On the verbal concordance to analyze the usages of a word in Japanese literature

国語研究のための索引作成システム

——新システムの基本思想を中心として——

靍岡昭夫

1. まえがき

これまでに，世の中には多くの総索引が作られている。その大半は，日本古典などの文学作品の語をカードに書き写して並びかえるという，手作業によって作られたものである。しかしこの十年来，電子計算機を利用して索引を作る方法が開発され発達して来ているので，われわれ国立国語研究所では，これを改良して言語研究に便利な索引や検索用言語データを作ることを考えたのである。そしていくつかのシステムが作られてきたが，今回，田中章夫，土屋信一，中野洋と本稿筆者の四人の共同研究により新しい用語索引作成システムが完成した。そこで，この索引システムの内容を報告するとともに，ここに至るまでの経過と今後の見通しについて述べることにした。

2. これまでの索引システム

○国立国語研究所での索引作成システム

前節で書いたように，国立国語研究所では，これまでにさまざまな索引システムとそれによる総索引を作成してきた。われわれは，言語研究のためには，文脈付き索引（検索すべき語＝key word が文脈中＝in contex にあるので，KWIC索引と呼ぶ）が便利であると考え，この方式を持つ索引を多く作ってきた。今回新たに作成したものを含め，現在までの索引作成システムは，入出力の方法という点で，つぎのように分類される。

①漢字テレタイプ入力（漢字かな）→漢字テレタイプ出力（漢字かな）(注1)

②フレキソ入力（ローマ字）→ラインプリンタ出力（ローマ字）(注2)

③フレキソ入力（片かな）→ラインプリンタ出力（片かな）(注3)

④漢字テレタイプ入力（漢字かな）→ラインプリンタ出力（片かな）(注4)

⑤漢字テレタイプ入力（漢字かな）→高速漢字プリンタ（漢字かな）・ラインプリンタ出力（片かな・ローマ字）

電子計算機を利用しての索引作りは，まず①から始まった。日本語は漢字かな交り文が一般的であり，それを処理することを最初に考えたわけである。そして，このシステムを使って芥川龍之介の「くもの糸」の索引が作られた。しかし，このシステムは，出力機である漢字テレタイプ印字機の印字速度が遅い（120 字／分）ので，短編作品の索引は作れても，中編・長編のものとなると非常に時間がかかって実用的でない。

そこで，出力を高速（90000～100000 字／分）のラインプリンタ（電子計算機の印字機）による②③④のシステムが作られたのである。②③では入力機にフレキソ（英数字・片かなタイプライタに紙テープさん孔機の付いたもの）を用い，片かなで入力すれば片かな索引に（③），ローマ字で入力すればローマ字索引（②）になるようになっている。また，④のシステムは，漢字（よみがな付き）とかなで漢字テレタイプによって入力し，ラインプリンタにより片かな索引を作るものである。これは，漢字かなで入力してあるので，漢字テレタイプで印字することもできるが，前に述べた理由からほとんど行われていない。③のシステムでは，主として日本古典の『遊子方言』『浮世風呂』『浮世床』『心中天の網島』『今昔物語集26巻・30巻』『当世書生気質』などの索引が作られた。また，④のシステムでは，昭和41年度から昭和48年度にかけて行われた「現代新聞の語彙調査」の長単位カナ KWIC が出力されたほか，昭和 48 年度から行われている「漱石・鷗外の用語研究」のために，『三四郎』『行人』『硝子戸の中』『夢十夜』『高瀬舟』『青年』の KWIC 索引も作られた。②のシステムは入出力がローマ字ということから日本語への応用という点で利用度が低いが，現在『夢十夜』ローマ字索引が出来ているほか『星の王子さま』など外国語の索引がいくつか作られている。

②③④のシステムではラインプリンタによる出力方式をとったために出力時間は短くなった。しかし，②はローマ字文しか適用できないし，③④は片かな出力のため読みにくく，同形見出しの識別もやっかいである。そこで，最近開発された高速漢字プリンタ（50000～100000字／分）を用いて漢字かなまじり文を高速出力するシステムの開発にかかったのである。

高速漢字プリンタを出力に利用するだけならば①や④のシステムの出力部分を変えればよい。しかし，どうせ言語データを作るならば，索引作りばかりでなく，さまざまな言語研究に使えるデータを作っておこうということと，さまざまな形のデータを扱える汎用性のある処理システムにしておこうということになって，⑤の新システムの開発が始められた。このシステムにおける，言語データ作成のための前処理（プレエディット）については本稿（次節）で，またそのデータを処理するプログラムシステムについては中野洋論文（本集16ページ以下）において書かれている。

このシステムによって作成された（作成されつつある）KWIC索引をはじめとした言語データはつぎのとおりである（語数は概数。*印は第一次出力—mini KWIC による校正用出力，**印は処理途中の未完成品）。

森　鷗外『寒山拾得』（4100語）
〃　『雁』（50000語）
〃　『山椒大夫』*（16000語）
〃　『渋江抽斎』**（150000語）
夏目漱石『草枕』*（58000語）
〃　『坊っちゃん』*（55000語）
〃　『こころ』**（150000語）

なお，昭和49年度から開始された「高校教科書の用語調査」のシステムもこのシステムを基礎としたものである（教科書調査のシステムについては別な機会に発表が行われる予定）。

