A system of the word count program 2

電子計算機による語彙調査　Ⅱ

——主として短単位処理について

斎藤秀紀

0　はじめに

国立国語研究所で行なわれている語彙調査も現在第二段階を迎え，短単位処理による調査が進められている。長単位処理による調査については，昭和41年度に起案され現在にいたっているが，調査の重点を速報性においた結果，付加情報は最小限必要なものに限定されている。それは，長単位処理自体，短単位処理のプレエディトとしての性格を持ち，単位切りの能率と電子計算機を使用する上での大量のデータを扱う問題点をさぐる目的があったためである。

以上の点から，短単位処理では，より多方面の分析にたえられるよう語種，品詞，活用形情報，その他漢字の仮名付け等，文法情報の付加処理を行った。また異なる二つの語の単位（長単位，短単位）の接続をはかるため，原文の内容を容易に参照できるよう配慮し，用例表を作成する。これらの用例表は，長単位の問題点であった同形異語処理，また短単位処理における同音異語の判定を可能にし，さらに資料としても，意味の研究や自動単位切りの問題等，語の認定における自動化の研究に対し，貴重な資料になる。

計算機による処理の概略は図1に示してあるが，細部については各担当者の報告を参照されたい。なお短単位処理のアウトプットとして予定しているものは，次の五種の語彙表である。この調査に使用した機械はＨＩＴＡＣ—3010形電子計算機一式，漢字等のデータ入力機器としては漢字テレタイプライターを使用した。

1　50音順短単位表

見出し語に語種，品種，活用コード及び見出度数を50音順に配列した

もの。

2　度数順短単位表

50音順の配列を度数順に再分類したもの。

3　活用形語彙表

各活用語について代表形（終止形）と度数を示し，変化形別の度数カウントを行なう。

4　語種品詞別語彙表

各見出し語を品詞別に分類し，度数，類内順位，類内使用率を示す。

5　50音順用例表

見出し語の用例をＫＷＩＣ形式で仮名印字したもの。

1.　システムの概要

短単位処理システムにおいて，特に留意したことは次の二点である。

1.　エラーデータ処理は，チェック点で判定記号を挿入し，ファイルからの分離を行なわず他のチェック済みデータと同一ファイルにまとめる。
2.　磁気テープフォーマットは形式を規順化し全体を印字処理とデータ処理関係の二種類に統一する。

システムの効率を上げるためには，演算時間，入出力時間の短縮をはかることが重要であるが，もっとも大きな障害となっているのは，いわゆる Man-machine communication と言われる人間と機械の Interface である。特に個々のデータについて原文出典を参照するエラー処理において密接な関係を示し，効率化の問題はこれら修正の方法いかんにかかっていることが多い。一般

1）中野　洋　語彙調査の類別語彙表について（国立国語研究所報告34）
2）江川　清　「活用形処理」の自動化に関する一方式（同上）
3）石綿敏雄　新聞用語調査の用例印字プログラム “COBOL-KWIC”（本報告）
また長単位処理の概要については国語研究所報告31，34，を参照されたい。

のバッチ処理では各ランごとにエラーデータを別ファイルに分離し，適当なサイクルで一括処理する方法が多くとられる。しかしメインシステムの進行に合わせる場合，いずれの場合も，見出し語の照合分離抽出する過程において，データのビット変化や脱落をそのまま再元して入力しなければならず，作業能率の向上はあまり期待できない。また，調査全体のシステムを長単位処理と短単位処理の二段階に分けた結果，エラーデータの種類が複雑になり，個々のエラー別ファイルの作成は無駄が多く，データ管理面においても問題が多い。そこで，修正処理については，任意の位置でデータの追加，削除ができることが望ましく同時にメイン処理に対し，割りこみ処理の形で早急に修正ランの挿入が実行できるような機能が必要になる。

短単位システムでは，以上の点を考慮し，エラーデータについては，個々に分離することはせず，チェック記号を挿入する方法をとった。これで従来の遂次処理方式と集中処理方式の両方式が可能になり，任意の位置でランの進行状態と適当に合わせて接続点を選ぶことができる。またエラーによっては，ビットの変化等再現しにくい状態のものも，テープ中のビット変換によって比較的らくに修正でき，作業段階でのデータ脱落等二重ミスをさけることができる。その他のエラーデータにも度数の修正を含んだ処置が可能になる。

システム構成の基本的形式はループ状態をなすもの，Tree（木）構造をなすものの二つに分けられる。木構造の特徴としては，前者に比べ各部門間の緊密な連絡を必要とせず，各々に独立した処理体系を組むことができる。これは，従来の行政組織をそのまま利用でき，調査の目標基準が各個人単位で計画され組織全体にまでおよばない場合に特に有効な形態であり，直線的な処理方法からは，高度に組織化された集団を必要としない等，利点が多い。Computer 利用技術としては最も低次のものであろうが，システムの納期等限られた期間内に目的を実行しなければならない場合，さらに従来の人手による組織を活用せざるをえない場合等この形式になることが多い。本調査のシステムにおいても基本的にはこの形式をとったが，末端では各々疎な関係にあっても調査目的の主たるものが資料作成におかれた場合，一応の目的は果たされるため問題は起

