

# 国立国語研究所学術情報リポジトリ

## 電子計算機による代表構文作成の試み

メタデータ	言語: Japanese 出版者: 公開日: 2019-02-15 キーワード (Ja): キーワード (En): 作成者: 麿岡, 昭夫, TURUOKA, Akio メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00001782">https://doi.org/10.15084/00001782</a>

# 電子計算機による代表構文作成の試み

鶴岡 昭夫

## I・方法論

### 1・1

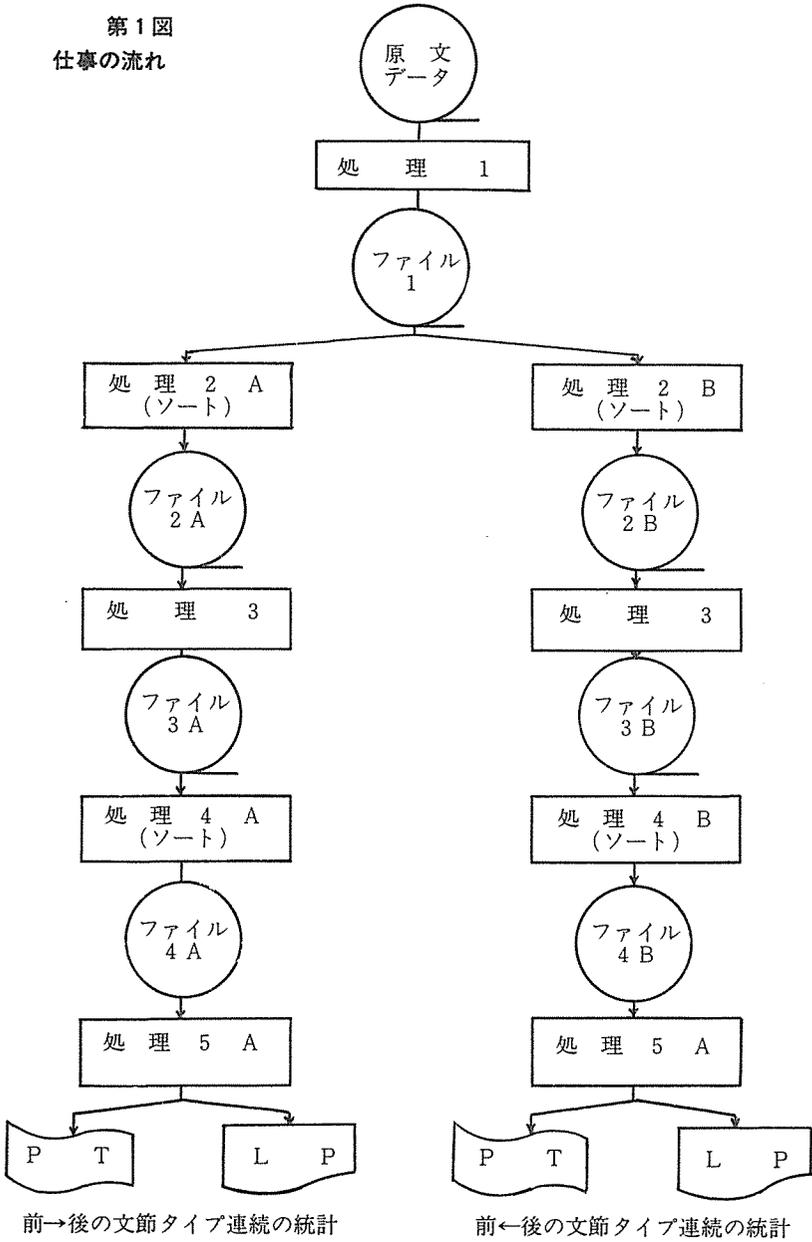
国立国語研究所には、現在、電子計算機が稼動している。<sup>(注1)</sup>そしてわれわれは、これを用いてさまざまな言語研究活動をしている。ここで、これから述べるのは、この研究のうちの、「漱石・鷗外の用語研究」の一環として開発を進めているシステムについてである。