○国立国語研究所外の索引作成システム

電子計算機を用いて索引を作る実験や作業は現在では非常に多くのところで行われている。それらの多くは特徴のある興味深いものであるが，言語研究を目的とした索引はそれほど多くはないように思われる。まとまった作品の索引では，『平家物語総索引』(金田一春彦・清水功・近藤正美編・1973. 学習研究社刊) があるが，これは従来の手作業方式の一部を電子計算機 (HITAC 10) にやらせるようにしたものである。そのほか植村俊亮「漢字かなまじり文KWIC」(『情報処理 10-5』1969年) で発表されたシステムおよびそれによる新聞長単位60万語のKWIC索引がある。また，KWIC索引を利用して分析した結果が「放送用語のKWIC索引」(管野謙『文研月報』昭和46年6月) にあげられている。

(注 1) 斎藤秀紀作成。「電子計算機と漢テレによる用語総索引の作成」(『電子計算機による国語研究』1968)
(注 2) 斎藤秀紀作成。および江川清作成。
(注 3) 土屋信一作成。「カナ入力による日本語文総索引の作成」(『電子計算機による国語研究Ⅳ』1972)
(注 4) 石綿敏雄作成。「新聞用語調査の用例印字プログラム"COBOL－KWIC"」(『電子計算機による国語研究Ⅲ』1971)

3. 新しい索引作成システム

3·1 言語研究の方向

電子計算機は，入力するデータとそれを処理するシステムとがしっかりしていさえすれば，われわれの必要とする形式・内容をもった言語データ (索引を含めて) を，忠実に作ってくれる。したがって，われわれが言語データを作るには，どのような言語研究をするか，そのためにどのような形式・内容をもった言語データにするか，そしてそれにはどんな情報を付加する必要があるか，ということを検討することが必要になる。

今われわれが考えている研究は，音素・文字・語 (自立語・付属語・語構成要素など，さまざまなレベルがある)・文節・句・節・文・文章といったいろ

いろな言語単位について，用法（出現環境・出現率など）を考察して行く計画である。

文字の出現環境（例えば連続確率など）や出現率などは，一文字が一単位となっている原文データの磁気テープ（MT）ファイルがあればそれをもとにして電子計算機で自動的に調査することができる。音素については漢字かなまじり文においては漢字にふりがなを付けておけば，それをもとにして自動的に全文かなのデータが得られるから，あとは文字の場合と同じように調査をすることができる。語や文節の研究には，文章から語，文節の単位で用例が得られるように単位切りをする必要があるが，これについては次項で述べる。文や文章の研究は，音素・文字・語・文節などをもとにしたやり方，本文から直接データ（文の種類わけなど）を得るやり方といった，さまざまな調査方法を考えている。(注5)

以上の研究は大きくわけると，MTファイルをもとにして電子計算機で検索・カウントなどを処理して行うものと，電子計算機で編集し，漢字プリンタでプリントアウトしたKWIC索引によるものとになる。そして従来，語彙調査は前者を中心に，索引作りは後者を中心に行われていたように感じられる。本システムは，どちらにも使える，というよりも文章を材料にした総合的言語研究ができるように開発されたものである。

(注5) これまで，このシステムで作られたデータを用いた研究は次のものが発表になっている。
○靍岡昭夫「電子計算機による代表構文作成の試み」(『ことばの研究5』1974)
○靍岡昭夫「文節タイプ連続の研究」(『電子計算機による国語研究Ⅶ』1974)
○米田正人「"文の長さ"の統計学的一考察」『電子計算機による国語研究 Ⅶ』1974)

3·2 単位切り

今までの索引は，ほとんど一種類の単位を用いて作られている。カードに転

写する手作業方式では単位の種類をふやせば，それだけ作業が増大するから，複数の単位を用いることはまずなかった。電子計算機を用いた索引作りも，ほとんど単一の単位で行われてきた。国立国語研究所で作られた索引の単位はこれまでに行われた用語調査で用いられた単位がそのまま用いられることが多かった。それらは次のようなものである。

- α単位（国立国語研究所報告4『婦人雑誌の用語』1953. 19ページ以降参照）
- β単位（国立国語研究所報告21『現代雑誌九十種の用語用字』1962. 6ページ以降参照）
- 長単位（国立国語研究所報告37『電子計算機による新聞の語彙調査』1970. 13ページ以降参照）
- 短単位（国立国語研究所報告37. 15ページ以降参照）

このうち，たとえば前節で示した作品のうち，『高瀬舟』『青年』などは短単位で単位切りがなされているが，それは次のようになっている（単位の切れ目は／で示す）。

／高瀬舟／は／京都／の／高瀬川／を／上下／する／小舟／で／ある／。／徳川／時代／に／京都／の／罪人／が／遠島／を／申し／渡さ／れる／と／……（『高瀬舟』より）

／純一／は／立ち留まつ／て／名前／を／読ん／で／見／た／。／自分／の／捜す／大石／絹太郎／と／いふ／名／は／上／から／二三／人／目／に／書い／て／ある／ので，／すぐ／に／見附かつ／た／。／…（『青年』より）

