことは、構文論の確実な資料を得る意味で大切なことと思われる。

(注2) 「城の崎にて」の語彙調査結果は、田中章夫「自動抄録処理におけるキーワードの性格」(国立国語研究所報告49「電子計算機による国語研究〈V〉」所収)の173ページ参照。

6. 句の結合型態の確率

この節では、これまでに取扱ってきた、二つの作品における句のエントロピーを使って、構文論で、しばしば、とりあげられる「句の結合形」のいくつかについて考察してみることにする。

はじめに、第3節で触れた「象は鼻が長い」型の結合であるが、これと、「象の鼻は長い」型の結合とでは、どちらの方が成立しやすいか見てみると、結果

は、表12のようになる。この表で、「形容詞系」としたのは、形容詞で始まる型の句を一括したものである。表12をみると、今回のデータで推定した限りでは、「象の鼻は長い」のタイプの構文の方が、「象は鼻が長い」よりは、圧倒的に成立しやすいということになる。

同様な方法で、第4節でとりあげた「ニ格一ヲ格一動詞」の順序と、「ヲ格一ニ格一動詞」の順では、どちらの方が成立しやすいかを調べてみると表13の通りであり、この表で「動詞系」としたのは、動詞で始まるタイプの句を一括したものである。表13からわかるように、いずれの場合も、推移確率は、常に「ニ格一ヲ格一動詞」の方が、「ヲ格一ニ格一動詞」よりも高い。したがって、今回得たデータの範囲では「ニ格一ヲ格一動詞」の系の方が、成立しやすいということになる。

表14は、文法で、しばしばとりあげられる「水が飲みたい」型の構文と、「水ヲ飲みたい」型の構文との対応についてであ

表12 「象は鼻が長い」型と「象の鼻は長い」型

データ	(先行句) → (後続句)	確率	データ	(先行句) ← (後続句)	確率
城の橋にて	(名+ハ) 1.64% (名+ガ) 3.57% (形容詞系)	0.59%	城の橋にて	(名+ハ) 2.38% (名+ガ) 4.11% (形容詞系)	0.98%
ストウ夫	(名+ノ) 7.50% (名+ハ) 7.38% (形容詞系)	5.53	ストウ夫	(名+ノ) 7.83% (名+ハ) 12.33% (形容詞系)	9.10
	(名+ハ) 2.00% (名+ガ) 3.65% (形容詞系)	0.73		(名+ハ) 2.68% (名+ガ) 6.85% (形容詞系)	1.84
	(名+ノ) 9.13% (名+ハ) 4.37% (形容詞系)	3.99		(名+ノ) 14.94% (名+ハ) 10.96% (形容詞系)	16.37

表13 「ニ格・ヲ格」と動詞の間の推移確率

データ	(先行句) → (後続句)	確率	データ	(先行句) ← (後続句)	確率
城の橋にて	(名+ニ) 4.59% (名+ヲ) 76.09% (動詞系)	34.90%	城の橋にて	(名+ニ) 5.43% (名+ヲ) 13.04% (動詞系)	7.01%
ストウ夫	(名+ヲ) 2.17% (名+ニ) 48.62% (動詞系)	10.57	ストウ夫	(名+ヲ) 1.83% (名+ニ) 9.87% (動詞系)	1.81
	(名+ニ) 8.15% (名+ヲ) 82.99% (動詞系)	67.65		(名+ニ) 2.51% (名+ヲ) 17.10% (動詞系)	14.55
	(名+ヲ) 4.54% (名+ニ) 68.84% (動詞系)	31.23		(名+ヲ) 4.35% (名+ニ) 14.80% (動詞系)	6.44

表14. 「水ガ飲みたい」型と
「水ヲ飲みたい」型(ストウ夫人)

(先行句) —— (後続句)	確率
(名+ガ) → (動+タイ系)	0.24%
(名+ヲ) → (動+タイ系)	0.57
(名+ガ) ← (動+タイ系)	8.33
(名+ヲ) ← (動+タイ系)	25.00

る。この表で「動+タイ系」としたのは、「飲みたい・飲みたいく・飲みたいか……」など、動詞に助動詞「たい」のついた型の句を一括したものである。表14の結果をみると、すくなくとも、「ストウ夫人」のデータからは、「水ヲ飲みたい」的なパターンの方が、「水ガ飲みたい」的なパターンよりも成立しやすいとみることができ（「城の崎にて」の方は、「(名詞+ヲ) — (動詞+タカッタ)」の接続形が一例あっただけで、この点についての結果を出すには至らなかった）。

つぎに、「城の崎にて」のデータにもとづいて「天気がいい日」型の構文と、「天気ノいい日」型の構文の推移確率を計算すると、表15の通りになる。また「雨が降る日」のタイプと「雨ノ降る日」のタイプについて、「ストウ夫人」のデータから推移確率を求めると表16のようになる。この二つの表を見ると、

表15. 「天気がいい日」型と
「天気ノいい日」型(城の崎にて)