きない。しかし，データは多方面に分枝される可能性が生じるため，各部門間の接続は特に柔軟性を持つことが要求される。

接続点についての，磁気テープ形式と処理時間とは大量のデータを扱う関係から，計算機の入出力時間を最小にするよう注意しなければならない。しかし，テープの編集，分類作業等の割合も全体としてかなりの部分をしめるため，項目の固定化と入出力時間の短縮という相反した関係の調整も必要になる。また項目の位置と桁の設計には分類処理のキーに制約されるため，直接関係のある項目については固定化しなければならず，さらにデータの冗長度が増す結果になる。しかし，キー配列構成を連続指定することにより，処理時の短縮をはかることは可能であり全体のバランスは保たれるものと考える。

一般には前述のとおり，各ランのアウトプットは共通ファイル的性格を持っており，データは使いやすさと同時に，用語調査と研究部門の両端において満足できる内容を持つことが必要である。またシステム拡張にともなうプログラム変更を無理なく吸収できる余裕と研究結果をシステムにフィードバックできる態勢等，常にシステムの拡張に対し柔軟性を保てることが必要である。

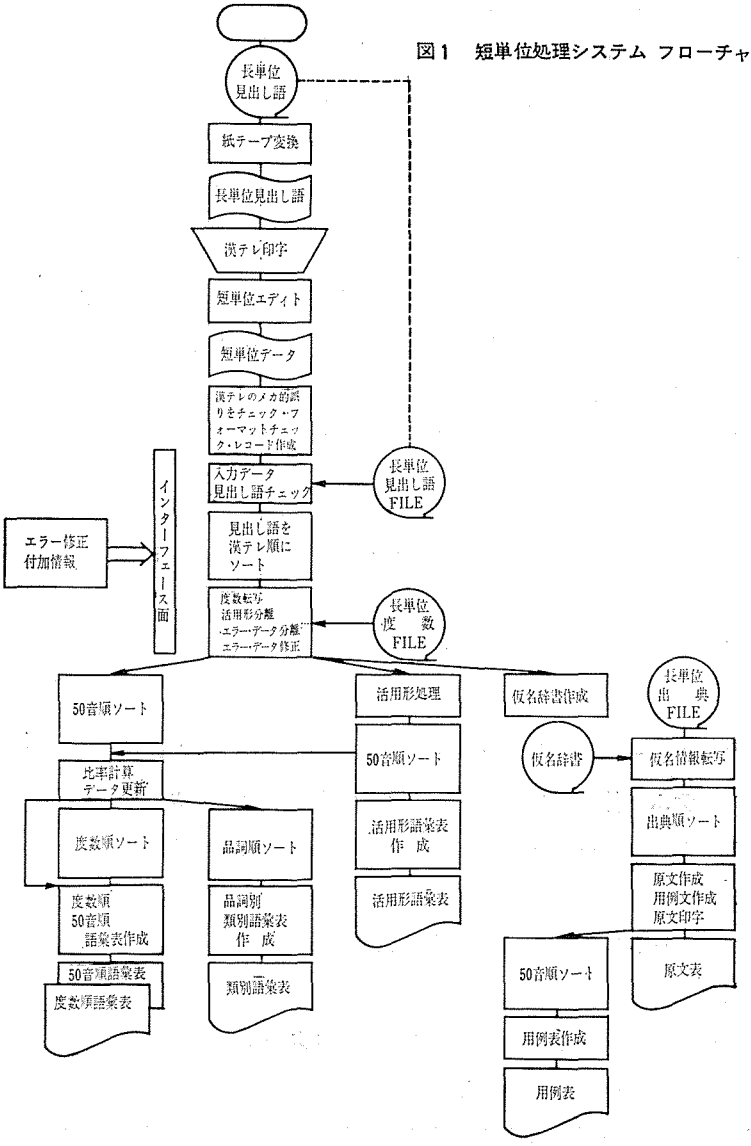
システム条件〔II〕は以上の点の考慮し設定されたものである。

形式の統一については，関係部門で作成されたプログラムを共通に利用できる利点もある。周辺処理と基本システムとの接続は，ラン進行に左右されない疎な関係にあることが望ましいが，データ形式の規格化によってデータ分類順序の変更，情報付加処理のいかんにかかわらず任意の位置で接続可能になる。周辺処理については，基本システムをささえ，システム拡張要素を含んだ多面的なものにすることができ，大規模な機械処理では不可欠のものである。また，機械辞書の共通利用も可能になり，同時に将来辞書方式による集中処理へ移行する場合の貴重な基本資料も得られよう。