しかし，今までの単位は，もともと語数カウントを中心とする語彙調査のために考えられたものであるので，ことばを探す場合には都合の悪いことが少なくない。たとえば，上にあげた『高瀬舟』や『青年』の索引からは，「高瀬舟」の「舟（ブネ，またはフネ）」，「高瀬川」の「川（ガワ，またはカワ）」や「立ちま留つた」の「留まつ（ドマツ，またはトマル）」などの語を検索することが困難である。

また，本研究は前項で述べたとおり，さまざまな言語単位について研究をす

る計画なので，一種類の単位では不足である。

そこで，今回開発した新システムでは，C単位，L単位，S単位の三種類の単位を用いることができるようになっている。

〔C単位〕

C単位は，文節に相当するもの，および記号（連続した記号は全体で一単位）である。文節の認定は，大すじ次のように定めた（本稿では切れ目を／で表す）。

(1)スペース（一字分以上の空白）で切る。

(2)助詞・助動詞のあとで切る。助詞助動詞は『現代語の助詞・助動詞』（国立国語研究所報告3）によるほか，「じゃ」「にゃ」「ちまう」「ちゃう」「てる」「とる」や，形容詞型活用語・形容動詞型活用語の連用形の直後の「ある」「ござる」を加える。「—の～」の形の体言句で，新潮国語辞典で一語扱いにされているもの（「板の間」「菜の花」など）は，「の」の後を切らない。

(3)助詞・助動詞を伴わないで，主語，連用修飾語，連体修飾語，接続語，独立語となっている自立語は一単位。体言の連体修飾はみとめないが，時間，場所，肩書きなどで，下位・上位のものの並列は切り離す（／5月／1日／正午に／　／東京都／北区／西ケ丘の／）。

〔L単位〕

L単位は，C単位（文節および記号連続）を，次のように切り離したものである（本稿では切れ目をIで表す）。

(1)記号は一個につき一L単位。

(2)助詞・助動詞（それらが連続している場合はその先頭のもの）の前で切る。助詞・助動詞はC単位の(2)で述べたとおりである。

(3)副詞語尾・形容動詞語尾は語幹から切り離さない（／はっきりと／　／すぐに／　／静かだ／　／立派です／　cf. ／友達Iと／　／人間Iだ／　／うさぎIです／）。

〔S単位〕

S単位は，L単位を，現代語で意味を担う最小の言語単位（最小単位）(注6)の一～二回結合によって分割したものである。その切り方のあらましは次のようになっている（本稿では切れ目を／で表す）。

(1)助詞・助動詞は一語一単位とする。C単位で切られなかった「菜の花」「板の間」などの「の」も一S単位にする。しかし，「この」「その」「どの」などの「の」は切り離さずに全体で一単位とする。

(2)和語の自立語は一要素を一単位とする。ただし，名詞・副詞・形容動詞語幹は二要素結合までを一単位とする(注7)。（｜走り／こむ｜　｜雨あがり｜）

(3)漢語(注8)の自立語は，二要素(漢字二字）の結合までを一単位とする。（｜社会／主義｜　｜駐在／員｜）

(4)外来語の自立語は，一要素（原語で一単語）を一単位とする。ただし，「ネクタイ」「クーデター」など，日本語で一単語と考えられ，分割できないものは切らない。（｜ウォーミング／アップ｜　｜アンダー／ショット｜）

(5)和語要素と漢語要素の混種語である自立語は，(3)と同様に二要素結合を一単位とする（｜重箱｜　｜場所｜）。それ以外の混種語は種によって切り離し，それぞれの中を(2)～(4)にしたがって処理する。ただし，「デモる」「タクる」などの活用語尾は切り離さない。（｜ナイロン／ザイル｜　｜水割り／ウィスキー｜）

(6)動詞型活用の接辞（「…がる」「…めく」など）や，形容詞型活用の接辞（「…がましい」「…っぽい」など）は一要素を一単位とする。また，形容詞・形容動詞の語幹に付く「さ」「み」「げ」も一単位とする。名詞性接辞は一要素として，(2)～(5)の中で処理されるので，外来語，および用言の中では一単位（｜プレ／オリンピック｜　｜お／美しい｜）となるが，体言の中では一次結合は切らない（｜お足｜　｜お体｜　cf.　｜お／父さん｜）。

(7)サ変動詞「する・ずる」は一S単位とする（｜びっくり／する｜　｜心配／する｜）。同じ位置に来る「できる」「いたす」なども同様に処置する。ただし，一字漢字に付いた「す・する・ずる・じる」などは切らない（｜愛する

I　I信じるI)。

(8)形容動詞語尾は語幹から切り離す。(I静か／なI　I立派／なI)

(9)副詞語尾「に」「と」は前と切り離さない(I割にI　Iじっとl)。ただし,「に」「と」を切り離して単独でも副詞として用いられるもの,三音節以上の擬音語・擬態語などに付く「に」「と」は切り離す(I割合／にI　Iすぐ／にI　Iぐらり／とI)。

(10)同音・類音の反復形である副詞・擬音語・擬態語は,反復部分が三音節以上の場合切り離す。(Iのっし／のっしI　Iのらり／くらりI)

(11)独立して用いられないものを含むもの,切るべき位置のわからないもの(略称も含む)などは,二要素以上でも一S単位とする。(IけだものI　I都区内I　I有頂天I　IPTAI　Iペ平連I)