(先行句) —— (後続句)	推移確率
(名+ガ) $\xrightarrow{3.57\%}$ (形) $\xleftarrow{52.83\%}$ (名詞系)	18.87%
(名+ノ) $\xrightarrow{1.67}$ (形) $\xleftarrow{52.83}$ (名詞系)	8.81
(名+ガ) $\xrightarrow{2.66}$ (形) $\xleftarrow{3.65}$ (名詞系)	2.01
(名+ノ) $\xrightarrow{3.77}$ (形) $\xleftarrow{3.65}$ (名詞系)	1.38

表16. 「雨が降る日」型と
「雨ノ降る日」型(ストウ夫人)

(先行句) —— (後続句)	推移確率
(名+ガ) $\xrightarrow{1.71\%}$ (動) $\xleftarrow{90.79\%}$ (名詞系)	15.50%
(名+ノ) $\xrightarrow{2.00}$ (動) $\xleftarrow{90.79}$ (名詞系)	18.20
(名+ガ) $\xrightarrow{1.79}$ (動) $\xleftarrow{8.24}$ (名詞系)	1.48
(名+ノ) $\xrightarrow{4.60}$ (動) $\xleftarrow{8.24}$ (名詞系)	3.79

「天気がいい日」型と「天気ノいい日」型では、「ガ格」の方が成立しやすく、「雨が降る日」と「雨ノ降る日」では、わずかに「ノ格」の方が成立しやすいという結果になっている。すなわち「(名詞+ガ・ノ) — (形容詞句・動詞句) — (名詞系)」の系において、「形容詞句」が選ばれたときは「ガ格」に、「動詞句」が選ばれたときは「ノ格」になりやすいというわけである。なお、表15・表16の「名詞系」は、名詞ではじまる句(文節)を一括して扱ったも

のである。

以上の結果は、今回とりあげた二つの短編小説から得たデータをもとにしたものなので、これをもって、日本語の構文の一般的な傾向を推しはかることは、もちろん、できないが、このように推移確率を用いて、構文上の問題を考察することの利点を理解してもらえれば充分である。単に、度数カウントのみによって構文上の数理的な傾向を分析することは、センテンスが、一面において、確率的な過程を経て成立してくることを考えると、場合によっては、かなり危険であると言わざるをえない。今までに挙げた諸例からもわかる通り、句と句との結びつく可能性というものは、文頭から文末方向へ、先行句を初期状態とした場合と、文末から文頭方向へ、後続句を初期状態とした場合とでは、かなり違ってくる。ということは、構文上のある事象が、句と句との連接や関連に基づいて生ずるとき、先行句からみた重みと、後続句からみた重みとは、必ずしも一致しない。むしろ、ある場合には、先行句から求めた傾向と、後続句から求めた傾向とが逆になることもありうることを示している。

簡単な話が、英語においては、QのあとにはUが来やすい、これは事実であるが、Uの前に来やすいのは、決してQではなくて、Oである。構文上の問題においても、たとえば、連体詞文節のあとには百パーセント体言文節が続くが、体言文節の前にもっとも現われやすいのは、決して連体詞句ではなく、表3でわかる通り「名詞+ノ」のタイプの句である。

今回の調査においても、さきに「城の崎にて」のデータに基づいて表15に示した「天気がいい日/天気ノいい日」型構文の確率過程を、「ストウ夫人」について調べてみると、表17のようになる。すなわち、この場合には、先行句からたどると「カ格」の方が成立しやすく、後続句からたどると「ノ格」の方が

表17. 推移確率の逆転例 (ストウ夫人)

(先行句) — (後続句)	推移確率
(名+カ) $\xrightarrow{3.66\%}$ (形) $\xleftarrow{68.04\%}$ (名詞系)	24.89%
(名+ホ) $\xrightarrow{3.01}$ (形) $\xleftarrow{68.04}$ (名詞系)	20.46
(名+カ) $\xrightarrow{6.85}$ (形) $\xleftarrow{3.46}$ (名詞系)	2.37
(名+ノ) $\xrightarrow{12.33}$ (形) $\xleftarrow{3.46}$ (名詞系)	4.26

成立しやすいという結果になる。したがって、すくなくとも「ストウ夫人」のデータの範囲では、このタイプの構文に「ノ格」が選ばれやすいか、「カ格」が選ばれやすいかは、きわめて微妙

で、どちらとも決めにくい。もし、これを単純な度数カウントによって分析したならば、おそらく、こうした事実はとらえられないに相違ない。

構文の上の、句と句とのつながり具合や、関連性を求めたり、あるいは、句の順序や呼応などを数量的に解釈したりするさいには、単純な度数カウントで分析するよりは、やはり、確率過程の観点を導入した方が確実である。

また、従来の構文論や語順研究は、どちらかという、実際の発話や文章にもとづかず、整理された例文や演釈的な方法をもって分析を進める傾向が強かったように思われるが、研究対象を確率的な手法で、実際の文章からとり出すとともに、その分析においても、確率論的な方法を考えてみることも、また、必要なのではないかと思うのである。