

新編日本語品詞列集成について

田中穂積、荻野孝野、荻野綱男
(電総研) (計量計画研) (東大)

0. はじめに

筆者らは、「日本語品詞列集成」を作成した。この作業は、最初 西村忠彦氏によって昭和50年に行なわれたが、今回より改良を加え、新たに作成したものである。ここでは、「品詞列集成」の概要と改良作業等について述べよう。

1. 品詞列集成の概要

品詞列集成とは、品詞つき単語列を品詞でソートし、所定の長さの品詞列とその品詞列を含む単語列に前後文脈をつけて出力・収録したものである。

○ 原データ

品詞列集成の基になったデータは、昭和41年の朝日・毎日・読売新聞1年分の記事をサンプリング抽出した100万短単位のうち、広告文等を除く37万単語である。これは、国立国語研究所の「新聞の語彙調査」¹作業で作成された原文テープと辞書テープをお貸しいただき、この原テープに種々の加工を行なったものである。

○ 全体形式

一出力見本

今回の形式

文例 (25単語)

千品詞列 記事番号		前文脈(最大7単語)	ルートリ (3単語)	後文脈(最大15単語)
名助形容名	(K058207)	八千円の負担となる。#	国庫助成	の少ない私立大学では経費の増大を入試の受験料や学費をあけること
名助形容名	(K061942)	いる。#季節によって	出さかり	安い野菜を積極的に利用するよう、消費者教育をする必要がある
名助形容名	(K064950)	に、長年の苦難をしのばせる	ヒタイの	深いシワ——十日から東京ではじまつ”海外日系人第七回大会”
名助形容名	(K064950)	に住む日本語教師の発言は、	大会参会者の	の強い拍手を呼んだ。#G ₁
名助形容名	(K067247)	一、普通外米(準内地米以外の	粒の	長いタイ、ビルマ米など)は現在米屋で自由販売しているが、G ₁
名助形容名	(K071212)	季節により山菜を変えるなど	の細かい	ところを見せているが「ブドウの味のする甘いおもしろい」とみられ、国際的にも割高なG ₂
名助形容名	(K076849)	#今回ルビー通貨の切下げは	世界銀行の	強い要請で行なわれたものとみられ、要請で行なわれたものとみられ、国際的にも割高なG ₂
名助形容名	(K081129)	る今回から見て初めてつかめる	北海道の	すばらしい景観を見逃さないでほしいとするG ₂
名助形容名	(K084607)	最後かもしれない”といふ。	の切ない感情を、	メルクーリーは巧みにだしている。#温泉と湖水のG ₂
名助形容名	(K087907)	#断られたという。#ジエニイの	世界で真に「歓喜の歌」が歌えるのはいつのことか	G ₁
名助形容名	(J105063)	思う。#和平主義には、すでに	のない差別の	G ₁
名助形容名	(J108358)	#しかし二人のそばには、	美しい世界がある…。#『舞台さま、よしつけのお願いをお聞き	G ₂
名助形容名	(J119702)	大勢の	友だちがいて、若い同士らしく浮き浮きと話しているので、なか	G ₂
名助形容名	(J122278)	輸出物資へと倍以上。#	の少ないアラスカにとってこんな有難い話もない。	G ₁
名助形容名	(J123168)	#この平均寿命は世界で最も	ノルウェー、スウェーデン、デンマークの北欧三国の男七一歳、	G ₁
名助形容名	(J125160)	#一問わてて面上を上げた	#一問わてて面上を上げた彼女の	G ₁
名助形容名	(J136276)	内も正午前荒れ出し、台風週辺	の強い雨が降りはじめた。風速は午後零時二十九分最大瞬間二三・一メ	G ₂
名助形容名	(J138887)	#ーム静かなチームが競く	弱いファン層に送れる放送界で、浪曲番組をふやす傾向がある	G ₂
		かき根の	弱い住まいも、やはり問題であろう。#全国的視点からの海上警備再	G ₁