文の集合(文章)の中で、統計的に文頭あるいは文末に来る確率(推移確率)の最も高い文節タイプ<sup>(注2)</sup>を初期値として、後または前に最も来やすい文節タイプを、順次つなげていくと、その文章において最も起りやすい文節タイプ連続で構成された文型が出来る。これをここでは代表構文と呼ぶ。このように推移確率を用いて代表構文を作ることについては、「文章における構文のタイプ」(田中章夫『計量国語学』65集に収録)に詳しく論じられているが、これからここに表わす小論は、その作業を電子計算機に行わせる方法、およびそれによって得られた結果、得ようとするねらい等について論ずるものである。

ここで電子計算機を使用した大きな理由は二つある。一つは、前に述べたように、電子計算機を用いての「漱石・鷗外の用語研究」が進行中であって、そのデータが大量に作られて来るから<sup>(注3)</sup>それを利用しようという消極的理由からで、もう一つは、大量のデータを用いて総計確率的な仕事をするのに電子計算機が適しているという積極的理由からである。

(注1) 国立国語研究所には、昭和41年以降HITAC3010型電子計算機(容量2万字)が稼動しているが、昭和48年度中にHITAC8250型電子計算機(容量9万6千字)が導入されることになっている。HITAC3010用のデータやプログラムは、そのまま

第1図  
仕事の流れ



はHITAC8250に用いることができないので、現在、国立国語研究所では、メーカーのバックアップのもとにコンバート（変換）作業を進めている。ここでは3010型用のプログラム、データとして解説した。

(注2) ここに言う「文節タイプ」とは、文節中で、自立語部分（接辞も含める）は品詞名、助詞・助動詞部分は語形という形にして文節の種類を表わしたものをいう。例えば「行きました」を〈動詞+ました〉、「父のです」を〈名詞+のです〉という形で表わしたものがそれである。

(注3) 「漱石・鷗外の用語研究」のために作成した、または作成中のデータには、カナ索引作成用データと類別索引作成用データとがあるが、ここで使用するのは後者の方である。現在の所では、『寒山拾得』『雁』『山椒大夫』『渋江抽斎』『草枕』『坊っちゃん』の六点が計画されている。このうち、『寒山拾得』はすでに完成しており、他の五つもその作業は順調に進行して、昭和48年から49年度にかけてつぎつぎに完成する予定である。

## 1・2

この作業の流れは第1図のとおりである。すなわち、各種のプログラムやサービスルーチン<sup>(注4)</sup>を用いて、原文データを文節タイプ二連続データに変え、それを適当に並べかえたり、同じ文節タイプ連続を合計したりしてその確率を計算してそれをラインプリンター（LP）や紙テープ（PT）に出力するのがこの作業のあらましである。

### ㊤ 処理1と入出力ファイル

処理1はプログラムを用いて、原文データを入力し、それから文節タイプ二連続データ（データ1）を作るものである。

ここで入力に用いるファイル（データを集合させたもの）は、前述のとおり「漱石・鷗外の用語研究」用のものである。<sup>(注5)</sup>これは、作品の文章を要素（語より小さな単位）に分けたもので、その要素が作品順に並んでいる。1データは64桁の固定長で、各情報は下の第2図のとおりである（カッコ内の数字は桁数を表わす）。

第2図 入力データのフォーマット

単位 (3)	見し 出 (20)	読仮 み名 (20)	品 詞 (6)	ペジ ー (4)	行 (2)	文 種 (1)	題 (1)	段 落 (4)	エ ラ ー (2)	E / i (1)
-----------	-----------------	------------------	---------------	----------------	----------	---------------	----------	---------------	--------------------	--------------------

まず、この入力データの要素1内至数個から文節タイプをワークテープ（第1ワーク）に作るのが第一段階であるが、その際、本文以外の題・見出し・注など（単位情報が0001～9999の間の数でないもの）は無視する。また、本文であっても、見出し情報が“，”（カンマ＝句点に相当）であるものは、これも無視する。<sup>(注6)</sup> こうして無視したもの以外のデータの、単位情報、読み仮名情報、<sup>(注7)</sup> 品詞情報を<sup>(注8)</sup>をもとにして文節タイプを作るのである。

