

国立国語研究所学術情報リポジトリ

文節タイプ連続の研究：
「寒山拾得」のデータを使って

メタデータ	言語: Japanese 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 麿岡, 昭夫, TSURUOKA, Akio メールアドレス: 所属:
URL	https://doi.org/10.15084/00001038

文節タイプ連続の研究

——「寒山拾得」のデータを使って——

霧 岡 昭 夫

I はじめに

この小論は、先に著わした『電子計算機による代表構文作成の試み』（国立国語研究所論集「ことばの研究5」昭和49年3月・所収）を発展させた研究をまとめたものである。

上記論文では、文節タイプの二連続データを用いたので、三連続データを用いるとどうなるかという興味が持越された。また、地の文と会話文を一緒にカウントして分析したために、それぞれの傾向が分析できなかった。

ここでは、改良したシステムによりそれらの分析を行う。入力データは森鷗外の『寒山拾得』を用いた。『寒山拾得』のデータ量は、総センテンス220（地の文130、会話文90）、文節総数1683（地の文1232、会話文451）、異なり文節数200（地の文142、会話文99、両方に見られるもの41）である（注1）。

文節タイプのデータ作りは「電子計算機による代表構文作成の試み」に述べたとおりである。すなわち、文節（橋本文法による）を、自立語と付属語とに分けて、自立語部分には品詞名を、付属語部分にはその語形のままをあてたデータを原文ファイルから作り、これを文節タイプデータとするのである。このデータは固定長25桁にし、他に、地の文・会話の文の文種の別を示す情報を1桁加えてある。文節タイプが25字より短い場合は残余部分にⓈが埋められ、また、25字より長い場合は——殆んど起こらないことであるが、万が一起こった場合——「OVER・nn（nnは00～99までの二桁の整数）」の記号を埋め、そのnnの番号と実際の文節タイプを紙テープにパンチし、漢字テレタイプ印字にかけて文字化しておくようにしてある。

以上のようにして、『寒山拾得』の原文データ

唐の貞観の頃だといふから、西洋は七世紀の初……

という冒頭部分から、文節タイプデータ

㊦ 名+ノ 名+ノ 名+ダト 動+カラ 名+ハ 名+ノ 名 …………… [C]

が得られる(注2)。㊦ というのは文頭・文末を示す“くぎりデータ”である。

引用句の処理は、「」・『』・()の中を会話文のファイルBで処理し、地の文中ではその部分に“引用”という自立語相当のものを入れてファイルAで処理するようにする。例えば、

「お頭痛は」と僧が問うた。「あ癒りました。」実際聞は……という文から

㊦ 引用+ト 名+ガ 動+タ ㊦ 引用 ㊦ 副 名+ハ …………… [A]

㊦ 名+ハ ㊦ 感 動+マシタ ㊦ …………… [B]

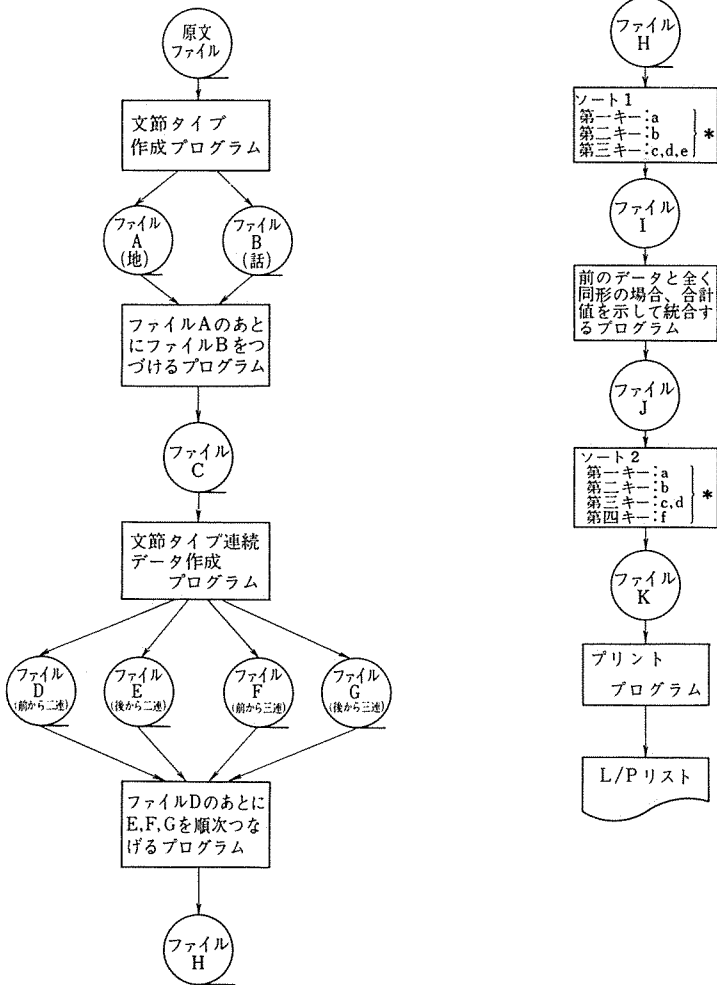
のように二本のファイルに文節タイプを作って行き、文節タイプ作成作業の済んだ後にファイルAのあとにファイルBをつなげてファイルC(→図1)を作るのである。