—図1— (左順篇)

ルートキー ①②③④ ④ ⑤ ⑥
主キー (品詞キー)

前回の形式

094948	動陳名格形容名動陳名格形容名	(A025811)	この	しょん	たの	心	サンフ
094949	連陳名格形容名付名動陳名格形容名	(A102484)	合併化	たなし	た市	の	。#
094951	名格形容名付名格形容名	(J165085)	たなのはの	に金回	まじめ	いい	し
094976	名格形容名付名格形容名	(C106935)	たなのはのドの	に金回	いい	いい	い
094979	付格形容名付格形容名	(C158582)	後	いた市	いい	いい	いい
094988	名格形容名動名名名名	(C184971)	後	えに木	まじめ	いい	い
094992	名格形容名動陳名格形容名	(A016177)	後	浦	くらんど	いい	い
095013	読名名格形容名名格形容名	(C161021)	後	葵	のぐひ	のくび	のくび
095206	数読名格形容名名格形容名	(J186891)	後	省	と都	と都	と都
095245	名格形容名経格名格動詞	(A008885)	後	多	が研究	の	の
095248	名格形容名経格名格動詞	(C187271)	後	数	と腐敗	の	の
095253	名格形容名経格名格動詞	(C114438)	後	久	派	の	の
095273	名格形容名経格名格動詞	(C100786)	後	政	政権	の	の
095280	名格形容名経格名格動詞	(C134163)	後	だつ	の	の	の
095285	陳付名格形容名付名格動詞	(A020947)	修正	た	た	た	た
095287	サ動名格形容名動名格動詞	(C184515)	連手	た	た	た	た
095327	動説名格形容名陳名動名付名	(A045967)	修	た	た	た	た
095330	句名格形容名付名付名付名付名	(A054228)	推進	た	た	た	た
095334	句名格形容名付名付名付名付名	(A055110)	あり	た	た	た	た

—図2—

出力形式の比較のために前回の形式も添付した。相違の詳細については4で述べる。なお、前回の出力見本は正順ソートのものであるが、これと全く逆に第1キーの位置を右側にして逆順ソートした右順索引である。述語をキーにして係り受け等を見ると時によい。

○ 規模

品詞列集成の規模を数の上からみてみよう。

表1 品詞列集成の規模の比較

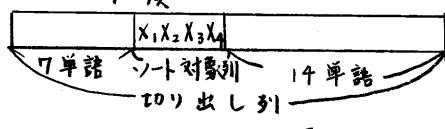
項目	今回の内容	前回の内容
原文の語数	366,687語	984,977語
原文の文数	14,576文	30,360文
印刷した行数(原本)	366,687行 ²⁾ " (普及本)	³⁾ なし
3品詞列の種類	2,980種	5,078種
2品詞列の種類	327種	427種

○ 品詞列集成の使い方

利用者は、調べたい長さ(1~3まで)の品詞列で探す。その品詞列の中に特定品詞(助詞・助動詞・連体詞)が含まれていればその特定品詞に応じる単語の語形で探すこともできる。例えば、図1を<名><助>,<名><助動>,<名><助の><形容>で探したりである。

2. 形式の詳細

○ ソート順



(ソートキー)

図1の①②③⑦ --- ソート対象列 $X_1 \sim X_4$ の品詞である(品詞キー)。

図1の④⑤⑥ --- $X_1 \sim X_3$ のいずれかが特定品

詞(助詞・助動詞・連体詞)の時のみの語形である(語形キー)。

ソートは上述の語形キーと品詞キーの組合せで行なわれる。キーの順序は①②③④⑤⑥⑦である。各キーのソート順は①②③⑦が

空名サナ付数ア連副接動用形陳助記説句井不

の品詞名順、④⑤⑥が国研コード順(ひらがな・カタカナのみのソート、それ以外は一つに扱う)である。

○ 単語の単位

国研の長単位に近い単位である。著者テーブルの1長単位内に含まれる短単位列を後に述べるルールでまとめたものである。便宜的にこの単位からなる列を単語列とよぶ。

	また	金融	引き締め	に	批判的	な	多く	の	銀行家	も、		図5のC
(本書の単位)												
(長単位)												図5のA
(短単位…												図5のB ₁ , B ₂ , …
前回の単位)												