ラン1

前述のとおり，前処理的性格を持つ長単位処理と短単位処理とは接続面に中間エディット作業が入り，エラーの種類を複雑にすることになった。またシステム構成面から，長単位情報の転写等見出し語の一致を必要とする場合が多い。

図1　短単位処理システム フローチャート



短単位処理では，エディット作業の能率化をはかるため長単位処理で出現した語の異なりについて作業を行なってある。総出現度数を知りたい場合，長単位語より度数の転写を受け，短単位語について再集計を行なわなければならない。

エラーの種類については，打鍵ミス，漢テレの誤動作によるさん孔ミス等，またエディット作業によって生じる単位切り，付加情報及び校正漏れが主な誤りである。このランでは，主に見出し語についてチェックし，以下に続く各チェック部門のエラーの揺れを少なくする前段階的処理を目的とする。しかし，処理過程で必要に応じ，チェック項目を省くことができるがシステムには影響を与えない。チェックの条件としては，短単位作業のリスト表と再入力される短単位データとは一致する前程において行い，不一致の見出し語は全て修正の対象として処理を行う。なお不一致の原因として考えられるものは次の三つの場合である。

1）長単位エラーを短単位で修正したもの。
2）短単位処理で新たに誤りの発生したもの。
3）長単位エラーの修正不能なもの。

検出されたデータは，各々チェック点で表示記号を挿入し，他のデータと区別し，システム条件のI）を満足するよう照合済みのデータとの分離は行なわない。しかし，エラーファイルに分離することは，任意の位置で可能である。また入力データについては，50音順に分類されていることが望ましいが，ランダム順序のデータについても処理可能である。

分類済みの場合

エラー記録用ファイルは使用されず，マスターに全て記録される。データの判別はエラー表示記号によって行なう。

U　長単位テーブルに見出し語が存在するにもかかわらず，短単位データの入力がなかったもの。

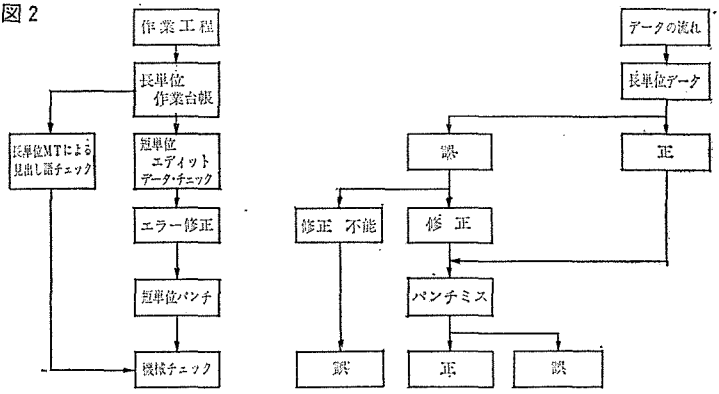
＊　入力された短単位データと長単位テーブルとの間に一致する見出し語がなかった場合。

ランダムの場合

エラー専用の磁気テープを使用する。上記の＊のデータが分離され，エラー用ファイルに書き出される。入力のチェック範囲は，見出し語の照合に止まり，テーブルから脱落リストの転写を受けることは不能である。

これらの処理上の選択は計算機上の操作卓から行ない，処理については，全て自動的に行なわれる。図2は短単位処理で新たに発生すると予想されるエラーの種類と発生点である。なおこのランの処理は一紙半年分単位に実行される。

図2



ラン2

前記ランで1サンプリング単位ずつ処理された短単位データを一本の磁気テープに併合する。このランでは，50音順，漢テレ順いずれも処理可能である。これらの処理の選択指示は計算機の操作卓上の割り込みボタンによって行なう。サービスルーチンを使用することもできるが，特別の場合以外はこのプログラムを使用する。

ラン3

このランではプログラム処理はおよそ次の通りである。

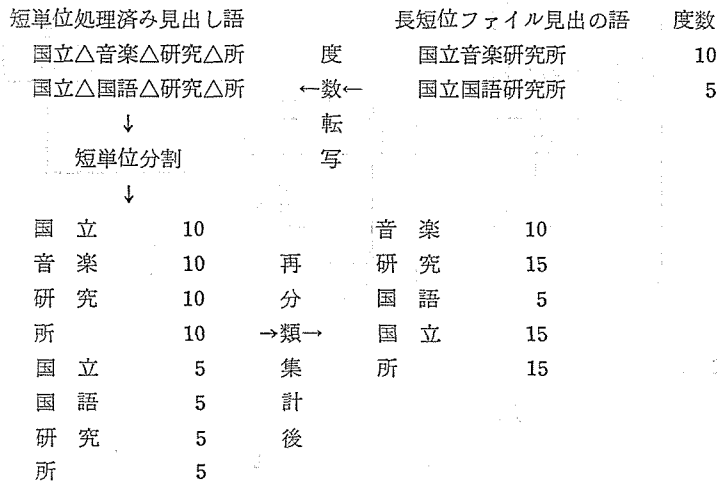
1）長単位度数ファイルを利用し，短単位レコードに度数を転写する。

2）エディット記号に従って，長単位から短単位に見出し語の分割を行う。

3）必要に応じエラーデータの分離を行う。

短単位レコードに長単位度数を転写し，必要に応じラン1で処理された不一致データを分離する。度数転写については，システム構成上，短単位で部分集計を行い，それに長単位度数を加算して総度数を求める。度数の転写方法は，いずれも両ファイルとも見出し語を漢テレコード順に分類しておき，照合して一致した語に度数を転写する。次の例は度数の配分の状態を表わしたものである。