単位情報は、その要素が文節頭であることを示す「CLS」と、文節内の助詞助動詞部分の先頭であることを示す「△LS」と、文節の先頭でも、助動詞動詞部分の先頭でもないことを示す「△△S」の三種類がある。<sup>(注9)</sup> 単位情報が「CLS」であるもの、及びそれに続いて次の「△LS」や「CLS」までの間にある「△△S」は自立語部分であるから、その品詞情報にもとづいて品詞名をワークエリアに書き込む。この時、後から書き込まれるものを優先させれば、複合語の品詞は、最後の要素の品詞で決まるから、「立ち上る」「出席する」等は動詞に、「うすぐらい」は形容詞に、また「貯金箱」は名詞になっている。また「うつくしげ」「さびしがる」はそれぞれ、名詞性接辞、動詞性接辞となってしまうから、それを名詞、動詞になおすようにすればよいのである。次に、「△LS」、及びその後が続いて次の「CLS」の前に来る「△△S」は、助詞助動詞であるから、ワークエリアで自立語の品詞名を書き込んだ後に順々に続けて行けば文節タイプが完成する。そして次の「CLS」単位を読みこんだらワークエリアのデータを第1ワークに書きこむ。

〔例1〕	(単位)	(読みがな)	(品詞)	(文節タイプ)
	CLS	りよ	8	} <名詞+と>
	△△S	きゅういん	8	
	△LS	と	R	} <動詞>
	CLS	いう	E	
	CLS	かんり	1	} <名詞+が>
	△LS	が	R	
	CLS	い	E	} <動詞+たそうである>
	△LS	た	P	
	△△S	そうで	P	
	△△S	ある	P	} <.>
	CLS	.	Y	

文節タイプが完成したら、第二段階として、そのファイル（第1ワーク）をもとに引用句（会話部分等）の処理を行なわなくてはならない。それは、引用句が、文の中で1文節相当の機能を果していながら、その中が1文ないし数文から出来ているからである。

〔例2〕 父は「ああ。そうか。わかった」と言った。  
もし、上の例2のような文を、電子計算機で単純に“。”を文末として切っていくと、“父は「ああ」「そうか」「わかった」と言った”の三文に切れてしまうし、それに伴って文節タイプの連続確率の計算も不正確になってしまうのである。これを防ぐためには、第1ワークのデータを別なワーク・テープ（第2ワーク）に転写していき、「」や『』内の部分を〈引用句〉とか〈引用句+と〉とかの文節タイプ相当句におきかえ、同時にその引用カッコの中の文を他のワークテープ（第3ワーク）に移しておき、第1ワークのデータの終了後、第2ワークの後に続けるようにする。<sup>〔注10〕</sup>このようにすることにより、〔例2〕の文は次のように処理されるのである。

〔例3〕 〈名詞+は〉 〈引用句+と〉 〈動詞+た〉 〈。〉 ……………  
… 〈感動〉 〈。〉 〈副詞+だ〉 〈。〉 〈動詞+た〉 〈。〉

このようにして処理のすんだファイル（第2ワーク・テープ）から、文節タイプ二連続データのファイル（ファイル1）を作るのが第三段階で、処理1の最終段階である。

文節タイプ二連続データは、1データ81桁の固定長で、各情報は、第1文節タイプ情報40桁、<sup>〔注11〕</sup>第2文節タイプ情報40桁、E/iマーク1桁である。その作り方は、第2ワークに出来ている文節タイプを一つずつ読み、その一回前に読んだデータを第1文節タイプ情報とし、今読んだものを第2文節タイプ情報とする、ということを順にくりかえして二連続の文節タイプデータを作るのであるが、一番最初は、第1文節タイプ情報に文頭表示記号①を入れておく。また、文の始まる前、すなわち文頭句の前にも①記号のはいるようにする。この①語号は、文節タイプ一つ分に相当するものとして、推移確率計算に用いる。

以上のやり方で〔例1〕を処理すると次のようになる。

〔例 4〕	通し番号	(第 1 文節タイプ情報)	(第 2 文節タイプ情報)	E / i
	1	<Ⓓ>	<名詞+と>	E / i
	2	<名詞+と>	<動詞>	E / i
	3	<動詞>	<名詞+が>	E / i
	4	<名詞+が>	<動詞+たそうである>	E / i
	5	<動詞+たそうである>	<.>	E / i