(12)固有名詞(人名・地名)は一要素を一単位とする。(I森／鷗外I　I東京／都I)

(13)数(算用数字・漢数字・ローマ字のほか「幾度」「何人」の「幾」「何」などを含むは)一字一単位とする(I一／回I　I百／人I)。ただし,一,二,……十……というように数え進むことのできないもの,位取りを表す十,百,千,万,億……などは前と切り離さない(I一旦I　Iふたりl　I五十万／円I)。

単位切りの作業は,本文のC単位の切れ目に黒鉛筆で／(スラッシュ)を入れて行き,それが済んだもののL単位の切れ目に青鉛筆で／を入れ,最後にS単位の切れ目に赤鉛筆で／を入れるという,三段階方式をとって行われる。なお,複合してできた接続詞(例えば「それに反して」「そのほか」など)や,連体詞(例えば「こういった」「そうした」など)のようなものはその認定が人によって,また時によってゆれるおそれがあるので,今までに述べたCLS単位で切れるものは切っておき,あとから連語コードを付けていくようにした。例えば,「それに反して」の場合は,／それI に／反しI て／と切っておいて,それぞれに接続詞の連語コードを付けておくのである(連語コードにつ

いては次項でのべる)。連語が多く集まれば，それを全部電子計算機に登録して自動的に連語コードを付けることも可能になる。

以上のような規則で処理したものは次のようになる。

／唐｜の／貞観｜の／頃｜だ／と／言ふ｜から／，／西洋｜は／七／世紀｜の／初／日本｜は／年号｜と／云ふ／もの｜の／やつと／出来／掛かつ｜た／時｜で／ある／。／……（『寒山拾得』より）

／親譲り｜の／無／鉄砲／で／小供｜の／時｜から／損｜ばかり／し｜て／居る／。／小学／校｜に／居る／時分／学校｜の／二／階｜から／飛び／降り｜て／一／週間｜程／腰｜を／抜かし｜た／事｜が／ある／。／……（『坊っちゃん』より）

上の単位切りで，／はC単位の切れ目であると同時にL単位，S単位の切れ目でもあり，｜はL単位，S単位の切れ目でもある。したがって／をCLS，｜をLS，／をSと表わすと『寒山拾得』と『坊っちゃん』は，

〔寒山拾得〕	C	L	S
CLS唐	]	]	]
LSの	]	]	]
CLS貞観	]	]	]
LSの	]	]	]
CLS頃	]	]	]
LSだ	]	]	]
Sと	]	]	]
CLS言ふ	]	]	]
LSから	]	]	]

〔坊っちゃん〕	C	L	S
CLS親譲り	]	]	]
LSの	]	]	]
CLS無	]	]	]
S鉄砲	]	]	]
Sで	]	]	]
CLS小供	]	]	]
LSの	]	]	]
CLS時	]	]	]
LSから	]	]	]

ということになる。Cから次のCの前までがC単位，Lから次のLの前までがL単位ということになる。

(注 6)　国立国語研究所報告21参照

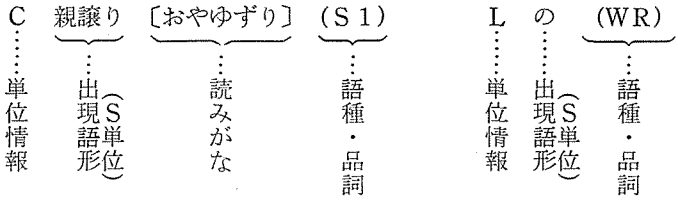
(注 7)　体言類を二要素の結合を一単位としたのは，「屋根」「場合」「割合」「仕

事」など，二要素の結合全体で一語のように使われるものが少なくないという理由からである。しかし，そのために「食べすぎる」は「食べ／すぎる」と切れるのに「食べすぎ」は切れないので「食べすぎ」から「すぎ」が検索できないという問題がある。

(注 8)　本稿にいう漢語，外来語には，それぞれ和製漢語，和製英語などを含める。

3·3　付加情報

このシステムは，多種の情報を処理する能力が備えられている。それらは，つぎのようになっている。



単位情報は，C，L，S単位の先頭，すなわち黒い／の直後の語にはCと書き，L，S単位の先頭，すなわち青い／の直後の語にはLと書き，S単位の先頭すなわち赤い／の直後の語にはSと書いておく。この情報をもとに，C単位のデータ，L単位のデータ，S単位のデータが自動的に出来る。

出現語形のあとに，読み仮名が必要な場合は〔　〕の中にひらがなで書き入れる(注9)。そしてすべての出現語形のあと（　）内に語種，品詞の情報を書き入れ，またその品詞が動詞や動詞型活用の接辞（「がる」「めく」など）である場合は活用型，活用行の情報も付け，複数のS単位が複合して他の品詞となっているものには，それぞれに連語コードも付ける（連語コードの付け方は14ページ参照)。付加情報は，表1に示したようなものがある(注10)。

単位切りの済んだ原文に書き込む空白はほとんどないから，単位切り作業と情報付加作業の間に原稿用紙に清書する作業がはいる。清書は出現語形一S単位ごとに一行ずつ書き，13ページの例のような形で情報を付ける。先に単位切りの例としてあげた『寒山拾得』と『坊っちゃん』の漢字テレタイプさん孔