○ 品詞分類

国研の品詞分類を基にキを加えたものである。

表2 品詞対応表

略品 詞	品 詞 名	元の品詞		語例
		單品詞	多品詞	
空	空品詞			
名	名詞	1, 2, 5, 8, 9, 1と5, 1と8, 2と8, 3, 1と9, 1とP		こと, もの, 東京都, 午前, ..., 訴訟, 汚職, ...
サ	サ変動詞語幹			
ナ	形容動詞語幹	4	1とCとDとL	同じ, 明らか, 必要, ...
付	付属語(接辞)	6		ごろ, 方, ら, 氏, ...
数	数詞	7		一, 一部, 一日, ...
ア	数字(算用・ローマ)	X		0, 1, 27日, ...
連	連体詞	D		この, その, 大きな, ...
副	副詞	C	1とC, 2とC, 9とC	いま, どう, 結局, ...
接	接続詞・感動詞	A, B	AとC	では, また, しかし, ...
動	動詞	E	DとE	いる, ある, し, ...
用	動詞類	+	1とE, 2とE	考え, オリ, 調べ, ...
形	形容詞	L, -	LとP	ない, よい, 多い, ...
陳	助動詞	P		た, だ, れ, ...
助	助詞	R		の, を, ので, には, ...
記	記号・符号	Y	読点, コンマ	※M〇,], ...
読	読点(コンマ)	Y	句点, ピリオド, 小数点	,
句	句点(小数点)	Y		.
#	文区切り			#
不	無関係・不明	Z		>, ここで, the, ...

(品詞の説明)

多品詞…辞書テープ中の短単位に二つ以上の品詞がついているもののうち判別文法による品詞の区分けをしなかったもの。表示の略品詞外になる可能性があるので、多品詞を含む品詞については注意が必要である。

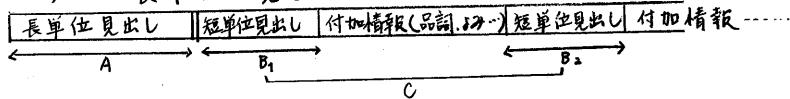
2品詞…2種のもの…品詞分類の数を減らすために一緒にしたものである。多品詞同様、着目の単語が実際は何品詞なのか念頭におかなければならぬ。