#### ② 処理 2 と入出力ファイル

処理 2 は、処理 1 で得たファイル 1 を入力ファイルとして、HITAC3010コード順に配列させる処理（ソート処理）である。ソートにはサービスルーチンを用いる。

ソートでは、並びかえる基準の情報（キー）を、第 1 キー、第 2 キーの二つ指定しておけば、全データは第 1 キーの情報の配列順に並び、その中で、第 1 キーの情報の等しいものは、第 2 キーの配列順に並ぶということになる。

前→後への文節タイプ連続を調べる場合（処理 2 A）は、入力データの第 1 文節タイプ情報を第 1 キー、第 2 文節タイプ情報を第 2 キーとすれば、文節タイプ連続で前部分が優先された配列法となってファイル 2 A が得られる。前←後の文節タイプ連続の場合（処理 2 B）は、第 1 キーを第 2 文節タイプ情報、第 2 キーを第 1 文節タイプ情報とすれば、後部分の文節タイプを中心として配列されたファイル 2 B が得られる。出力ファイルのファイル 2 A、ファイル 2 B のデータ長は入力ファイルと同じく固定長 81 桁である。

#### ③ 処理 3 と入出力ファイル

ファイル 2 A、ファイル 2 B では、第 1 文節タイプ情報も第 2 文節タイプ情報もともに等しいデータ、すなわち同一の文節タイプ連続のデータは一個所に集まって並んでいる。処理 3 は、ファイル 2 A またはファイル 2 B を入力して同一の文節タイプ連続は一つにまとめ、合計数を出すようにし、ファイル 3 A またはファイル 3 B を出力するものである。出力ファイルのデータ長は、カウント情報（10 桁）のふえた、固定長 91 桁である。

#### ④ 処理 4 と入出力ファイル

処理 3 で得たファイル 3 A の、第 1 文節タイプ情報の順番は変らせないう

にし、その第1文節タイプ情報が同形の場合その中で並ぶ第2文節タイプ情報の順番を、頻度の高いものの順に並びかえるのが処理4Aである。この作業も処理2と同様にサービスルーチンのソートを用いるのであるが、上の処理目的から、第1キーを第1文節タイプ情報、第2キーをカウント（下降順）と定める。たとえば、一部分を示すと、

(カウント)		(第1文節タイプ)		(第2文節タイプ)
⋮		⋮		⋮
0000000028	下 降 順	<名詞+に>	—  致	<動詞+て>
0000000013		<名詞+に>		<動詞+た>
0000000013		<名詞+に>		<名詞+を>
0000000011		<名詞+に>		<動詞>
0000000009		<名詞+に>		<名詞+の>
0000000005		<名詞+に>		<名詞+が>
0000000005		<名詞+に>		<名詞+と>
0000000004		<名詞+に>		<動詞+ても>
0000000008	下 降 順	<名詞+には>	— 致	<名詞+を>
0000000004		<名詞+には>		<名詞+と>
0000000002		<名詞+には>		<名詞+や>
⋮		⋮		⋮

というようになるのである。このようにデータを収めたものがファイル4Aである。前←後の文節タイプ連続を調べる処理（処理4B）では、ファイル3Bを入力ファイルとし、第2文節タイプ情報を第1キー、カウント情報を第2キーとしてソートする。それで得られるファイル（ファイル4B）のデータは、一部分を示すとつぎのようになっている。

(カウント)		(第1文節タイプ)		(第2文節タイプ)
⋮		⋮		⋮
0000000005	下 降 順	<動詞>	—  致	<名詞+も>
0000000004		<㊦>		<名詞+も>
0000000001		<動詞+た>		<名詞+も>
0000000001		<動詞+だ>		<名詞+も>
0000000001		<名詞+は>		<名詞+も>

0000000002	下 降 順	<名詞+が>	<名詞+や>	一 致
0000000002		<名詞+や>	<名詞+や>	
0000000001		<接続>	<名詞+や>	
0000000001		<動詞>	<名詞+や>	
⋮		⋮	⋮	

