An automatic checking system by Kanzi-teletypewriter of input data

漢テレ・入力データのチェック

木　村　　繁

§0. システムの設計と入力データ・チェック

§0.0 Computer による情報処理システムの設計

Computer は，入口から送り込まれた情報を読み，その情報に，大量の記憶装置を使い，高速度の算術演算および判断操作を行うことによって，計算・分類・翻訳のような処理を施して，出口へ結果を出す情報処理機械である。

Computer はその特性のためしばしば人工頭脳などといわれているが，人間と異なり，意志や自発的判断にもとづいて自発的にものを実行することも，人間のことばを直接理解することもできない。そのため，機械の実行できる単純な細かい操作に分解した一連の処理手順(プログラム)，また処理すべき情報(データ)を読み込み可能な bit 情報の集まりとして与えてやらなければ動かすことができない。

多量のデータを，このような Computer によって情報処理を行なうためには，次のようなシステム設計を行なう必要がある。

i） データの収集と資料化

ii） Computer の読み込み可能な input データの作製

iii） Computer による情報処理，その結果として人間の読解できる形での Output

iv） Output の利用，分析。

計算機システムにおいては，ソフト・ウエア，ハード・ウエアが完備されていないと有効に働かないと言われている。ここで，システムという角度から見るならば，ソフト・ウエアは単にプログラム作りだけではなく，情報処理に伴う各作業ブロック間の人と情報の流れを組織体の経営としてとらえる時に生ずる各種の管理・運営の技術（たとえば，品質管理，日程計画，人員配置，コストと効率の問題 etc.）などを含めて考えねばならない。

このように処理過程をシステム(EDPS)として把握するとともに，与えられた制約条件のもとにシステム全体としての目的とする効率を高めるように処理過程が設計され運営される必要がある。

§0. 1 入力データ・チェックの必要性

一般に EDPS の効率を考えるとき，情報の正確さと処理の迅速さが問題になる。Computer のみならず機械にデジタル的な情報を処理させる場合，誤動作があっては困る。特に Computer では時間の持つ経済的な意味は重大であるから，誤りは直ちに検出され，短時間に処置できるように，コード自体にチェックの機能をもたせたり，Computer の回路にチェック機構をもたせている。

EDPS における速度に関しての最大のネックは入出力装置である。単にオン・ラインとしてばかりではなく，オフ・ラインとしてのデータ作成に非常に時間を要する。この遅い機器によって作られたデータを有効に処理するためには，次のことが満たされるようにシステム設計を行なわなければならない。

（1） まずエラーを早く検出し，エラーの箇所をはっきりさせること。

（2） 一部に誤りがあっても，システムのダイナミックな作業進行の過程において修正可能であって，全体として正しく効率のよい処理ができること。

（3） エラーの原因を追求し，発生源を押さえ防止できること。たとえば漢テレを保守するとか，作業者の慣習的エラーをなおすとかする。このためには，フィード・バックが有効に働くような工程間の連絡・作業日程の設計が必要である。

入力データチェックは，エラーを早期発見し，その影響を比較的狭い範囲の波及にとどめるばかりではなく，ソース・データ入力以降の計算機処理においてやたらとチェックを設けて処理手順を複雑にしないためにも，厳重に行なう必要がある。厳重といっても，システムの効率における要件となる速度と精度の2面を考慮すると，誤りが何らかの原因で生じた場合，それを見のがすチャンスが確率的に実用上無視でき，システムの処理結果としても必要な精度を満たし，かつ経済性に裏付けられる程度と考える。

また，単に設計時に予知できたり，想定したエラーばかりではなく，運用時における新事態の発生に対しても，エラー対策の検出・訂正・追加・削除等がフレキシブルに可能なシステム設計が必要であろう。

§0. 2　入力データのエラーの原因と対策

Computer による語彙調査の第1次計画（我々は急行コースと呼んでいる）は，図1のようなゼネラル・フローのもとで行なっている。以下は，この調査において，漢テレ，入力データのチェックを行なってきたことについての報告である。

漢テレ・入力データのエラーとしては次のようなものがある。

（1）　機械的工学的理由によって生ずる誤印字，誤穿孔〔漢テレ・エラー〕

（2）　穿孔する人が違う文字を打鍵したエラー〔ミス・タッチ〕

（3）　原稿に記入された内容が原資料と違っていたり〔転写エラー〕，プリエディット（前編集）で単位切りや出典・層別などの情報を付加する時ルール違反があって，入力データとしての様式に合っていないエラー〔フォーマット・エラー〕