3. 作業手順

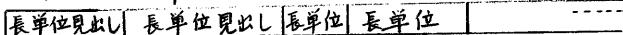
○ 原テープの形式

この作業は原テープの加工から始るので原テープの形式を明示しておこう。

辞書テープ…長単位を見出しとし短単位ごとの附加情報がついたもの



原文テープ…長単位切り込みに原文



一図5-

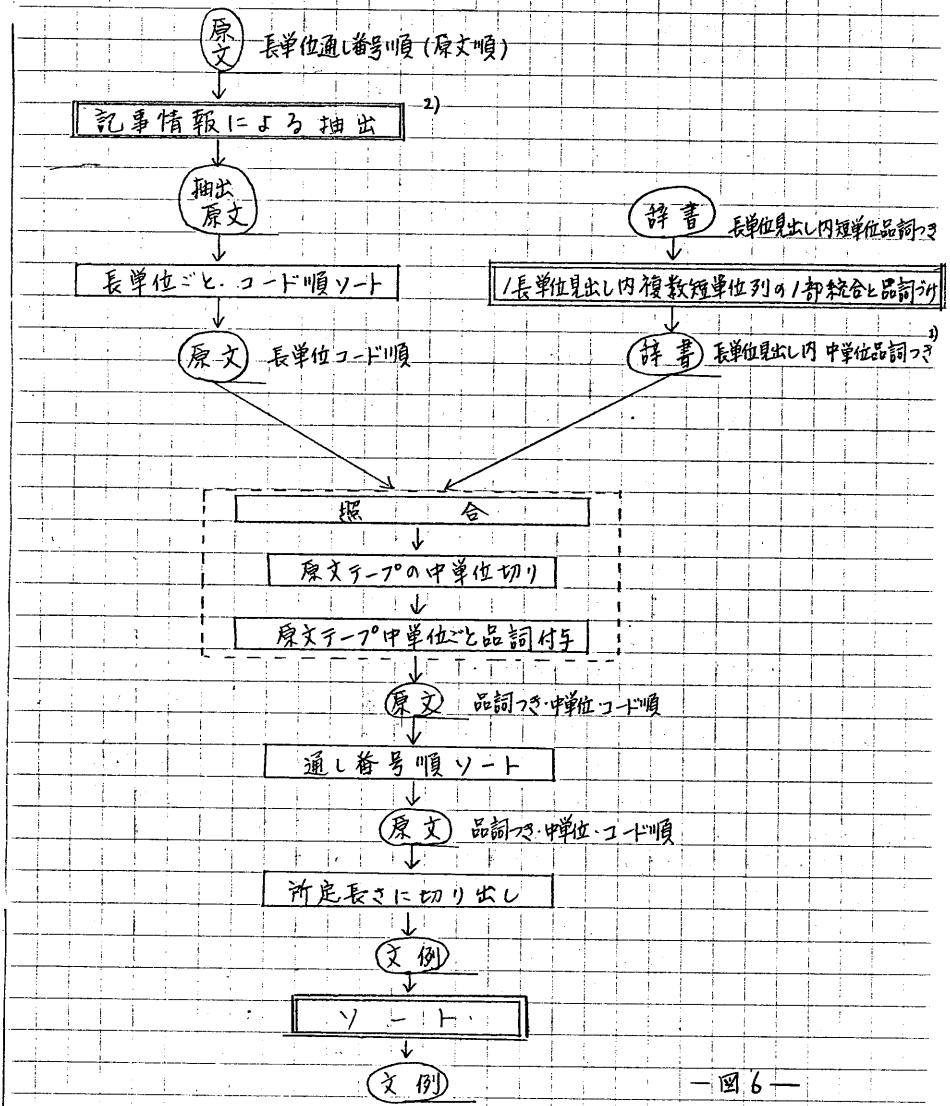
○ 作業概要

- (1) 上記の辞書テープを使って原文テープを辞書見出し右側の単位に切り直し、
その単位ごとの品詞を付与する。
- (2) この品詞付き原文テープを「切り出し列」として所定の長さに切る。
- (3) この品詞付き「切り出し列」を所定の位置のキーでソートする。
- (4) ソートされた品詞付き「切り出し列」の集合を全部あるいはサンプリング
して、所定の形式で出力する。

今日は、上記作業の前に以下の作業をし、原テープにキを入れた。

- ① 原辞書テープの書き換え —— 1長単位内(A)に含まれる短単位列(B₁, B₂)まとめてられるものは1単位(C)にする。
- ② 原文テープの記事情報によるデータの抽出 —— 広告文等を除く。

以上の作業内容を図示する。下図の 部分は、前回作業への追加・変更部分である。



④ 中単位とは
1長単位内で
まとめられる短単位
列からなる新しい
単位を指す。

4. 前回作業との相違点

この作業は前回の品詞列集成の改良版でもあるから、変更部分を明らかにしてみよう。

4.1 主な相違点

- (1) 番組・広告記事等の削除 —— 表6からもわかるように、前回のものは〈名詞〉の連続、〈数字〉の連続、〈名詞〉X〈記号〉の列等が上位を占め、構文的な資料としては不充分などころもあったので、対象データから番組・広告等を外した。
- (2) 長単位内短単位列の書き換之 —— 複合語の構成要素等はまとめて、単語単位と長単位に近いものにした。〈名〉X〈名〉X〈助〉は〈名〉X〈助〉となり見やすくなつた。表