㊤ 上で得たファイル4 A, ファイル4 Bに書き込まれたデータを, 人間が読める形にラインプリンター (LP) や漢字テレタイプ用の紙テープ (PT) へ出力する作業が処理5 (処理5 A, 処理5 B) である。

(注4) 電子計算機のメーカーが開発した, 作業用の処理システム。

(注5) 第2図のようなフォーマットを持ち, 作品出現順になっているデータならば, 「漱石・鷗外の用語研究」用のもの以外でもこのシステムに用いることができる。

(注6) 句点は, 文中休止を示すのみで, 文構成の上ではあまり意味がない。しかも, 句点はどんな句の間にでも来ることがあるし, その数も多いから, これを残す, すなわち一文節相当に扱うと, 文節タイプ連続の確率計算が不正確になってしまうからである。

(注7) 見出し情報を用いたのでは, 原文での表記が異なると, 例えば「まで」と「迄」のような異形同詞が別語扱いになってしまう。これを防ぐために読み仮名情報を用いるのである。なお, 記号データの場合, 読み仮名情報に記号のままはいつているので問題はない。

(注8) 品詞情報は6桁であるが, それは語種コード (1桁), 品詞コード (1桁), 活用コード (1桁), 補助コード (2桁) に分かれている。このうち, ここで用いるのは2番目の品詞コードである。品詞コードには, その要素のもつ文法的性格から, 次のうちのどれか一つがいつている。

1～9 (純名詞, 代名詞, 名詞性接辞, 固有名詞, 形容動詞語幹, サ変動詞語幹など, 各種の名詞) ・E (動詞) ・+ (動詞性接辞) ・M (口語形容詞) ・N (文語形容詞) ・- (形容詞性接辞) ・A (接続詞) ・B (感動詞) ・C (副詞) ・D (連体詞) ・X (数詞・数字) ・P (助動詞) ・R (助詞) ・Y (記号) ・Z (品詞不明)

なお, 上のうち, P (助動詞) ・R (助詞) ・Z (品詞不明) のコードを持つものは, 文節タイプにする時に品詞名に変換せずに, 本文で用いられた形のままで用いるものである。

(注9) 類別索引用データは, 単位切り作業が, ①原文を文節単位 (C単位) に切る, ②文節を自立語部分と助詞・助動詞部分に分ける (それぞれをL単位とする), ③L単位内を要素 (S単位) に分ける, の三段階で行なわれている。この単位切り作

業の結果、各要素（S単位）は、①C単位の先頭、②後半のL単位（助詞・助動詞部分の先頭）、③上のいずれでもないもの（L単位内で二番目以降のS単位）、の三種類になり、それぞれに「CLS」「ΔLS」「△△S」の単位情報が付くのである。

(注10) ここでは一応、引用句中の文を、あとで地の文と一緒にしてカウントする方法を取る。ただし、地の文だけ、あるいは引用句中の文だけで分析することは必要であり、可能である。将来は、それぞれのやり方でやって行く予定である。

(注11) HITAC3010コードは、6ビットなので、全部で64種類しかない。そこで国立国語研究所では、それを二字ずつ組み合わせた4096種の漢字テレタイプコードを用いている。したがって、HITAC3010コードで44桁は漢テレで20字分にあたる。なお、各文節タイプ情報で、データ長が40桁より短い場合は、余りの部分にスペース（Δ）がはいるようにする。

## II・分析

### 2・0

前項で述べた方法で、『寒山拾得』のデータを分析する。この分析結果には特に目新しいことはないかも知れない。しかし、ここでは、電子計算機で推移確率を出し、代表構文を作ることができる、ということを示せば主な目的が達せられるのである。この方法さえ確立しておけば、データソースとしての作品の種類が増えた時にはもっとさまざまな分析が出来るのである。

『寒山拾得』のデータ量は、総センテンス数259、文節数1729である(①および“.”は1文節と数えない)。

### 2・1

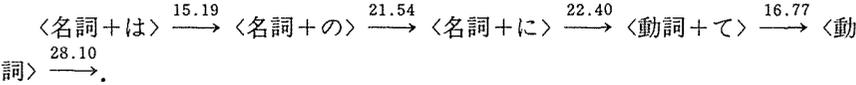
『寒山拾得』で、文の先頭に来やすい、すなわち、処理5Aで文頭表示マーク①のあとに来る回数の多い上位八つを、順に並べたものが右の表1である。この結果と、『ストウ夫人』『城の崎にて』の結果（「文章における構文のタイプ」10ページ）