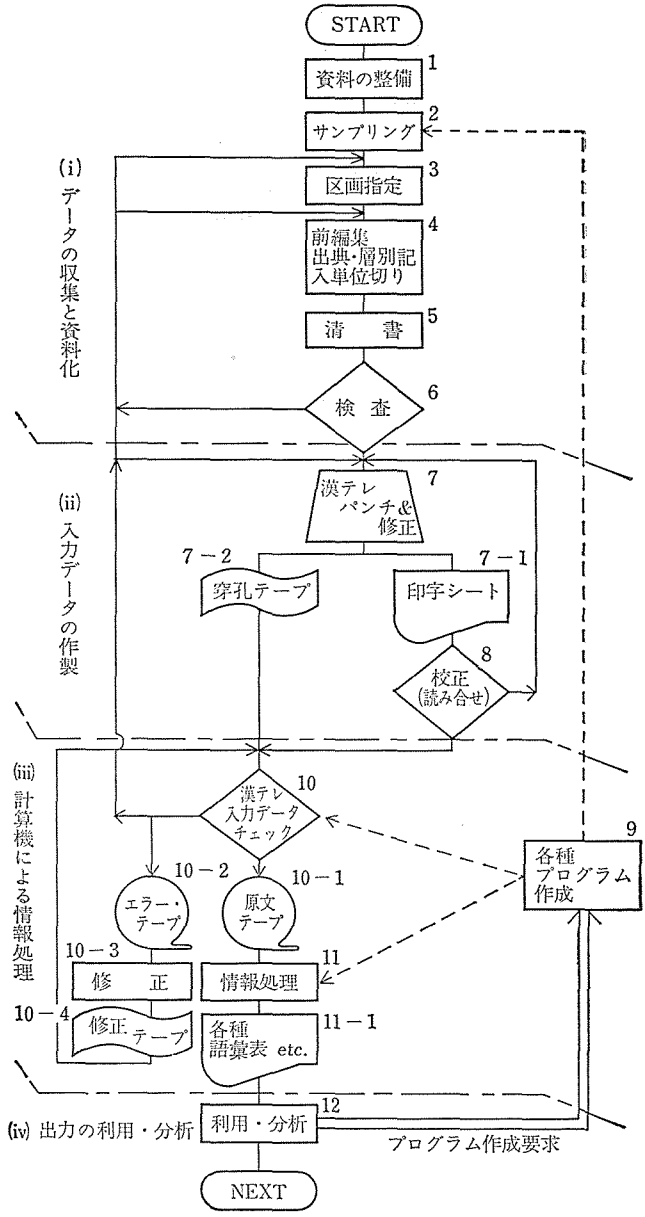
（4）　システムにおける流れとして，必要かつ十分な情報が過不足なく処理されたかという，システム管理的観点からのチェックによって発見されるエラー〔たとえば，ブロック・エラー〕

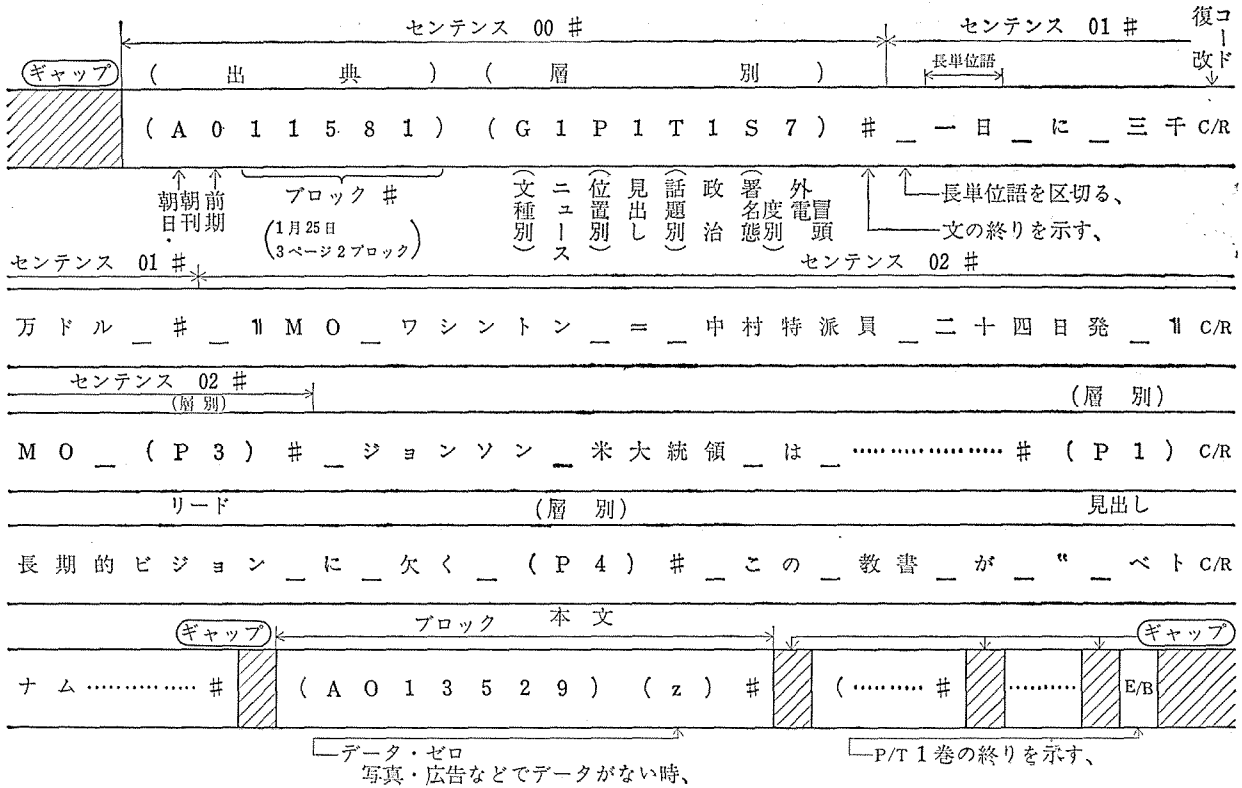
このうち，（3）の転写エラー，フォーマット・エラーは検査（図1の6）において原稿の上でチェックされる。原稿にもとづいて漢テレ穿孔した時に，同時にできる印字シートと，その原稿との校正（読み合せ検査）（図1の8）においては，上のエラーのうち（2）のミス・タッチと（1）の誤印字をチェックする。