なお、上の例で、[A]の中に㊦ 引用 ㊦ という連続ができたが、このような場合には最初の㊦を残して引用㊦の二つを消去するように実際はなっている(注3)。

以上のようにして作られた文節タイプデータのファイルCをもとにして、前から二連続、後から二連続、前から三連続、後から三連続の、各種文節タイプ連続のデータを作るのである。文節タイプ連続の各データは、処理を一本するために、データ長を統一してあり、文節タイプが三つ並ぶようになっている(図2のc・d・eにあたる)。三連続データはその欄に一つずつはいるが、二連続

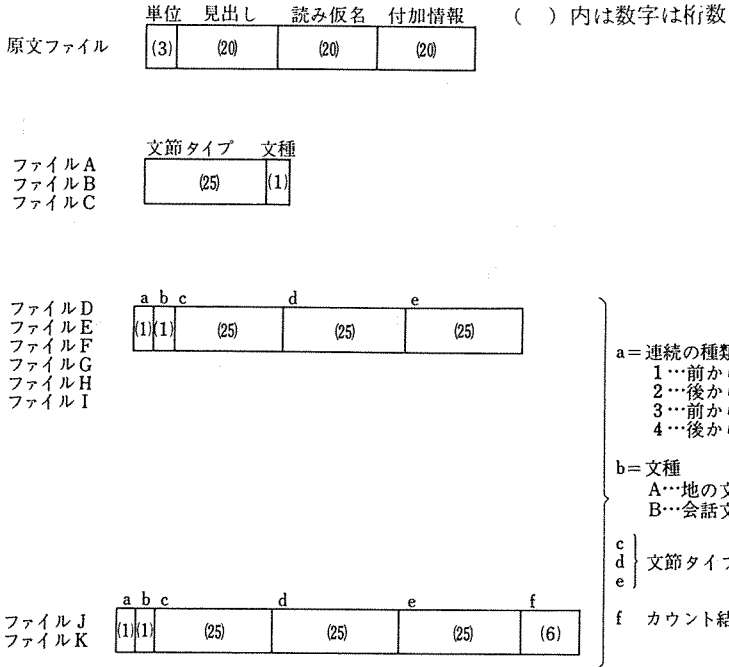
データの場合は、三つの欄のうち二番目と三番目を用い、先頭の欄にはⓈDを埋めておく。

図1. 作業の流れ

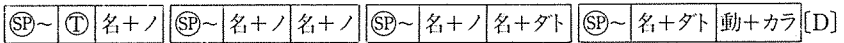


* は図2 参照

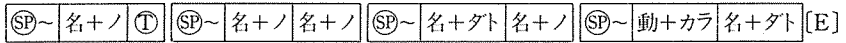
図2. 各ファイルのデータフォーマット



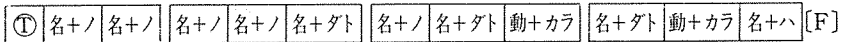
文節タイプデータ, 例えば先の [c] をもとにして, 前から二連続データ



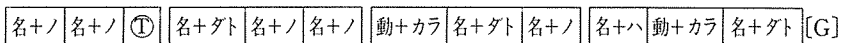
が作られる。この二番目と三番目を入れ換えると後から二連続データ



を得る。前から三連続データは, [c] をもとにして



となり, この一番目と三番目を入れかえると後から三連続データとなる。



連続データの作成が終わったらファイルDのあとにE, F, Gを続けて一本のファイル（ファイルH）にする。

ファイルHを、連続の種類別に、しかもその中を文種別に、そしてさらにその中の文節タイプ連続が五十音順に並ぶ（ソートされる）ように、第1ソートキーを連続の種類（図2のa）、第2ソートキーを文種（b）、第3ソートキーを文節タイプ連続（c・d・e）に定めてソートをする、それぞれの連続・文種の中で同じものが一個所に集まる。これを統合・カウントして、その値を第4ソートキーにして再度ソートし、L/Pに打ち出すと文節タイプ連続の一覧表が得られる。