表1 文頭に来やすい文節タイプ

順位	文頭句の文節タイプ	度数	生起率(%)
1	<名詞+は>	40	15.44
2	<引用句>	36	13.90
3	<副詞>	20	7.72
4	<名詞>	16	6.18
4	<感動詞>	16	6.18
6	<名詞+が>	15	5.79
8	<名詞+の>	15	5.79
8	<接続詞>	13	5.02
総文数259、文頭文節タイプの異り数=43			

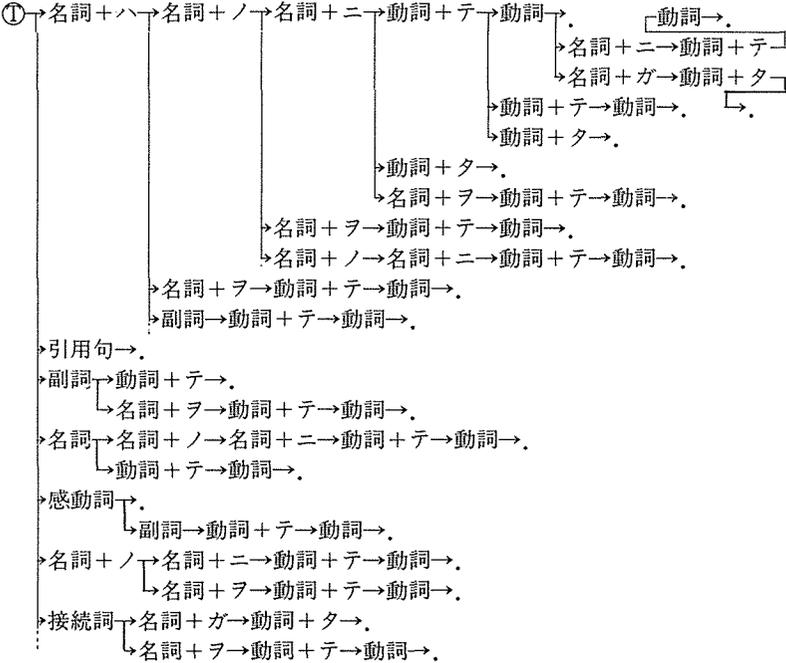
と比べてみると、『寒山拾得』では〈引用句〉が多いという以外は、大体一致している。すなわち、『寒山拾得』でも〈名詞+は〉が第一位を占め、また、〈引用句〉は厳密な意味での文頭句とは言えないからこれを除く223センテンスで考えると〈名詞+は〉〈副詞〉〈名詞〉〈感動詞〉〈名詞+が〉〈名詞+の〉〈接続詞〉の七種で135例60, 54%を占める、などのことがわかる。

つぎに、文頭に最も来やすい〈名詞+は〉からつぎつぎに、推移確率の最も高い文節タイプをつなげて、ピリオドが現われるまでの一文を作ると、次の構文が得られる(矢印上の数字は推移確率を示す百分比)。



これが文頭からたどった代表構文ということになる。

また、文頭表示①から、あとに続く文節タイプをどんどんつなげて行く、いわゆる樹型測度の、一部を示すと、次のようになる。



この樹型では、例えば〈名詞+に〉→〈名詞+を〉の順に並びやすい、という  
 ような構文上の問題が示唆されている。

## 2・2

文は先頭と末尾を持つ訳である。したがって、推移確率も、前項でのべたよ  
 うな、前→後の文節タイプ連続のほかに、前←後の順に追って行くこともでき  
 る。前節の処理5 Bで得られたLPおよびPTの出力結果は、それを調べるた  
 めのものである。