表 1　付加情報コード表

1ケタ目	2　ケ　タ　目	3ケタ目	4ケタ目	3あるいは5ケタ目
語種コード	品 詞 コ ー ド	活用コード	(活用形)(活用行)	連語コード
S和語	1純名詞 2(空番)	F四段・五段活用	ワわあ行	ササ変動詞の一部
T漢語	3サ変動詞語幹 4形容動詞語幹(派生形を含む) 5形容詞語幹(派生形を含む)	G上一段活用	ああ行	ケ形容動詞の一部
U外来語	6名詞性接辞，助数詞 7数詞	H上二段活用	かか行	コ固有名詞の一部
V混種語	8固有名詞 9代名詞 A接続詞	I下一段活用	がが行	セ接続詞の一部
W語種不要	B感動詞 C副詞	J下二段活用	ささ行	カ感動詞の一部
X数字	D連体詞 E動詞	K変格活用	ざざ行	フ副詞の一部
Y記号	＋動詞性接辞 －形容詞性接辞 M口語形容詞	V融合形，変則活用	たた行	レ連体詞の一部
Z語種不明	N文語形容詞 P助動詞，形容動詞語尾 Q(空番)	Z活用不明	だだ行	ド動詞の一部
%情報無視	R助詞 X算用数字，ローマ数字 Y記号 Z品詞不明 %情報無視		なな行 はは行 ばば行 ぱぱ行 まま行 やや行 らら行 わわ行 ん3ケタ目VZ	ヨ形容詞の一部

注　○すべての語に，語種コード・品詞コードがつく。活用コードがつくのは動詞と動詞性接辞だけである。連語コードは必要な時に，人間又は機械によってつけられる。

○語種コードのW（語種不要）は，品詞が，助詞・助動詞・固有名詞である場合に付ける。

○品詞コードの4，5に語幹とあるが，もともとは活用語の語幹であるが転成名詞の一部となっているものも含む。

用原稿はつぎのようになっている（実際にはこの他にページ情報，行情報，段落先頭情報，本文外の語の情報といったシフトコードが書かれるが，これについては中野論文25ページ参照）。

『寒山拾得』

C唐〔とう〕（T1）
Lの（WR）
C貞観〔じょうがん〕（T1）
Lの（WR）
C頃〔ころ〕（S1）
Lだ（WP）
Sと（WR）
C言ふ〔いふ〕（SEFは）
Lから（WR）
C，（YY）
C西洋〔せいよう〕（T1）
Lは（WR）
C七〔なな〕（X7）
S世紀〔せいき〕（T1）
Lの（WR）
C初〔はじめ〕（S1）
C日本〔にっぽん〕（W8）
Lは（WR）
C年号〔ねんごう〕（T1）
Lと（WR）
C云ふ〔いふ〕（SEFは）
Cもの（S1）
Lの（WR）

『坊っちゃん』

C親譲り〔おやゆずり〕（S1）
Lの（WR）
C無〔む〕（T6）
S鉄砲〔てっぽう〕（T1）
Sで（WR）
C小供〔こども〕（S1）
Lの（WR）
C時〔とき〕（S1）
Lから（WR）
C損〔そん〕（T1）
Lばかり（WR）
Cし（SEKさ）
Lて（WR）
C居る〔いる〕（SEIあ）
C。（YY）
C小学〔しょうがっ〕（T1）
S校〔こう〕（T6）
Cに（WR）
C居る〔いる〕（SEIあ）
C時分〔じぶん〕（T1）
C学校〔がっこう〕（T1）
Lの（WR）
C二〔に〕（X7）

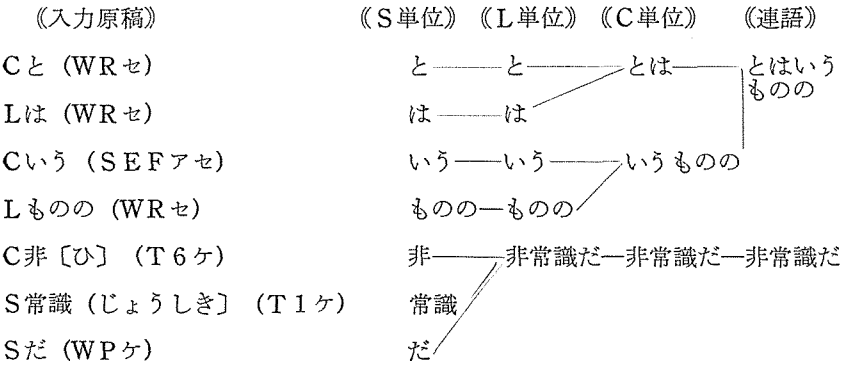
Cやっと（SC）　　S階〔かい〕（T6）

C出来〔でき〕（SEIか）　　Lから（WR）

S掛かっ〔かかっ〕（SEFら）　　C飛び〔とび〕（SEFば）

Lた（WP）　　S降り〔おり〕（SEIら）

S単位がいくつか集まって別な品詞となっている連語の処理は，たとえば，「とはいうものの」という接続詞や「非常識だ」という形容動詞の場合，つぎの左側の入力原稿のように情報を付けておけば，右側のように処理されるようになっている。