————— 文 節 タ イ プ 連 続 リ ス ト —————

連続種類	文種	文 節 タ イ プ			カウント
(a)	(b)	(c)	(d)	(e)	(f)
1	A	ⓈP	メイシ+ヨリ	ケイヨウ	000001
1	A	ⓈP	メイシ+ラシク	メイシ+ノ	000001
1	A	ⓈP	メイシ	ドウシ	000004
1	A	ⓈP	メイシ	メイシ+ガ	000003
1	A	ⓈP	メイシ	メイシ+ノ	000003
}	}	}	}	}	}
1	B	ⓈP	ドウシ+テ	ドウシ+テ	000006
1	B	ⓈP	ドウシ+テ	メイシ+ニ	000003
1	B	ⓈP	ドウシ+テ	ドウシ+レマシタ	000003
1	B	ⓈP	ドウシ+テ	ドウシ+レマスカ	000003
}	}	}	}	}	}
3	A	メイシ+ヨリ	メイシ+テ	副 副	000001
3	A	メイシ+ラシク	メイシ+ハ	①	000001
}	}	}	}	}	0 }
4	A	ドウシ+テ	メイシ+ニ	メイシ+ノ	000006
4	A	ドウシ+テ	メイシ+ニ	ドウシ+テ	000003
4	A	ドウシ+テ	メイシ+ニ	ドウシ+タ	000002
}	}	}	}	}	}

実際には、プリント時に次のような処理をするようになって
いる。①(c)(d)欄が前と同じもの場合は、(c)(d)欄は印字せずにⓈPをうめる、②(c)(d)欄が前と等しくない場合は、それまでの(c)(d)欄の等しいものの合計(延べ数)をプリントする。

注1) 旧稿では、㊦—引用—㊦という連続の場合の「引用」を、一文節相

当に数えていたが、これは「引用」を、形式的に自立語相当のものと認めた意味からいって適当でないと考えられたので、本稿では削除した。そのために旧稿と本稿のセンテンス数、文節数は一致していない。

注2) 記号は、理解の補助手段であるにすぎないし、また、記号を考えに入れると記号の現われた所で文節タイプ連続の正確な確率が求められなくなるので無視した。「」『』() ! ? などが文頭、文末や引用句を探すのに用いられただけである。

注3) →注1

II 分 析

0 ここでは、文節タイプ連続の研究方法の確立に主眼がある。分析に用いた短編小説『寒山拾得』はほんのテストデータでしかない。しかし、この方法を用いて他の森鷗外の作品を多量に分析すれば森鷗外の文の特徴が明らかになるし、さらに多くの日本文を分析すれば日本語の文一般の特徴を明らかにすることができると考えられる。以上の分析は、その大きな狙いの先駆けをなすものである。

1 文頭と文末

Iで述べたようにして打ち出された文節タイプ連続リストを見れば、前から二連続データをもとにすると㊦のあとにどんな文節タイプが来やすいか、すなわち、どんな文節タイプで文が始まりやすいかが解り(→表1)、後から二連続データでは㊦の前にどんな文節タイプが来るか、すなわちどんな文節タイプで文が終わりやすいかということが解る(→表2)。また、前から三連続データを見れば、㊦のあとに、どんな文節タイプ連続が起りやすいか、すなわち、文がどんな文節タイプ連続で始まりやすいかが解り(→表3)、後から三連続データを見れば㊦の前にどんな文節タイプ連続が来ることが多いか、すなわち、文がどんな文節タイプ連続で終わりやすいかが解る(→表4)。

『寒山拾得』の地の文で文頭に立ちやすい文節タイプについて考察してみる

表 1 文頭に立ちやすい文節タイプ

地の文			会話文			全 体		
文節タイプ	数	%	文節タイプ	数	%	文節タイプ	数	%
名+は	25	19.23	感動	17	18.89	名+は	40	18.18
副	16	12.30	名+は	15	16.67	副	23	10.45
接続	13	10.00	接続	9	10.00	接続	17	7.72
名+が	12	9.23	副	7	7.78	名+の	17	7.72
名+の	12	9.23	名	5	5.56	感動	17	7.72
連体	8	6.15	名+に	4	4.44	名+が	15	6.82
名+に	6	4.62	名+の	4	4.44	名+と	12	5.45
引用+と	6	4.62	名+では	3	3.33	連体	11	5.00
形動	4	3.08	名+と	3	3.33	名+に	10	4.55
			名+が	3	3.33			
			連体	3	3.33			