1 の品詞列要素のリバーターンの減少、表 7 の <名詞><名詞>、<名詞><接辞> 等複合語構成要素列の減少となつた。

(3) ソート順の変更 —— 品詞だけではなく、助詞・助動詞等については単語の語形ソートを組合せにため、それらの用例や機能分析に有効である。

(4) 品詞グウの変更 —— 前回、助詞は <格>、<係・副>、<終・接> の三つに分かれていたが、助詞については語形ソートを行なうので品詞の細分はやめて <助>とした。

(5) 出力形式の変更 —— キーによる単語列に少しでも長く前後文脈をつけるよう工夫した。図 1、図 2 によって出力形式を比較してみるとわかるように、

① 切り出し番号の削除 ② ソート対象列外の品詞名削除 ③ ソート対象列外の単語間空白削除

等が主な変更点である。

○ 表による形式の比較

以上の変更点を表の上でみると形式的には次のような相違となる。

表 3 品詞列集成の形式の比較

項目	今回の内容	前回の内容
単位	国研の長単位に近い単位	短単位
品詞分類 (空品詞を含めない数)	19種	21種
ソートキー	3 単語列の品詞とその 3 単語 列内に含まれる特定品詞 (助詞・助動詞・連体詞) を 持つ語の語形	5 単語列の品詞
切り出し列の形式 1行の長さ	25 単語	12 単語
第 1 キーの位置 8 単語目	左(あるいは右)から 8 単語目	左(あるいは右)から 3 単語目
印刷文例の形式 1行の長さ	最大 54 字(うち空白 4 字, 最大 25 単語の文脈が入る)	最大 43 字(うち空白 11 字, 最大 12 単語の文脈が入る)
第 1 キーの位置 15 字目	左(あるいは右)から 15 字目	左(あるいは右)から 8 字目
サンプリング	ソートずみ切り出し列の集合 から通し番号順に $\frac{1}{5}$ 均等抽出	5 単語列の異なりごとに先頭 の 1 文例を抽出