以上のようにプレエディットした原稿を漢字テレタイプで紙テープにさん孔して電子計算機処理にまわすのである。電子計算機の処理システムについては中野論文で述べられている。

(注 9)　作業では読み仮名を現代仮名づかいで入れた。それは，ルビの付いていない本文の場合，作業者への負担がより軽いと考えたからである。

(注10)　これらの情報のうち，語種・品詞・活用コードは，『新聞の語彙調査Ⅱ』（国立国語研究所報告38，11ページ）に発表されたものを基本にした。

4. あとがき

今まで述べてきたことにより，新しく開発されたシステムと，それによって作られた言語データが，他のシステム，データとどう違うかということが明ら

かになった。

これだけのシステムができたのは，電子計算機・漢字プリンタなど機械類の発達と，電子計算機による言語処理の研究の成果があったからである。

今後望まれるのは，これらのデータを蓄積し，それをもとに単位切りや情報付加の作業を自動的に行う方式（「一貫処理システム」と名付けている）の実用化によって，人手作業を軽減することと，入力のスピード化・簡単化——たとえば光学文字読取装置の実用化など——といったことである。これらも現代の技術開発の状況を見るとそう遠い将来ではないようである。

【寒山拾得】

あー	465	06	C	SB----	0007120	うーたー、ー「ー	あー	、ー給りーまし
間ーにー	462	01	C	WR----	0001690	ー此ー三日ーのー	間ーにー	、ーさー人
生憎ー	462	06	C	SC----	0002490	しーたー時ー、ー	生憎ー	こちらへーられ
逢うーたー	468	12	C	WP----	0012640	出ーたーのーにー	逢つーたー	時ー、ー
逢うーてー	462	11	C	WR----	0003150	、ー「ー寺に伺ー	逢つーてー	見るーか
青ー白いー	468	08	C	SM----	0012000	ー晴れーてー、ー	青ー白いー	目ーがー
垢ーつくー汚れる	463	04	C	WP----	0003940	でありーたー、ー	垢ーつきー汚れー	法
明家ーにー	470	08	C	WR----	0015240	ー「ー只今ーもー	明家ーにー	なつーて
明家ーにー	470	10	C	WR----	0015610	手ーのーあーたー	明家ーにー	連れーて
明家ーをー	471	02	C	WR----	0016190	ー忙しーけーにー	明家ーをー	出ーたー
諦めるー	467	12	C	SEIC--	0010890	なー人ーだーとー	諦めー	、ー別ーに
上げるーたりー	471	13	C	WP----	0017340	像ーにー香ーをー	上げーたりー	、ー掃
上げるーたりー	471	13	C	WP----	0017370	ー、ー燈明ーをー	上げーたりー	、ー其
足跡ーをー	470	12	C	WR----	0015830	上ーのー虎ーのー	足跡ーをー	指ーさし
足ーにーはー	473	09	C	WR----	0019870	ー出しーてー、ー	足ーにーはー	草鞋ー
足ーにーはー	473	10	C	WR----	0020000	ー被つーてー、ー	足ーにーはー	木履ー
味うーてー	462	02	C	WR----	0001910	さいーことーをー	味つーてー	、ー意気
与へるーのーであるー	468	10	C	WP----	0012460	ーにー満足ーをー	与へるーのーであるー	
頭ーにー	465	03	C	WR----	0006940	つーとー闇ーのー	頭ーにー	吹きー付け
頭ーをー	473	09	C	WR----	0019840	ーすー伸びーたー	頭ーをー	剝きー出し
頭ーをー	473	12	C	WR----	0020180	けーたー時ー、ー	頭ーをー	剝きー出し
当るーてー	473	07	C	WR----	0019760	つーてー火ーにー	当つーてー	ゐるーの
当るーてー	471	05	C	WR----	0016530	のーとー火ーにー	当つーてー	をりーま
あちらーのー	471	04	C	WR----	0016460	ますー、ー先刻ー	あちらーのー	厨ーで
跡ーからー	471	02	C	WR----	0016230	たー、ーそしてー	跡ーからー	付いーて
址ーはー	470	07	C	WR----	0015170	ー其ー伽藍ーのー	址ーはー	どうーなつ
跡ーをー	474	06	C	WR----	0021100	ー、ー跪いーてー	跡ーをー	見ー送つー
あなたーはー	463	09	C	WR----	0004390	つーたー、ー「ー	あなたーはー	台州ー
あなたーはー	466	02	C	WR----	0008120	ますーまいー、ー	あなたーはー	どちら
逢ふーうーとー	463	03	C	WR----	0003820	ーとー聞いーてー	逢はーうーとー	云つ
逢ふーたいーとー	463	07	C	WR----	0004300	「ーわたしーにー	逢ひーたいーとー	云
逢ふーにー	466	09	C	WR----	0008720	ー台州ーにーはー	逢ひーにー	往つーて
編むーだー	473	09	C	WP----	0019960	木ーのー皮ーでー	編んーだー	帽ーをー
阿羅漢ーですーね	470	07	C	WR----	0015130	ー、ー活きーたー	阿羅漢ーですーな	、
改めるー	468	06	C	SEIC--	0011690	ーはー衣服ーをー	改めー	輿ーにー乗つ
洗ふーせーてー	472	04	C	WR----	0017750	ーのー食器ーをー	洗はーせーてー	をり
洗ふーますー	472	08	C	WP----	0018140	ーがー食器ーをー	洗ひーますー	時ー、
或るー	472	01	C	SD----	0017460	、ーそのーうちー	或るー	日ー上座ーの
或るー	470	01	C	SD----	0014560	ー、ーするーとー	或るー	日ーふいーと
或るー	470	04	C	SD----	0014740	ございましー、ー	或るー	日ー出ーから
あるー	461	05	C	SEFE--	0000740	下ーにー郷ーがー	ありー	郷ーのー下ー
あるー	462	05	C	SEFE--	0002350	ーはー因縁ーがー	あるー	、ー関ーがー