表 2 文末に来やすい文節タイプ

地の文			会話文			全 体		
文節タイプ	数	%	文節タイプ	数	%	文節タイプ	数	%
動+た	49	37.69	動+ます	9	10.00	動詞+た	49	22.27
動	31	23.85	名+でございます	8	8.89	動	34	15.45
名+である	11	8.46	動+れますか	4	4.44	感動	11	5.00
動+ない	6	4.62	動+ました	4	4.44	名+である	11	5.00
動+のである	6	4.62	名+で	3	3.33	動+ます	9	4.09
動+たのである	3	2.31	名+か	3	3.33	名+でございます	8	3.63
形容	3	2.31	動+れました	3	3.33	動+ない	6	2.73
名+ほどである	2	1.54	動	3	3.33	動+のである	6	2.73
			動+ませう	3	3.33	形容	5	2.27

と、最も先頭に来やすい文節タイプは「名詞+は」で、ついで「副詞」・「接続詞」・「名詞+が」・「名詞+の」・「連体詞」の順で続く。これを『ストウ夫人』『城の崎にて』と比べてみると(注4)、第一位が「名詞+は」であることが同じであるうえ、上位六つが順番こそ違っているが、全く同じのものであるということも解る。そしてその上位六つでの生起率は、『ストウ夫人』で70.77%、『城の崎にて』で71.26%であったが、この『寒山拾得』でもやや低いながら66.5%とそれ程の違いがないということが言える。

『寒山拾得』の会話文の先頭は、第一位に「感動詞」が立つ。これが目立つ以外、第二位からはそれ程地の文との違いは感じられない。

文末ではどうかというと、地の文では、「動詞+た」が第一位である。そして前述の二つの作品と比べてみると文体、特に敬体でない点は『城の崎にて』に近いことが解るが、「動詞」の終束率が23.85%とやや高い（『城の崎にて』は13.87%）こと、「名詞+である」「動詞+のである」「動詞+たのである」のような「である」で終るものが多いということ、「動詞+なかった」の形で終わることが少ない（表にはないが1例）という点などは『城の崎にて』と異なる点といえる。むしろ、「名詞+です」「動詞+のです」「動詞+たのです」等の多く現われる『ストウ夫人』の方に近いという面も持っているようである。

会話文の文末では、地の文とかなり違っている。上位九つの文節タイプのうち、過去形「……た」のものが7.78%と非常に少なくなっていることがまず目につく。また、上位九つまでに、否定形「ない」「ず」等を含むものがみられないことも解る。そして、「名詞+で」「名詞+か」のように、名詞に助詞が付いたもので文を終るものがかかなりあり、特に「名詞+で」のような不完全な終止をするものもある。

次に前から三連続データを用い、文頭に来やすい文節タイプ連続を見てみる。まず、地の文では、第一位の「副詞→動詞+て」がわずか5例、生起率3.85%で、第二位の「名詞+は→名詞+を」「名詞+は→副」「引用+と→動詞+て」と1例差である。全話文においては第一位の「感動詞→㊦（これは感動詞一語の文であることを意味する）」が11例で12.22%と多い。この点が地の文と会話文との大きな違いであろう。また、表3で地の文と会話文との両方に現われるものは「名詞+は→名詞+を」「名詞+は→名詞+の」の二つだけで、かなり両者の間に違いがあることが解る。以上の二点以外には、地の文・会話文とも、文頭に來る文節タイプ連続は、さまざまなものが2~3%の生起率で用いられていると言うことができる。

文末に現われやすい文節タイプ連続、すなわち㊦の前に並びやすい文節タイ

表 3 文頭に

地の文				前文節タイプ
前文節タイプ	後文節タイプ	数	%	前文節タイプ
副	動+て	5	3.85	感動
名+は	名+を	4	3.08	名
名+は	副	4	3.08	名+に
引用+と	動+て	4	3.08	名+の
名+が	動+た	3	2.31	名+は
名+の	名+に	3	2.31	名+は
名+の	名+では	3	2.31	名+は
名+は	名+に	3	2.31	名+は
名+は	名+の	3	2.31	名+は
名+は	形動	3	2.31	接続
副	名+は	3	2.31	形動+でございます
連体	名+が	3	2.31	感動
				副
				連体
				動+ました

※自立語の略号は、「名」は名詞、「動」は動詞、「形」は口語
「接続」は接続詞、「形動」は形容動詞をさす。他に、「引

表 4 文末に

地の文				前文節タイプ
前文節タイプ	後文節タイプ	数	%	前文節タイプ
動+て	動	14	10.77	Top
名+に	動+た	9	6.92	名+は
名+を	動+た	9	6.92	動+て
名+が	動+た	7	5.38	形
動+て	動+た	7	5.38	Top
名+が	動	5	3.85	Top
文形	名+である	4	3.77	名+に
引用+と	動+た	4	3.77	Top
動+て	動+のである	4	3.77	名+と
				名+が
				動+て
				名+を
				名+を