4.2 長単位内短単位列の書き換え

4.1 に示す相違点のうち、(2) の長単位内短単位の書き換えについて少し詳しく述べてみよう。

切り出し列を長くとりたい、名詞の連続等は 1 品詞に書き換えてはどうか。これがこの作業の出発点であった。ただし、長単位間の品詞列も書き換えると、

----- この名林いすれし ----- のようなものもあるので書き換えは長単位内短単位列を

対象とした。図5 書き書きテープの形式 のBから複数個のものが書き換える対象となる。

○ 長単位内短単位列の異なりパターンと頻度

まず、対象となりるものの中のパターンを検討した。表4がパターン別頻度の一部で

KOSUU PATTERN あり、図7がパターン別例である。

表4

* ZISYO TAPE NO TANTANI NO PATTERN *

1	1111116	電話 (1) 土木 (1) 技術 (1) 者 (6)
6	111111	都 (1) 計画 (1) 検査 (1) 所長 (6)
51	11111	都 (1) 教組 (1) 委員会 (1) 長 (6)
2	11111	都市 (1) 交通 (1) 課 (1) 研究 (1) 所 (6)
32	11116	国連 (1) 中心 (1) 王義 (1) 的 (6) な (P)
1	11117	娘 (1) 芝居 (1) 騒動 (1) 記 (6) の (R) 卷 (1)
298	1111	フランダンブルグ (1) 協奏 (1) 曲 (1) 第 (6) 4 (X) 番 (6) ト (1) 長調 (1)
1	1111Y1	バイオリン (1) 協奏 (1) 曲 (1) 第 (6) 1 (X) 番 (6) 変 (6) ロ (1) 長調 (1)
1	11121	オルガン (1) 協奏 (1) 曲 (1) 第 (6) 4 (X) 番 (6)
8	1112	ピアノ (1) 協奏 (1) 曲 (1) 第 (6) 2 (X) 番 (6)
4	111611	化学 (1) 反応 (1) 速度 (1) 論 (6) (0116) 一 (Y)
16	11161	新人 (1) 会 (1) 创立 (1) 五 (7) 周年 (1) 記念 (1)
1	111666	ピアノ (1) 協奏 (1) 曲 (1) 一 (7) 番 (6)
9	11166	歌謡 (1) 曲 (1) ベスト (1) テン (7)
4	111676	勤続 (1) 年数 (1) 別 (1)
236	1116	
1	1116P	
1	1116R1	
1	1116X611	1 名詞・名詞 10549
1	1116X6611	2 名詞・接頭語 5484
2	1116X6	3 固有名詞・名詞 4189
1	1116Y	4 固有名詞・固有名詞 4066
1	111711	5 サ変語幹・動詞 2915
1	11176	:
1	1117	:
1887	111	:

図7 パターン別例

参考のために、上位の短単位列と書き出してみる。

○ 規則作成と適用

このパターン表と頻度と例を充分に検討した上で、書き換えるべき範囲を決り規則を作成した。書き換えるべき範囲は基本的に複合語レギュラのものとした。規則の一部が表5にある。規則数140、規則の一般形は $a_1 \dots a_m \rightarrow b_1 \dots b_n$ ($2 \leq m \leq 5$, $1 \leq n \leq 2$) である。左の規則を所定の順に適用して書き換えて行った結果は表6 図8に示す通りである。

* TANTANI NO KECUGOO NO YOOSU (OGINO) *		
711 1 0010	1111116	1 曲芸 (1) 師 (6)
1167 1 0021	111111	曲射 (1) 痘 (6)
116XX 1 0022	111161	曲 (1) 独楽 (6)
11XX 1 0023	11116	計画 (1) 素 (6)
11X 1 0024	11117	計画 (1) 性 (6)
117 1 0025	1111Y1	計画 (1) 局 (6) 長 (6)
11 1 0026	11121	計画 (1) 通 (6) だっ (P)
22 1 0030	1112	計画的 (6) な (P)
66 1 0040	111611	計画 (1) 内容 (1)
77 7 0050	1116	計画 (1) 部 (1)
88 8 0060	11117	計画 (1) 作成 (1)
CC C 0061	1111	計画 (1) 者 (6)
EE E 0070	1111Y1	立案 (1)
RR R 0080	11121	3さ (E)
XX X 0090	1112	3し (E)
YY Y 0100	111611	どおり (6)
81176 1 0101	11161	造船 (1)
6XX11 8 0102	111666	肝腎 (1)
6XX16 8 0103	11166	違 (2)
7Y771 1 0104	111676	機 (6) 数学 (1) 夜話 (2)
1776 7 0105	1116	
1X61 1 0106	1116P	
1X66 1 0107	4P	
616E E 0108	1116R1	
616P 4P 0109	1R1	
6X11 8 0110	1116X611	
6X16 8 0111	1116X6	
7161 1 0112	1116Y	
7171 1 0120	111711	
762E E 0121	1117	
8YXX X 0122	111	
X1X1 1 0123	111E1	
X6Y1 1 0124	111E	
12E E 0125	111R1	
16E E 0126	111R8	
16P 4P 0127	111R	
表5 規則	111X1	1