0001

(坊っちゃん)

云ふ	281	06	C	SEF9--	0184380	は御授業に及はん	云ふ	から，おれはか
云ふ	263	14	C	SEF9--	0101810	ふと，いかんぜ．	云ふ	なら僕丈に話せ
云ふ	287	02	C	SEF9--	0209920	は遠慮するがい．	云ふ	ならフランクリ
云ふ	379	15	C	SEF9--	0584470	入つて行かんと．	云ふ	位謹直な人が，
云ふ	304	11	C	SEF9--	0284360	と考へて見たが，	云ふ	なら人を驚ろか
云ふ	275	12	C	SEF9--	0154360	したか，云へ」「	云へ	て，入れんもの
云ふ	354	13	C	SEF9--	0484050	する許り，大きく	云へ	ば世の中の為に
云ふ	353	08	C	SEF9--	0478290	んの言葉を借りて	云へ	ば，正に大違ひ
云ふ	257	11	C	SEF9--	0074060	さうだ．文学士と	云へ	ば大学の卒業生
云ふ	331	04	C	SEF9--	0390270	くなつた．延岡と	云へ	ば山の中も山の
云ふ	275	10	C	SEF9--	0153990	やあろ」「馬鹿あ	云へ	．バツタが一人
云ふ	275	11	C	SEF9--	0154310	たづらをしたか，	云へ	」「云へて，入
云ふ	342	08	C	TEF9--	0431580	働いて，人が何か	云ふ	と，ちやんと逃
云ふ	301	02	C	TEF9--	0268420	て見たら韋駄天と	云ふ	怪物ださうだ．
云ふ	343	14	C	TEF9--	0437380	，其際関係したと	云ふ	話だが，成程見
云ふ	331	04	C	TEF9--	0390470	中だ，赤シヤツの	云ふ	所によると船か
云ふ	330	09	C	TEF9--	0387170	より重大な責任と	云へ	ば，数学の主任
言ふ	307	03	C	SEF9--	0295530	はだまつて何にも	言は	ない，赤シヤツ
言ふ	367	06	C	SEF9--	0537380	から，余計な事を	言は	ずに絵筆でも●
言ふ	330	07	C	SEF9--	0386740	が，一判然とは今	言ひ	にくいが一まあ
言ふ	354	06	C	SEF9--	0481790	ない．何てうまく	言ひ	抜けられる様な
言ふ	334	02	S	SEF9--	0400220	をあげてやると御	言ひ	たら，難有うと
言ふ	307	07	C	SEF9--	0296710	僕今一寸失念して	言ひ	落しましたから
言ふ	244	05	C	SEF9--	0013690	おれを勘当すると	言ひ	出した．其時は
言ふ	350	02	C	SEF9--	0463630	見て，済だ済だと	言ひ	直して居る．と
言ふ	380	07	C	SEF9--	0586340	と見えて，わつと	言ひ	ながら，尻持を
言ふ	313	02	C	SEF9--	0323560	ドンナマドンナと	言ふ	といてるでなも
言ふ	334	01	C	SEF9--	0400040	ぢやけれ，お婆の	言ふ	事をきいて，赤
いふ	270	10	C	SEF9--	0132610	isrightと	いふ	英語を引いて説
いふ	361	08	C	SEF9--	0511430	が始まつた．議と	いふ	から藤間か何で
いふ	379	13	C	SEF9--	0583940	へ泊つて悪るいと	いふ	規則がありますす
いふ	242	03	C	SEF9--	0004020	此質屋に勘太郎と	いふ	十三四の倅が居
いふ	317	02	C	SEF9--	0334150	田さんの方がえと	いふ	ぞなもし」「つ
いふ	289	12	C	SEF9--	0221280	顔を掻いた．何と	いふ	猪口才だらう．
いふ	371	15	C	SEF9--	0555550	か．辞表を出せと	いふ	なら公平に両方
いふ	344	08	C	SEF9--	0440120	器の事を瀬戸物と	いふ	のかと思つて居
いふ	334	04	C	SEF9--	0400880	うたつてる．謡と	いふ	ものは読んでわ
いふ	282	14	C	SEF9--	0190870	，此野だは，どう	いふ	了見だか，，赤
いへ	281	06	C	SB----	0184480	はかう答へた．「	いへ	，ちつとも心配
云へる	275	13	C	SEI9--	0154650	て自分のした事が	云へ	ない位なら，て
云へる	294	01	C	SEI9--	0238250	へ，公然と名前が	云へ	ない位な男だか
云へる	321	11	C	SEI9--	0353940	てないから何にも	云へ	ないが全く美人