※表 3 と同じ

来やすい文節タイプ連続

会 話 文			全 体			
後文節タイプ	数	%	前文節タイプ	後文節タイプ	数	%
Top	11	12.22	感動	Top	11	5.00
名+の	3	3.33	名+は	名+を	6	2.73
名+と	2	2.22	副	動+て	6	2.73
名+が	2	2.22	名+は	名+の	5	2.27
名+でございます	2	2.22	名+は	副	4	1.82
Top	2	2.22	連体	名	4	1.82
名+から	2	2.22	名	名+の	4	1.82
名+の	2	2.22	引用+と	動+て	4	1.82
名+を	2	2.22				
連体	2	2.22				
Top	2	2.22				
名+を	2	2.22				
形動	2	2.22				
名	2	2.22				
Top	2	2.22				

形容詞, 「文形」は文語形容詞, 「副」は副詞, 「感動」は感動詞, 「用」は引用句, 「Top」は文頭, 文末表示である。

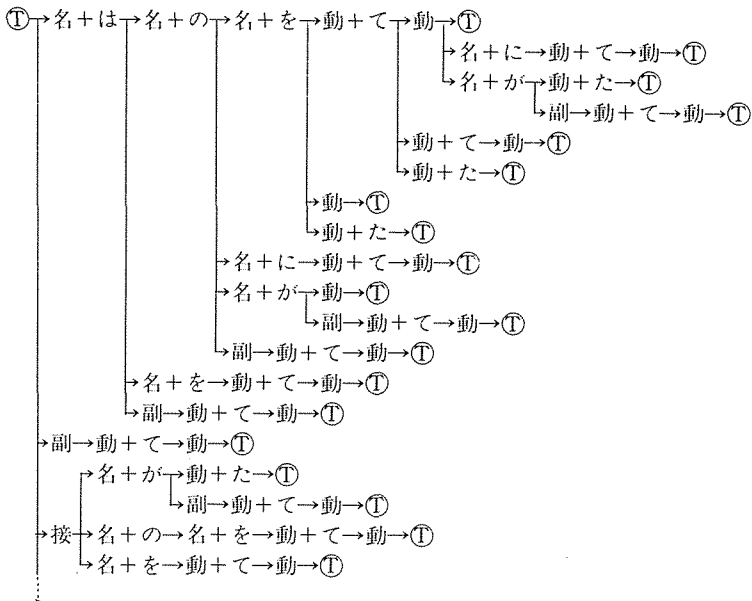
現われやすい文節タイプ連続

会 話 文			全 体			
後文節タイプ	数	%	前文節タイプ	後文節タイプ	数	%
感動	11	12.22	動+て	動	15	6.82
名+でございます	3	3.33	Top	感動	11	5.00
動+れました	2	2.22	名+を	動+た	9	4.09
名+でございます	2	2.22	名+に	動+た	9	4.09
形動+ございます	2	2.22	動+て	動+た	7	3.18
名+は	2	2.22	名+が	動+た	7	3.18
動+れますか	2	2.22	名+が	動	5	2.27
動+ました	2	2.22				
動+ます	2	2.22				
動+ます	2	2.22				
動+ます	2	2.22				
動+ます	2	2.22				
動+ませう	2	2.22				

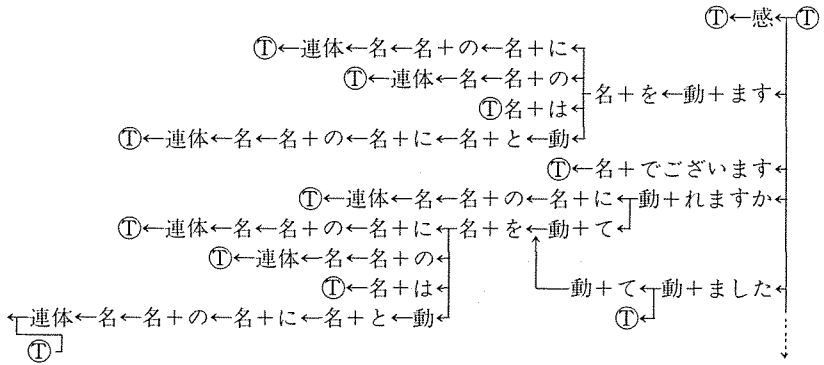
ブは表4に示すとおりである。地の文では「動詞+て→動詞」の終わり方が最も多く、次いで「名詞+に→動詞+た」「名詞+を→動詞+た」「名詞+が→動詞+た」「動詞+て→動詞+た」という、順当なものが上位を占めていて、その上位五つで終束率 35.37% を示している。一方、会話文では、「㊦→感動詞」という終わり方が11例と圧倒的で、3例の2位以下を引き離している。文頭でもそうであったが、文末でも第2例以下が3~2例で並ぶ。このような傾向は会話文の特徴であるとも推測される。