図8 書き換え例

表6 パターン別書き換えの結果

5 出力結果の数量的検討

4で述べた作業の相違が出力結果にどのように影響してかかるか、数字の上でみよう。

○ 頻度表による比較

(品詞列) -表 6-

	前回		今回			
	品詞	度数	割合	品詞	度数	割合
1	名詞	302603	30.7	助詞	99625	27.2
2	数字	141659	14.4	名詞	98015	26.7
3	格助詞	99302	10.1	動詞	40679	11.1
4	記号	85798	8.7	助動詞	23959	6.5
5	接辞	71483	7.3	読点	22772	6.2
6	動詞	51167	5.2	区切り	16653	4.5
7	数詞	42415	4.3	記号	14574	4.0
8	区切り	35123	3.6	句点	11683	3.2
9	読点	30083	3.1	数詞	11620	3.2
10	助動詞	29704	3.0	副詞	5651	1.5
11	係助詞	22584	2.3	形容詞	4169	1.1

数字・記号・接辞が減少している。これは作業相違(1)広告文等の削除(2)短単位列の書き換元による。助詞が1位になり、(1)品詞づけの変更により係助詞・格助詞・終助詞を一つにまとめたことと、(2)により〈名詞〉の連続がまとめてから〈名詞〉の頻度が減ったためと思われる。

(2 品詞列) -表 7-

	前回		今回			
	品詞列	度数	割合	品詞列	度数	割合
1	名詞・名詞	104472	10.6	名詞・助詞	66085	18.0
2	数字・数字	85640	8.7	助詞・名詞	35969	9.8
3	名詞・格助詞	65079	6.6	助詞・動詞	34569	9.4
4	格助詞・名詞	39706	4.0	読点・名詞	15847	4.3
5	記号・名詞	37722	3.8	動詞・助動詞	13939	3.8
6	名詞・接辞	36529	3.7	句点・区切り	12735	3.5
7	名詞・記号	31718	3.2	助詞・読点	10872	3.0
8	格助詞・動詞	26649	2.7	区切り・名詞	9980	2.7
9	記号・数字	24698	2.5	名詞・名詞	8688	2.4
10	接辞・名詞	24251	2.5	記号・名詞	8263	2.3

前回の〈名詞〉<名詞>、〈名詞><接辞>、〈数字><数字>、〈記号><名詞>等上位に入っているものが(2)短単位列の書き換元によりまとめてから、今回の上位列は日本語の構文的並びを示しているようである。

(3 品詞列) -表 8-

	前回		今回			
	品詞列	度数	割合	品詞列	度数	割合
1	数字・数字・数字	45746	4.6	名詞・助詞・名詞	27334	7.5
2	名詞・名詞・名詞	43672	4.4	助詞・名詞・助詞	26452	7.2
3	名詞・格助詞・名詞	27869	2.8	名詞・助詞・動詞	21950	6.0
4	名詞・格助詞・動詞	18670	1.9	助詞・動詞・助動詞	11604	3.2
5	数字・数字・記号	17235	1.7	読点・名詞・助詞	11088	3.0
6	名詞・記号・名詞	17210	1.7	助詞・動詞・助詞	10919	3.0
7	記号・数字・数字	16482	1.7	助詞・読点・名詞	6316	1.7
8	格助詞・名詞・格助詞	16126	1.6	区切り・名詞・助詞	5919	1.6
9	名詞・格助詞・格助詞	15647	1.6	動詞・助詞・動詞	5639	1.5
10	数字・記号・数字	15194	1.5	句点・区切り・名詞	5355	1.5

2品詞列でみた場合と同じことか言える。

6 利用例

○ 頻度表による日本語文作成

3品詞列の頻度順表を使って、日本語の品詞の連続の典型を作つてみよう。文区切りから出発して、その後に何がえやすいかを順にみていく。ここでは一番頻度の高いものをとつて、かつ一度とつたパターンの場合には次の頻度のものとつけていく。

(左順) 3品詞列頻度表より) 井名助名動陳句井

(右順) 3品詞列頻度表より) 井名助詭名助名動陳句井

もつともう少し日本語文の品詞の連続ができあがつた。このような検討を繰り返していくつかの文パターンを作ることも可能であろう。

左順 3品詞列頻度表 表9

②名助名	27334	③助名助	26452	④名動助	21950	⑤助動陳	11604	或名助	11088	助動助	10919
助詭名	6316	①#名助	5919	動助動	5639	句#名	5355	助動名	5297	⑦陳句#	5209
名詭名	4822	動名助	4513	陳名助	4423	名名助	4332	名助詭	4228	⑥動陳句	3571