表2 文末に来やすい文節タイプ

順位	文末句の文節タイプ	度数	終束率
1	〈動詞+た〉	49	18.92
2	〈引用句〉	36	13.90
3	〈動詞〉	34	13.13
4	〈感動詞〉	11	4.25
5	〈名詞+である〉	10	3.86
5	〈名詞+でございます〉	10	3.86
7	〈動詞+ます〉	9	3.47
8	〈名詞+のである〉	6	2.32
文総数=259,文末文節タイプ異り数=65			

まず、『寒山拾得』では、どんな  
 文節タイプが文末に来やすいか、す  
 なわち、ピリオド“.”の前に来る確  
 率の高いものはどのような文節タイ  
 プであるかを示すのが右の表2であ  
 る。前の『ストウ夫人』『城の崎に  
 て』の結果と比べてみるまでもなく、  
 常識的な結果が出ている。そして、  
 〈引用句〉を除く223文のうち、上

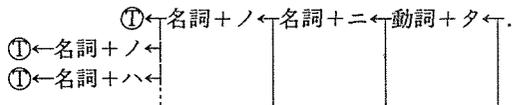
位7位で終わるものが53.36%、また〈動詞+た〉と〈動詞〉のいずれかで終  
 わるものが37.22%を占めている。ただ表2結果と、『ストウ夫人』『城の  
 崎にて』の結果との違いは、『寒山拾得』に〈引用句〉が多いというほか、『寒  
 山拾得』では、否定形で終わる文節タイプが上位に来ない、ということである。  
 これは『寒山拾得』の文体的な特徴と言える。

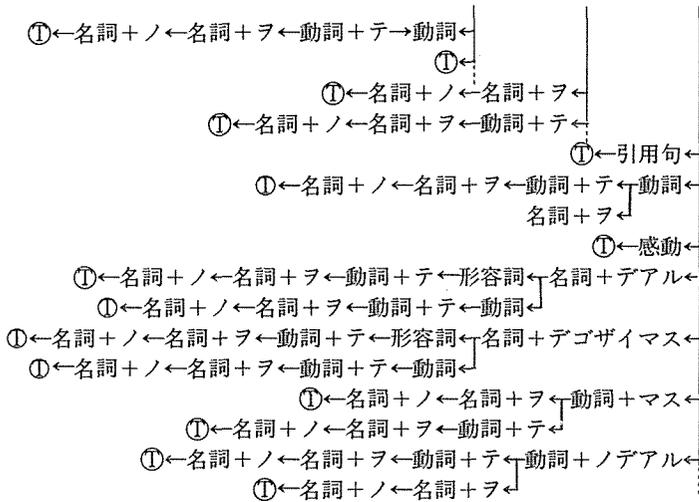
つぎに、ピリオド“.”から、前に来やすい句を順につなげて行くと、

①<sup>11.54</sup>←〈名詞+の〉←<sup>19.23</sup>〈名詞+に〉←<sup>16.67</sup>〈動詞+た〉←<sup>18.92</sup>.

という文型が得られる(矢印の上の数字は推移確率を示す百分比)。これが『寒  
 山拾得』の、文末句からたどった代表構文ということになる。

最後に、文末からの樹型の一部を示すとつぎのようになる。





上のようにすると、例えば〈名詞+デアル〉、〈名詞+デゴザイマス〉の前には、動詞や形容詞の連体修飾がかかる事が多いといったような、構文にかかわる問題が明らかになって来るのである。(実際には、〈名詞+である〉の場合が10例中8例、〈名詞+でございます〉の場合が10例中6例まで、動詞・形容詞・またはそれに助動詞の付いたもので終わっている)。

### III・展開

#### 3・0

以上述べた方法は、すでに実用段階にある。あとは入力データソースとしての各作品のファイルが完成すれば、それぞれの、および全体の処理が出来るのである。その結果についてはそれが出次第明らかにすることとし、以下で、その方法に手を加えて各種の研究に利用する構想について述べる。

#### 3・1

前節で表わした代表構文は、文節タイプ連続の統計結果を電子計算機で打ち出し、それを手作業で並べて作ったものである。この作業に、代表構文作成ルーチンを作って組み込めば、代表構文を自動的に打ち出すようになる。代表構文作成ルーチン自体は、文節タイプ連続の推移確率の高いものを順につなげて