図3



ラン4

前記処理で分割され，50音順に再分類されたデータを短単位の語形ごとに度数集計を行い，同時に分散された磁気テープを一本に併合する。このランのシステム中の位置は，予期しない，プログラム変更やシステム修正が発生した場合，極力この部分で吸収し，システムのバッファーとしての機能を持たせることにある。これによって，他のプログラムにまで変更が波及するのを防ぎ，こ

の部分で集中的に操作することを目的とする。

以上システムの前部分の概要である。印字部については，全てオプションとなるため，磁気テープフォーマットその他は機能本位に構成されている。以下のプログラムは主として用語調査における度数順語彙表・50音順語彙表を作成するランである。図3は印字部のフローチャートである。

作業手順はシステム・フローに示す様に，ソート及び3つのプログラムより構成されている。第1のプログラムは比率計算用パラメータにより，順位，比率等を計算するプログラム。第2のプログラムは度数順及び50音順の語彙表を，アウトプットする語彙表作成プログラムである。第3のプログラムは見出し語処理のプログラムで紙テープにパンチした後，漢テレにより印字するプログラムである。

ラン5

1紙1年分マージ済みデータを度数順ソート（下降順）し，そのデータをマスター・ファイルとする。パラメータとしては，前段階のプログラムにより集計された全体延べ語数・部分延べ語数・度数情報を使用して全体順位，比率・累積比率・部分順位・比率・累積比率を計算，その他の情報と共にニューマスタ・ファイルを作成する。なおパラメータ指示により任意の度数迄ファイルを作成し不必要な部分については，任意に削除することができる。

パラメータのフォーマット

×……×A……AB……B

7桁　7桁　7桁

××××　：語彙表のためのアウトプットする度数の値を示す。

AAAAAAA：全体の語に対する延べ数語を示す。

BBBBBBB：部分の語に対する延べ語数を示す。

ラン6

度数順語彙表・50音順語彙表を作成するためのフォーマット変換及びライン

プリンタにより語彙表をアウトプットし，見出し語について見出し語処理用ファイルを作成する。度数順語彙表・50音順語彙表・見出し語処理用ファイルはそれぞれマニアルの指定により，度数順語彙表又は50音順語彙表のいずれかをアウトプットし，見出し語処理用ファイルに対しては必要・不必要を指定する。語彙表は見出し語25個で1ページ分とし，200ページごとにリランが取られる。

語彙表の内容は，ラインカウンター（1ページ25行），判別コード（活用形処理済みデータは＊印，活用形処理において出て来たエラーデータは＃印），見出し語（見出し語は漢テレコードで入力されているので，漢字・ひかガナ・カナ文字変換によりカナ文字で次の表示がなされる。）。

ローマ字・数字	ローマ字・数字
盤内記号	—バイナイ—
盤外記号	—バンガイ—
特殊文字	—トクシュ—
情報無視	ノウカウント
エラーデータ	・・・・・・

コード番号・度数・全体順位・全体比率*・全体累積比率・部分順位・部分比率**・部分比率等である。50音順語彙表の場合，全体・部分ともに累積比率は印字しない。なお印字フォーマットは印字形式を参照。

ラン7

ラン6でアウトプットされた見出し語処理用ファイルにより，紙テープ出力フォーマット変換およびパンチ・アウトする。紙テープは50ページ分1巻とす

---

* パラメータで指定した全体延べ語数で度数を割り比率を計算する。単位は0/00（パーミル）で小数点以下4位を四捨五入した値。以下各比率は全てこの方法に従う。

** パラメータで指定した部分延べ語数で度数を割り比率を計算する。

る。なおこの時にリランが必要に応じて取られる。

印字形式は（度数順語彙表）表1の通りである。

紙テープにパンチされた見出し語は漢テレ印字する。印字フォーマットは印字形式を参照

「50音順語彙表」は，表2の印字型式である。

2. エラー処理

エラー処理については，従来の付属的な位置からメイン部と対等の状態になるよう考える。通常エラー処理システムを考える場合，メイン部と任意に接続でき，更新ランの周期と同期させやすい構成が必要であり，同時にエラーデータの削除，修正が容易に行なえることが望ましい。また，場合によってはエラーの移動が多くなるため，修正部門との受けわたしリストを正確にチェックする管理面への配慮が必要である。リストの製表については，直接修正用台帳に利用できるよう構成し，他の台帳への転写その他の手作業は入れないようにすべきである。表3は，短単位関係のエラーを抽出したものである。