右順 3品詞列頻度表 表10

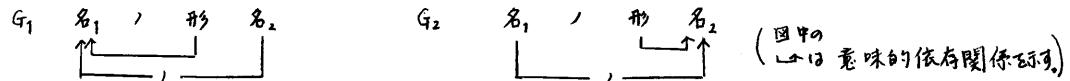
④名助名	27334	⑤助名助	26452	⑥名動助	21950	③助動陳	11604	⑦詭名助	11088	助動助	10919
③助詭名	6316	①#名助	5919	動助動	5639	句#名	5355	助動名	5297	①陳句#	5209
名詭名	4822	動名助	4513	陳名助	4423	名名助	4332	名助詭	4228	②動陳句	3571
動陳名	3362	動陳助	3355	動句#	3258	助動句	3146	記名配	2918	名配名	2813

○ 「名詞関係依存の原則」考察

田中らが先に発表した⁸⁾上記原則とこの「品詞列集成」を使って考察してみた⁹⁾。これは、〈名₁×助₁の〉×形×〈名₂〉の二つの〈名〉の意味的依存関係から〈形〉の修飾先を決定するというものである。

(1) 〈名₂〉/〈名₁〉が成立すれば、〈形〉は前の〈名₁〉を修飾する。—G₁型—

(2) 〈名₁〉/〈名₂〉が成立すれば、〈形〉は後の〈名₂〉を修飾する。—G₂型—



上記原則について、「品詞列集成」を使って調べたところ反例なしで一般性が確認された。図1に示すように、〈名×助×形〉の部分から〈名〉<助の><形>〈名〉の形を探し1例ずつ検討した。考察結果の詳細と文例は文献9)にもある。

7 問題点

○ 品詞づけ

原テープあるいは原テープ加工の段階で品詞づけの誤りが生じている。

例えば	誤り品詞	正しい品詞
すい	名詞	→ 助動詞
この	名詞	→ 連体詞
なら	動詞	→ 動詞か助動詞

等である。これらは、前回の品詞列集成から見つけたもので原テープの誤りと思われる。今回作業以前に見つかりたもので頻度の多いものについては、部分的な修正を行なったが、その後見つかりたもので

では 接続詞 → 助詞 か 助動詞+助詞

等は未修正である。

また、長単位内短単位列書き換えて、助詞を含まない文相当の短単位列、

潜水艦 潜航せず
1 3 E P
1

のようなものに短単位列書き換え規則が適用され、誤ってまじめられ品詞づけされたものもある。

以上のようない内容で、誤り品詞からなる単語列も含まれているので、利用者は注意してほしい。

最後に、原データを御提供下さった国立国語研究所、並びに前回作業を手がけられた西村忠彦氏（東京農工大）、今回作業を御援助下さった洲一博バターン情報部長に謝意を表する。

文献

- 1) 国立国語研究所：“電子計算機による新聞の語彙調査”，国研報告 37，1970
- 2) 3) 田中穂積、萩野秀野、萩野綱男：“新編日本語全品詞列集成 左順 右順”，電総研推論機構研究室，1979
- 4) 5) 同上：“新編日本語品詞列集成 左順 右順”，同上，1979
- 6) 7) 西村忠彦、萩野綱男：“日本語品詞列集成 左順 右順”，電総研，1977
- 8) 田中穂積、萩野秀野：“形容詞もしくは形容動詞の連体句と含む名詞句の意味処理”，情報処理学会大会予稿，昭和54年
- 9) 同上：“形容詞もしくは形容動詞の修飾先の名詞を決める原則について”，文部省特定研究報告，昭和54年