2 推移確率と近似文

文頭・文末表示の㊦を初期値として、連続確率の高い文節タイプを後、または前に順次つなげて行くと樹型測度が描かれる。二連続データをもとにして、文節タイプどうしの連続確率で描いた樹型の一部を示すとつぎのようになる。(途中でループしてしまうものは、2回目に第2位をとることにした。また、そのあとが頻度1で分散してしまうものは()で囲んで以下を省略した。)
 <前から二連続データによる、会話文の樹型の一部>



〈後から二連続データによる、地の文の樹型の一部〉



この樹形で、文頭、または文末の①から、矢印の順に文末あるいは文頭の①までをたどっていくと一続きの文型が得られる。このうち、最も連続確率の高い文節タイプ連続でできた文、すなわち各樹型の一番上に現われた文を一次近似文と呼ぶ。『寒山拾得』の地の文で文頭からたどった一次近似文は、

$$\textcircled{1} \xrightarrow{19.23} \text{名+は} \xrightarrow{18.18} \text{名+の} \xrightarrow{21.57} \text{名+を} \xrightarrow{37.50} \text{動+て} \xrightarrow{19.70} \text{動} \xrightarrow{32.63} \textcircled{1}$$

ということになる(注5)。地の文で文末からたどった一次近似文は

$$\textcircled{1} \xleftarrow{10.58} \text{名+の} \xleftarrow{20.19} \text{名+に} \xleftarrow{17.57} \text{動+た} \xleftarrow{37.69} \textcircled{1}$$

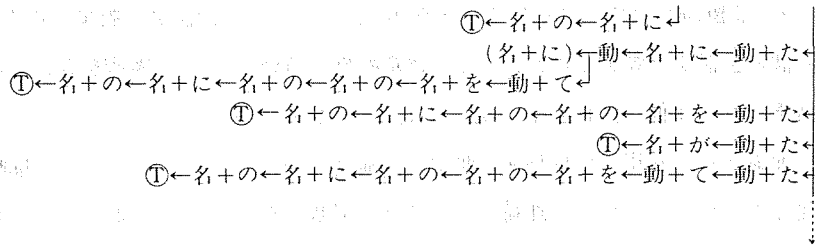
$$\textcircled{1} \xleftarrow{10.58} \text{名+の} \xleftarrow{21.15} \text{名+を} \xleftarrow{17.57} \text{動+た} \xleftarrow{37.69} \textcircled{1}$$

の二つである。

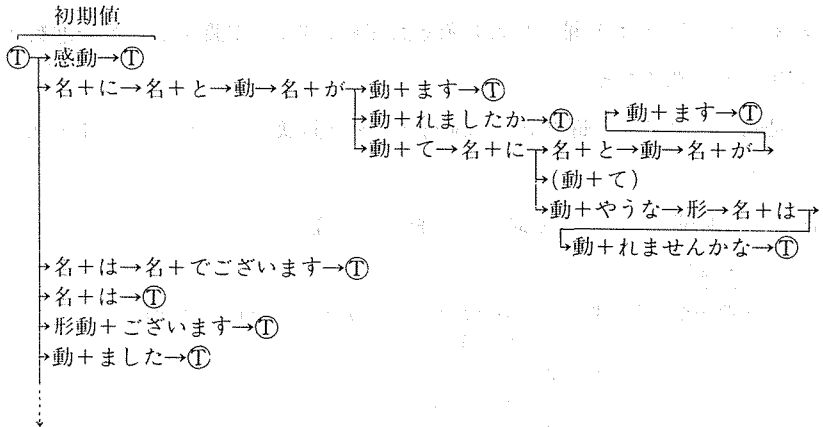
会話文では、文頭、文末からの一次近似文は、それぞれ

$$\textcircled{1} \xrightarrow{18.89} \text{感動} \xrightarrow{64.71} \textcircled{1}$$