表3の見出し語は漢テレによって別印字したものである。ＢＯＯＫ#，ＰＡＧＥ，ＷＯＲＤについては，このシステムのラン1で述べた短単位エディット用原稿の情報を表わす。度数については，長単位度数ファイルより，見出し語の一致したものに転写を行なう。その他見出し語の不一致のものについては，記号“U”で区別する。他の度数を持たない部分は，短単位処理で新たに処理された，長単位エラーの修正されたもの，短単位エラー等であり記号＊で表わす。＊記号については，短単位処理によって新たに挿入された見出し語であり，主に長単位処置におけるパンチミス，単位切りミス等である。表3では02の見出し語は短単位エディットで　阿→あ　に変更され入力されたものである。長単位台帳に登録されている『阿倍野川大斎場』は短単位データとしては

入力されないため記号が付加されており，08，09の見出し語についても同様に処理されている。18，19，20については，『アサ1005』は『アサ』『1005』に細分割されているが，u記号と＊記号の状態では長単位単位切りミスであろう。単位切りミスはほとんどの場合u，＊は対応ずけられるはずである。この修正ランでは，当分人手による修正にたよるが，将来は自動処理の方向にもって行く予定である。

3，システムの問題と今後の方針

システム設計の要点としては，次の点を明らかにすることが必要である。

1）調査の目的と目標になる具体的なテーマ。

2）確保可能な人員と機材の能力の範囲。

この2点は相互に影響しあうものと思われるが，研究上の試行錯誤を含み多方面の研究者が集まる場合，1）については，要求を同時に満足しようとするため，当初の目標が不明確になりやすい。また機械応用の経験に対して歴史のあさい場合，2）の査定を誤り初期の目的が達せられない場合も生じる。一般にシステム構成の上で設計以前に調査思想の統一，作成資料の利用方法，組織の体制等機構上の統一をはかっておくことが必要である。これらの過程において，調査の目的，使用範囲，前回の調査から次期に対する流れ，さらにこれが確認され，長期計画の上に立っての個々の調査目標の決定，また長期計画の修正へと拡張されていくからである。

一般に調査目的と用語表の関係は密接である。用語表の作成については，利用の対象から次の二つの場合が考えられよう。

1）一般に語の用法を調べるため，多数の用例を必要とする場合。

2）文字または語の統計的性格を調べる場合。

1）については，個々の目的に合った第2次情報の付加に耐えられる第1次資料的性格を持つものでなければならない。なぜならば，これらの資料は現在一般に対象が言語学方面のみならず Computer 利用の言語情報処理研究者に

も広がっており，特に具体的な問題を持った人々が多く，これらの要求にも答え得るものでなければならないからである。普通，単語またはそれに近い語形に分割され，集計された後では，調査対象の傾向を調べる場合に一つの意味を持つにしても，失なわれる情報量が多く，第1次資料としての性格が失なわれるからである。2）については，1）から任意な再処理が可能である。しかし逆の場合は必ずしも成立しないことは前述の点からも明らかであろう。

現在では Computer 利用の用例表の作成は，大体 KWIC 形式になることが多い。使用上の注意としては，収集されたデータは最大公約数的になり収録漏れについての扱いが問題になり，データ数増加との悪循環を断ち切る方法が必要である。

以上の2面を満足させるため，調査段階を二つに分けることが必要であろう。前部分では，第1次資料の作成，後部分ではそれらの資料を使っての分析を含む処理である。これで内外の利用者に対し，必要な段階ごとの資料を提供でき，かつ過去の記録の蓄積という，将来再処理の必要性に対するデータの保存としての性格も可能になる。これは調査資料の基本的性格としては，公刊物としての面と資料の記録性の2面を持っているものと考えるからである。

また第1次資料を利用し，さらに情報の付加を行なう場合，システム構成面からシソーラス利用の可能性を持たせるため，単一辞書方式による処理方法を考えてみたい。

単一辞書による集中処理方式の利点としては，およそ次の点で有利であると考える。

1）システム構成を調査部門と辞書作成部門の二つに大別できる。
2）情報付加，エラー修正等の集中処理が可能になり，辞書その他の誤りを多方面から同時にチェックできる。
3）辞書項目と調査の段階内処理とを一致させることが可能。