行くだけで、それほど困難はないように思われる。ただ、この場合、つぎつぎにつなげて行くと途中でどもの所へもどってしまい、永久に繰り返すつづける（ループする）ケースが考えられる。<sup>(注12)</sup> 例えば前に「名詞＋ノ←名詞＋ガ」が現われていて、後の方で「動詞→名詞＋ノ」というものが来ると、また「名詞＋ノ→名詞＋ガ」がつながるというケースである。機械で代表構文を自動的に作るには；このループを防ぐために、一度用いた文節タイプは二度目には用いずに文節タイプ連続の度数がそれにつぐものを用いるとか、同じ文節タイプが二度現れたらそれに続く文節タイプを第二位のものにするとかの方法を講じておく必要がある。これらの方法は、一度用いた文節タイプを計算機の中に登録しておけばよい。

このようにすれば代表構文を自動的に作成できる。

(注12) このようなループ現象は、データ量の少ない場合に起る蓋然性が高く、機械化作業と関係がある訳ではない。しかし手作業の場合はループすることが解ってから対策を講ずればよいが、機械の場合、ループするのが解るのは、終わりのない文を打ち続ける時であるから、それからでは対策が間に合わないのである。

### 3・2

代表構文が出来れば、その自立語部分の品詞にあてはまりやすい単語をあてはめてみると一種の代表文が出来る。『寒山拾得』の語彙調査の結果（『電子計算機による国語研究Ⅴ』145ページ）を、前節の代表構文にあてはめれば、

①→名詞＋ハ→名詞＋ノ→名詞＋ニ→動詞＋テ→動詞→。

閻　は　事　の　僧　に　云　つ　て　ゐ　る　。  
 (1)           (2)           (4)           (1)           (2)

①←名詞＋ノ←名詞＋ニ←動詞＋タ←。

閻　の　事　に　云　つ　た　。  
 (1)           (2)           (1)

ということになる（カッコ内の数字は品詞内の順位）。この場合、各文節タイプごとにはいりやすい単語、例えば「名詞＋ハ」の場合の名詞にはどんな語が多いかを調べて、それをあてはめればもっと精度の高い代表文になるように思われる。そのためには、別に、原文ファイルから、文節タイプと自立語語形とを備えたファイルを作成してカウントをしなければならない。

### 3・3

今まで述べた方法および結果は、文節タイプ＝連続データを用いたものであった。この方法は二つの文節タイプ間の推移確率によっているのであるが、そのために、推移確率の高い所で変化が乏しくなってしまう。代表構文として示したものが3～5文節と短いのも、接続助詞を持つ文節タイプの度数が低いことに帰因するものと考えられる。

このような不自然さを少しでも解消するには三連続の文節タイプのデータを用いた処理を行なうのが一つの方法であろう。

三連続データをもとにした推移確率にはいろいろな方法が考えられるが、初期値をABCとすると、次はBCD、さらにCDEという具合に、前のデータの後二つと後のデータの前二つを揃えてつなげ、ABCDEという連続を得るやり方が一番妥当なものと思われる。

文節タイプ三連続データは二連続データ（ファイル4A、ファイル4B）から作る。二連続データから三連続データを作るには次のようにすればよい訳である。

AB	
BC	ABC
CD	BCD
DE	CDE
⋮	⋮

このやり方で〔例4〕（163ページ）のデータを処理すると次のようになる。

#### 〔例5〕

（通し番号）	（第1文節タイプ情報）	（第2文節タイプ情報）	（第3文節タイプ情報）	
1	①	<名詞+と>	<動詞>	E/i
2	<名詞+と>	<動詞>	<名詞+が>	E/i
3	<動詞>	<名詞+が>	<動詞+たそうである>	E/i
4	<名詞+が>	<動詞+たそうである>	<.>	E/i
5	<動詞+たそうである>	<.>	<①>	E/i

この三連続文節タイプデータの処理も二連続の場合と同じようにやればよいように思われるのであるが、第1～第3文節タイプ情報が合計120桁（40×3）あり、ソートのキー数の限界（88桁）を大幅に越えてしまうので、ソートをする事が出来ない。そして、現在まだ、三連続文節タイプデータの処理方法は開発段階にしかない。