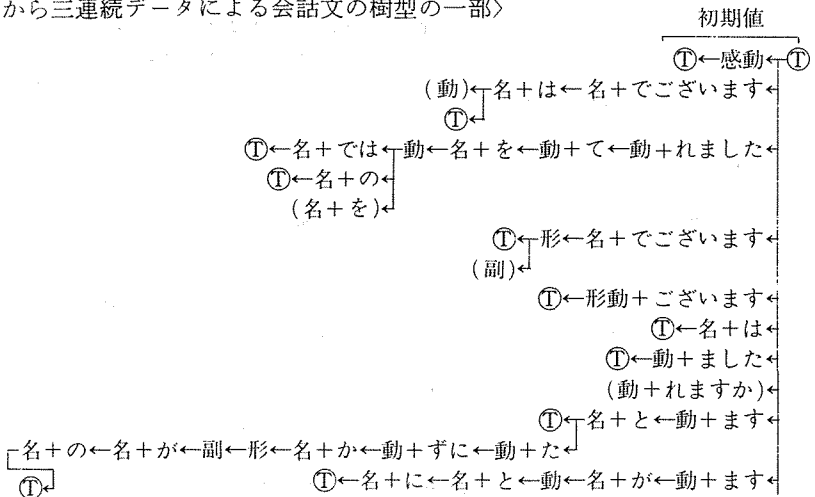
$$\textcircled{1} \xleftarrow{100.00} \text{感動} \xleftarrow{12.22} \textcircled{1}$$



〈前から三連続データによる会話文の樹型の一部〉



〈後から三連続データによる会話文の樹型の一部〉



のよりも現実に起こりやすい、不自然でない文型であることがわかる。また、その頻度もあまりかたよりを見せていないもので、このことから各種の文章を扱えば、その文章ごとに特徴的な近似文の合成されることが期待できそうである。そして、これらの樹型・近似文を見わたしてみると、「名詞+を」というタイプの文節は、「名詞+に」「名詞+へ」などの文節よりも前に来ることが少い、などといった、構文にかかわる問題が明らかになってくる。それについては次項で、一部述べる。

注4) 『文章における構文のタイプ』(田中章夫・「計量国語学」65)昭和48年6月・所収)による。

注5) 以下、矢印の上の整数は、出現数、下の数字は出現率(単位%)を示す。

3 構文にかかわるいくつかの指摘

文節タイプの連続データを用いると、前項末で述べたように、構文にかかわるさまざまな様相が明らかになる。

二連続データを使って、「名詞+が \leftrightarrow 動詞+たい系(助動詞「たい」及びその活用形を含んだ付属語部分を有するもの)」、「名詞+を \leftrightarrow 動詞+たい系」を調査すれば、「水が飲みたい」型、「水を飲みたい」型の比率が解る(しかしそのような連続は『寒山拾得』には現れていない)。また、三連続データの「名詞+の \leftrightarrow 名詞+は \leftrightarrow 形容系」と「名詞+は \leftrightarrow 名詞+が \leftrightarrow 形容系」とを比較すれば、「象は鼻が長い」型構文と「象の鼻は長い」型構文との使用状況が解る(これも『寒山拾得』には現れないが――)。

「雨の降る日」型の構文と「雨が降る日」型の構文とは、『寒山拾得』において表5のような傾向が見られる。すなわち、連体修飾語となる動詞句の主語には「――の」の方が出現度数が高い。また、前からたどって「名詞+の」をもとにして、「動詞系連体語」を経て「名詞系(名詞または名詞+助詞・助動詞)」と並ぶ確率は5.88%であり、「名詞+が」から「動詞系連体語」を経て「名詞系」と並ぶ確率4.34%よりやや高い。同様にうしろからたどって「名詞系」から「動

詞系連体語」を経て「名詞+の」と並ぶ確率 0.98% も、「名詞系」から「動詞系連体語」を経て「名詞+が」となる確率 0.49 より高い。これらのことから、この『寒山拾得』においては、「雨の降る日」型の構文のほうが「雨が降る日」型よりも多く用いられているとすることができる。

表 5 「雨が降る日」型と「雨の降る日」型 (地の文)

文節タイプ連続	出現 度数	推 移 状 況		確 率 (%)	
		前→後	前←後	前→後	前←後
名+が 動系連体語* 名系**	3	69→32→3	3←96←611	4.34	0.49
名+の 動系連体語 名系	6	102→29→9	6←96←611	5.88	0.98

* 自立語が動詞で連体修飾語となっている文節

** 自立語が名詞であるすべての文節

さらに極端なことは、「背の高い人」型と「背が高い人」型の場合である(→表 6)。すなわち、『寒山拾得』では、連体修飾語となる形容詞句の主語にはすべて「名詞+の」が来ている。