1）における調査部門は，従来の語彙表作成処理を中心としたもので，処理上では辞書作成部門と対になるものである。これらは，システム管理上二部門

に分けたものであって，さらに次の二種のシステムに分けることが必要になろう。

イ）情報検索システム

ロ）数値統計システム

これら二種のシステムは，辞書作成，データ処理部門とも共通に利用できるように，分析用としてサブプログラムシステムの性格を持たせることが必要である。特に用語調査における情報検索システムの必要性は，大量の用語を，システムの進行に合わせて分析するために，従来の語彙表を中心とした人手による方法では問題が多く，特に特定の条件を持つ用語の分析を行なう場合には，致命的なものとなる。これは，不特定多数の研究者が，関連部門で研究されたデータを自由に利用しようとする場合にも同様であり，任意の条件で必要な情報を抽出する技術が必要になってくる。

数値統計処理システムは，大量の調査用データの概略を知るうえに，また語の相互関係を具体的な数値で表わす場合に必要であり，機械処理とは密接な関係にある。また言語が，時間や環境によって変化し，強い規則性を持つと同時に同義性等の意味のあいまいさを持ち，試行錯誤を必要とするため数値によっても確定した値を求めることはできない。しかし統計量による相対値を示すことは可能になり，同時に，集合による状態のパターンから確定できるものと，それ以外のものの分離の目やすに使用でき，一般現象から具体的構造をさぐる場合に，調査対象の範囲を限定するのに有効である。

その他，システム管理面からは，多人数の人々の関係をシステムに接続させる問題がある。通常は，各分野ごとに研究グループをもうけ，それぞれを一種のサブシステムとして独立させ，辞書の共動作成面で結合させる。これによって，管理体制の縦割りから，横への関係に拡張でき，いまだ機械化の不能な部門についても，関接的にシステムへの接続が可能になり，さらに辞書の共動利用という面から，新たな研究分野の発展も期待できるようになる。

一般的な調査の形態ではシステムの効率化と調査精度が問題にされるが，用語調査では，研究という平行処理を持つため，特率については，ある程度制限を受けることが多い。

辞書を使って情報付加を行なう場合，処理段階を必要に応じ分割することができるが，エラーその他の修正更新のさいには，辞書の種類だけ処理の増加をまねくことが多い。特に共同利用のため種々の情報を蓄積して行く過程で，人手を使う場合には，修正の情報管理には不手ぎわを起こしやすく，辞書作成面でのエラーの扱いが問題になる。

この点辞書の作成を集中的に行った場合，語彙表の利用者による多面的チェックも可能になり，誤りチェックを多部門から同時に実行でき，修正に対しては一カ所でコントロールできる利点もあり，データ管理面からも比較的更新しやすいものとなる。

また最近記録媒体としての外部装置も，容量，速度の面で満足できるものが多くなり，将来辞書内容の増加によって，シソーラスとして発展させることも考えられ，これらの方面からも実用化の研究が必要であろう。

その他，計算機利用のためのプログラムの使用は，従来のアセンブル言語から問題向き言語の一種であるＣＯＢＯＬ等のコンパイラーの使用を考える必要があろう。ＣＯＢＯＬは事務計算用に開発されたもので，非数値的情報（文字，記号類）も扱いやすく，言語情報処理にも十分たえられる機能を持つ。表現は自然語に近い英文で記述され，書きやすさと共に，他人にも理解しやすい論理構造をとることが可能である。さらに，電子計算機の機種に関係しない共通言語的性格を持っておりシステム変更にともなうプログラム変換作業をさけることができる。

またＣＯＢＯＬの一命令は，ほぼアセンブラーの一処理単位に相当し，プログラム作成，修正間時を大幅に短縮でき，システム増加にともなう人員の確保と共に，新人教育における教育期間の短縮とを同時に行なえる利点がある。以上の点からもＣＯＢＯＬ等一般コンパイラー言語への切換えは積極的に検討する必要があろう。

4. 結び

以上で短単位処理に関するシステム構成の説明を終える。新聞の用語調査システムは長単位処理システムと対になるものであるが，これらのプログラムの開発によって，解決しなければならない種々の問題が明らかになった。これらの大部分は，現在の言語情報処理の問題点ともあいつうじ，解決方法を外部に求めることも多くなることが予想される。また，逆に外部の要求に答え得る基礎研究部門の充実は，用語調査のあるべき姿の一面を持っていると思われ，将来の調査についても定性的に，より厳密な調査が必要になるであろう。今回の調査がこれらの調査に対し，基本的なデータを与え，さらに作成された資料が一般言語研究者に有効に利用されれば幸いであると思う。

この報告の最後に，この処理で使用された磁気テープフォーマットを示しておいた。一見無駄が多く非能率的な形式であるが，なるべく使いやすく一次情報の保存をはかったつもりである。最後に，このシステムの設計は言語計量調査室，第一資料研究室の全員によって検討された。また，設計の中核をなした木村繁氏は設計の中途で他に転出された。研究補助員の花井夕起子氏にはプログラム作成面で種々の協力を受けた。特に後部門については花井氏におうところが多い。