0067

(坊っちゃん)

	297	03	あなたが迷惑ならよしませうと受け	合つ	た．君大丈夫かいと赤シヤツは念を押
	297	05	疑ぐつてる．擲りながら男だ．受け	合つ	た事を裏へ廻つて反古にする様なさも
	298	08	してやるんだが，口外しないと受け	合つ	たんだから動きがとれない．人がこん
	359	11	夜番をやるのかい」「まだ枡屋に懸	合つ	てないから，今夜は駄目だ」「それぢ
	364	01	な師範生が，十五六の中学生と組み	合つ	てゐる．止せと云つたら，止さないか
	302	04	かの連中は隣り同志で何だか私語き	合つ	て居る．手持無沙汰無聊なのは鉛筆の
	367	14	ない．夫でなければあ云ふ風に私語	合つ	てはくすくす笑ふ訳がない．校場へ出
	294	04	ち，虫の好かない奴が親切で，気の	合つ	た友達が悪漢だなんて，人を馬鹿にし
	355	06	返したり，押し返されたりして揉み	合つ	て居る．前方から静かに静かにと声を
	305	14	は赤シヤツ党だ．こんな連中が寄り	合つ	て学校を立てて居りや世話はない．お
	324	13	だらう．箱根の向だから化物が寄り	合つ	てるんだと云ふかも知れない．おれは
逢う	272	10	い事を云ふから「校長にはなつたら	逢つ	た，暑い時には散歩でもしないと宿直
	327	15	んなに隠さないでもよからう，現に	逢つ	てるんだ．よく嘘をつく男だ．是で中
	255	11	中から小倉の制服を着た生徒に沢山	逢つ	たが，みんな此門を這入つて行く．中
	261	07	は通町で氷水を一杯奢つた．学校で	逢つ	た時はやに横風な失敬な奴だと思つた
	327	10	て，交際をつづけて居る，野芹川で	逢つ	た翌日杯は，学校へ出ると第一番にお
	323	05	愛へ下りて見たら，又うらなり君に	逢つ	た，おれは会議や何かでいざと極まる
	350	11	た，野だが，鈴ちやん逢ひたい人に	逢つ	たと思つたら，すぐ御帰りで，御気の
	246	11	てさめて仕舞つた．こんな婆さんに	逢つ	ては叶はない．自分の好きなものは必
	266	08	に踏まりさうだが，かう骨董責めに	逢つ	てはとても長く続きそうにない．其う
	301	13	はないと思つてたが，うらなり君に	逢つ	てから始めて，矢つ張り正体のある文
	336	11	なり君にもうらなりの御母さんにも	逢つ	て詳しい事情は聞いて見なかつたのだ
	356	09	，矢つ張り東京迄出掛けて行つて，	逢つ	て話をする方が簡便だ．清の心配は察
青い	357	03	みる所は頗る珍らしいものだ．あの	青い	実が段々熟してきて，黄色になるんだ
	361	03	おれの頭の上で，ぱかりと割れて，	青い	烟が傘の骨の様に開いて，だらだらと
蒼い	379	14	寧な言葉を使つてる．顔の色は少々	蒼い	，「取扱上不都合だから，蕎麦屋や団
	309	10	らなり君で，おれが，かう云つたら	蒼い	顔を益蒼くした．おれは即夜下宿を引
	301	10	，温泉へ行くと，うらなり君が時々	蒼い	顔をして湯壺のなかに膨れて居る．僕
	258	02	黄顔色の悪るい男が居た．大概顔の	蒼い	人は痩せてるもんだが此男は蒼くふく
	258	06	るんですと教へて呉れた．それ以来	蒼く	ふくれた人を見れば必ずうらなりの唐
	374	01	のない明々白々たる理由だが，狸が	蒼く	なつたり，赤くなつたりして，可愛想
	258	03	の蒼い人は痩せてるもんだが此男は	蒼く	ふくれて居る．昔し小学校へ行く時分
	258	05	うらなりの唐茄子許り食べるから，	蒼く	ふくれるんですと教へて呉れた．それ
	309	10	，おれが，かう云つたら蒼い顔を益	蒼く	した．おれは即夜下宿を引き払つた．
仰ぐ	256	07	生徒の模範になれの，一校の師表と	仰が	れなくては行かんの，学問以外に個人
	305	04	す．どうか成るべく寛大の御処分を	仰ぎ	たいと思ひます」と云つた．野だの云
青島	283	12	針の様に●がつてる，向側を見ると	青島	が浮いてゐる．是は人の住まない島だ
	284	09	も這入つてるなら迷惑だ．おれには	青島	で沢山だ．あの岩の上に，どうです，
青空	289	05	めたと云ふのか方角がわからない．	青空	を見て居ると，日の光が段々弱つて来
仰向	280	06	いと引いたら，そいつは，どたりと	仰向	に倒れた．ざまを見ろ．残る一人が一
仰向け	272	14	を跳ねのけて，頭と尻持を突いて，	仰向け	になつた．おれが寝る時に頭と尻持
	287	11	おれは一匹で懲りたから，胴の間へ	仰向け	になつて，さつきから大空を眺めて

0003