つぎに、「AがBに……する」型と「BにAが……する」型、「AがBへ……する」型と「BへAが……する」型、「AがBを……する」型と「BをAが……する」型、「AがBを……する」型と「BをAに……する」型、「AへBを……する」型と「BをAへ……する」型などのそれぞれの比較結果が明らかになる(→表 7)。

さらに「名詞+が」のタイプの文節について考察すれば、「名詞+が」は、「名

表 6 「背が高い人」型と「背の低い人」型 (地の文)

文節タイプ連続	出現 度数	推 移 状 況		確 率 (%)	
		前→後	前←後	前→後	前←後
名+が 形容連体語*** 名系	0				
名+の 形容連体語 名系	3	102→3→3	3←24←611	2.94	0.49

*** 自立語が形容詞で、連体修飾語となる文節

表 7 「一が」「一を」「一に(へ)」相互の前後関係 (地の文)

文節タイプ連続	出現度	推移状況		確率 (%)	
		前→後	前←後	前→後	前←後
名+が 名+に 動系***	5	69→5→5	5←67←450	7.25	1.11
名+に 名+が 動系	4	96→5→4	4←32←450	4.17	0.89
名+が 名+へ 動系	0				
名+へ 名+が 動系	1	9→1→1	1←32←450	11.11	0.22
名+が 名+を 動系	3	69→3→3	3←90←450	4.35	0.67
名+を 名+が 動系	0				
名+に 名+を 動系	9	96→9→9	9←90←450	9.38	2.00
名+を 名+に 動系	4	104→4→4	4←67←450	3.85	0.89
名+へ 名+を 動系	0				
名+を 名+へ 動系	1	104→1→1	1←8←450	4.35	0.67

*** 自立語が動詞である文節

詞+を」「名詞+に」などよりも前に来る確率が高い、とすることができる。まず第一に、「名詞+が」文節は、「名詞+を」の前に来るのが3例で、「名詞+を」の後に来る例はなく、「名詞+が」は「名詞+を」の前に来る確率は圧倒的に高いと言えそうである。第二に、「名詞+が」文節は、「名詞+に」文節の前に来る場合が5例で、「名詞+に」文節の後に来る場合の4例より、僅かではあるが、「名詞+に」文節の前に来るほうが多く、また、「名詞+が」・「名詞+に」「動詞系」を前からたどった確率・後からたどった確率のどちらもが、「名詞+に」「名詞+が」「動詞系」の前・後からたどった確率よりもそれぞれ高くなっていることが解る。以上のことから「名詞+が」は「名詞+を」「名詞+に」よりも前に来ることが多いと言えそうである。ただ、「名詞+が」「名詞+へ」「動詞系」と並ぶ例が見当たらないのに対して、「名詞+へ」「名詞+が」「動詞系」と並ぶ例が1例見られるが、これだけのデータでは「名詞+が」が「名詞+へ」の後に来やすいとは明言はできない。とにかく「詞+が」と「名詞+を」「名詞+に」「名詞+へ」との前後関係については、もっとデータ量を多くして考察をする必要があるように感じられる。

III おわりに

以上のように、文節タイプ連続を考察すると、文頭、文末の様相、文体的特徴、構文の成立に関するさまざまな様相が解る。もちろん、この論文は単一の作品の中での考察がほとんどで、これをもって日本語の特質とすぐに結びつけるのは危険である。しかし、いろいろの文章を分析してそれぞれの特徴を示せばそれはそれなりの効果があり、また、それを総合して、日本語の特質にせまることも必要なことであると思われる。そしてまた、文節タイプを、特に文頭から追った場合、著者・話者の潜在的な構文意識（癖・好みなど）の現われることが期待される。この研究は、それらのことを目指しているのである。

最後に、この研究に電子計算機を利用した点での利点を述べる。電子計算機では一担プログラムを組んでおけば、あとはどんな長いデータ（原文）を用いてもすべて自動的に最終的な文節タイプ連続リストを、前から二連続、後から二連続、前から三連続、後から三連続と四つ打ち出してくれる。手作業で行なうカード取り（上の四つの表を作ろうとすれば一度に4枚必要）、カード分類、集計、製表等の作業は全く必要としない。研究者は、校正（エラーデータ修正）、簡単なオペレート（電子計算機操作）をして、あとはリストをひくだけとなる。

電子計算機を用い、大量データを確率統計的に処理する方法による日本語の各種文章の分析、およびそれをもとにした日本語全体の分析は、以上述べたようなやり方で進められている。現在、入力データの整備、プログラムの改良を行なっている段階で、他の作品による文節タイプ連続の研究も近い将来発表する予定である。