図2　印字部フローチャート

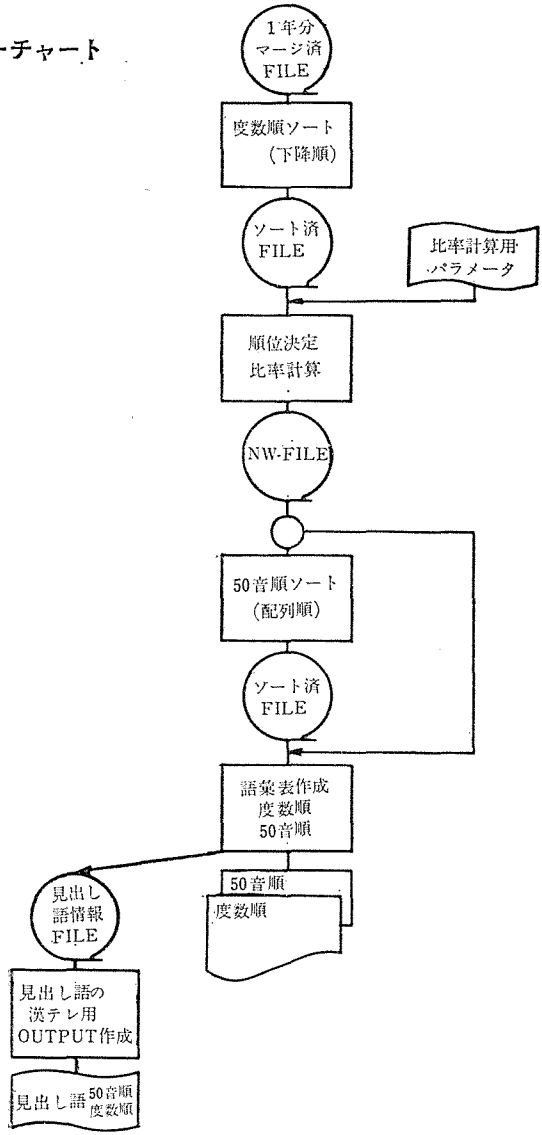


表 2

		ミダシ	コード	ドスウ	ジユンイ	ヒリツ	ジユンイ	ヒリツ
01	ア	ア	U０００	23	3353	.024	(2822)	.53
02	ア	ア	W８００	20	3735	.021		
03	ああ	アア	ＳＢ００	12	5713	.013	(4723)	.028
04	アート	アート	U０００	6	9648	.006	(7686)	.014
05	アートシアター	アートシアター	U０００	6	9648	.006	(7686)	.014
06	愛	アイ	Ｔ０００	89	961	.095	(773)	.206
07	愛	アイ	W８００	6	9648	.006		
08	会い	アイ	ＳＥＦ１	9	7196	.010		
09	相	アイ	Ｓ０００	5	10844	.005	(8567)	.012
10	合い	アイ	Ｓ＋Ｆ１	17	4271	.018		
11	あい	アイ	Ｓ000／ＳＥＦ1／Ｔ000	16	4500	.017	(3787)	.037
12	相営み	アイイトナミ	ＳＦＥＤ	5	10844	.005		
13	相営む	＊アイイトナム	ＳＥＦＤ	5			(8567)	.012
14	アイカ	アイカ	Ｔ０００	7	8664	.007	(6947)	.016
15	相変ら	アイカワラ	ＳＥＦＦ	7	8664	.007		
16	相変る	＊アイカワル	ＳＥＦＦ	7			(6947)	.016
17	愛好	アイコウ	Ｔ０００	9	7196	.010	(5862)	.021
18	愛国	アイコク	Ｔ０００	6	9648	.006	(7686)	.014
19	愛さ	アイサ	ＶＥＫ５	6	9648	.006		
20	あいさつ	アイサツ	Ｔ０００	44	1906	.047	(1610)	.102
21	愛し	アイシ	ＶＥＫ５	10	6615	.011		
22	愛情	アイジョウ	Ｔ０００	28	2852	.030	(2417)	.065
23	合図	アイズ	Ｖ０００	5	10844	.005	(8567)	.012
24	アイスパレス	アイスパレス	U０００	7	8664	.007	(6947)	.016
25	愛する	＊アイスル	ＶＥＫ５	20			(3153)	.046

表 1

		ミダシ	コード	ドスウ	ジュンイ	ヒリツ	ルイセキ	ジュンイ	ヒリツ	ルイセキ
01	ない	ナイ	ＳＬＭ０／ＷＰＰ０	2067	49	2.198	461.705			
02	〃	—バンナイ—	ＹＹ００	2035	50	2.164	463.869			
03	時	ジ	Ｔ０００	1965	51	2.089	465.958	(12)	4.557	89.632
04	なる	＊ナル	ＳＥＦＦ	1948				(13)	4.518	94.150
05	十	ジュウ	Ｘ０００	1880	52	1.999	467.957	(14)	4.360	98.510
06	東京	トウキョウ	Ｗ８００	1851	53	1.968	469.925			
07	する	スル	ＳＥＫ５	1837	54	1.953	471.878			
08	が	ガ	ＷＲ００	1796	55	1.910	473.788			
09	と	ト	ＷＲ００	1740	56	1.850	475.638			
10	いう	＊イウ	ＳＥＦ１	1682				(15)	3.901	102.411
11	六	ロク	Ｘ０００	1681	57	1.787	477.425	(16)	3.899	106.309
12	いう	イウ	ＳＥＦ１	1674	58	1.780	479.205			
13	て	テ	ＷＲ００	1629	59	1.732	480.937			
14	者	シャ	Ｔ０００	1617	60	1.719	482.656	(17)	3.750	110.059
15	だ	ダ	ＷＰＰ０	1610	61	1.712	484.368			
16	区	ク	Ｔ０００	1591	62	1.692	486.060	(18)	3.690	113.749
17	月	ガツ	Ｔ０００	1584	63	1.684	487.744	(19)	3.674	117.423
18	年	ネン	Ｔ０００	1568	64	1.667	489.411	(20)	3.636	121.059
19	この	コノ	Ｓ９００	1442	65	1.533	490.944	(21)	3.344	124.403
20	八	ハチ	Ｘ０００	1407	66	1.496	492.440	(22)	3.263	127.666
21	お	オ	Ｓ０００	1387	67	1.475	493.915	(23)	3.217	130.883
22	七	シチ	Ｘ０００	1304	68	1.386	495.301	(24)	3.024	133.907
23	い	イ	Ｓ０００／ＳＥＦＦ／ＳＥＧ２／Ｔ０００	1298	69	1.380	496.681	(25)	3.010	136.918
24	的	テキ	Ｔ０００	1290	70	1.372	498.052	(26)	2.992	139.909
25	第	ダイ	Ｔ０００	1263	71	1,343	499.396	(27)	2,929	142,839

エラーリスト

		CODE		BOOK#	PAGE	WORD	DOSU
01	あ倍野大斎場	＋5シＬＦＰ７ＰＴＢＩ９	*	1	3	04.0	
02	阿倍野大斎場	，セシＬＦＰ７ＰＴＢＩ９	U		3	.	
03	哀愁	２０ＷＦ	*	400	5	02.0	
04	ある	＋5オ.	*	400	5	03.0	
05	哀愁ある	２０ＷＦ＋5オ.	U		5	.	
06	哀歓	２０７６	*	1	6	01.0	
07	愛歓	２１７６	U		6	.	
08	愛国主義教養	２１ＡＧＦＡ７Ｇ８Ｉ７Ｇ	*	1	7	08.0	
09	愛国主義教養	２１ＡＧＦＡ７Ｇ８ＩＧＫ	U		7	.	
10	〔あた〕◆云会山	１シ＋..１／Ｉ）３１５ＡＣ７	*	1	9	02.0	
11	愛◆云会山	２１Ｉ）３１５ＡＣ７	U		9	.	
12	人員	２Ｔ２.	*	1	16	04.0	
13	機会	７Ｅ５Ａ	*	1	16	05.0	
14	握員	２２２.	U		16	.	
15	握機会	２２７Ｅ５Ａ	U		16	.	
16	あげ	＋５＋Ｖ	*	400	22	06.0	
17	あげら	＋５＋ＶオＧ	U		22	.	
18	アサ	＋４.０	*	1	24	08.0	
19	1005	カ１カ０カ０カ５	*	1	24	09.0	
20	アサ1005	＋４.０カ１カ０カ０カ５	U		24		

M／T　FORMAT

A	冊数番号	3ch.
	頁　情報	4
	語　番号	3
B	〔度　数	7
WORD	見出し語	
	漢テレ　C1	40＋VARi.
	仮　名　C2	VARi.
	漢テレ　C3	VARi.
	ル　ビ　C4	VARi.
D1	配列情報	5
	付加情報　I	
	iNDEX	3
	付加情報個数	1
	語種名	1 ｝×5
	品詞名	1 ｝×5
	活用形名	2 ｝×5
	付加情報　II	
	位置情報	1

A	B	C1. C2. C3. C4.	D1	D2	D3		Dn	E／I

WORDの内容　　△→SPACE

●→WORD　MARK

C1 →二億七千八十八万三千三百三十五株△～△●

C2 →に△おく△ななせん△ひゃく△はちじゅう△はち△まん△さんぜん△さんびゃく△さんじゅう△ご△かぶ△●

C3 →二△億△七千△百△八十△八△万△三千△三百△三十△五△株△●

C4 →二〔に〕△〔おく〕△七〔なな〕千〔せん〕△百〔ひゃく〕△八〔はち〕十〔じゅ〕△八〔はち〕△万〔まん〕△三〔さん〕千〔ぜん〕△三〔さん〕百〔びゃく〕△三〔さん〕十〔じゅう〕△五〔ご〕△株〔かぶ〕△●