ここで問題となるのは，印字が正しかったが穿孔が間違ったというようなマシン・ミスが，校正では検出できないということである。この対策としては，別の漢テレで再印字させてから校正を行なう方法も考えられる。

また，カード検孔機を使っての検査と同じような原理で紙テープの検孔も考えられる。しかし，現在は漢テレの検孔装置ができていない。また，仮にできているとしても，漢テレにはキーが多いので打鍵のスピードの点でカード検孔機の効用と同一に考えられないし，単に原稿との一致だけではなく検査（図1の6）の作業内容までも含めて校正（図1の8）を行なえることから，検孔装置によるチェックがこの場合必ずしも適しているとはいえないかも知れない。

〔図 1〕 語彙調査(急行コース)ゼネラル・フロー




センテンス 00 #
センテンス 01 #
復コード改
ギャップ
（出典）（層別）
長単位語
（A 0 1 1 5 8 1）（G 1 P 1 T 1 S 7）# 一日 に 三千 C/R
朝日・朝刊
前期
ブロック #
（1月25日 3ページ2ブロック）
（文種別）ニュース
（位置別）
（見出し）
（話題別）政治
（署名度別）外電
冒頭
長単位語を区切る、
文の終りを示す、
センテンス 01 #
センテンス 02 #
万ドル # ¶ MO ワシントン = 中村特派員 二十四日発 ¶ C/R
センテンス 02 #
（層別）
（層別）
MO （P 3）# ジョンソン 米大統領 は ……# （P 1）C/R
リード
（層別）
見出し
長期的ビジョン に 欠く （P 4）# この 教書 が " ベト C/R
ギャップ
ブロック 本文
ギャップ
ナム ……# （A 0 1 3 5 2 9）（z）#
（……#
……
E/B
データ・ゼロ
写真・広告などでデータがない時、
P/T 1巻の終りを示す、

漢テレ印字シートの校正がすむと，その印字シートに対応する穿孔テープを漢テレの読み取り部にかけて，修正しながら再穿孔・印字する。この時，読み取り部におけるメカニカルなチェック機構によって，パリティ・チェック，XYチェックがなされる。(詳細は，松本論文の附6—3を参照—P. 79)

以上述べたチェックの方法は人の判断力に頼る検査・校正などや，オフ・ラインとしての漢テレのメカニカルなチェック機構によるものである。漢テレ・入力データの Computer によるチェック方法については次の章で述べる。

§1. 語彙調査における入力データのチェック

まず，入力データのフォーマットを例文で示し，次に概括的にチェックの種類を，そして個々のチェックの方法を述べる。

§1.0 データのフォーマット

A. 例文……前のページ(P. 139)参照

B. データの作成の仕方

B 1. 出典……原文ブロックの先頭にある( )内の漢テレ7字で，その順序に従って，次の意味を示す。

1) 第1字……調査対象の新聞3紙・朝夕刊別を示す。

2) 第2字……0は前半(1～6月)，1は後半(7～12月)を示す。

3) 第3～7字……この5字は紙面1／2段を抽出単位としてサンプリングされたブロック#を示す。〔注〕第1月の1日(前半なら1月1日，後半なら7月1日)の第1ページ第1ブロックを00000とし，第6月の31日，16P・第30ブロックを89279とするシーケンシャルなナンバーをブロック#と呼ぶ。抽出比1／60で1紙半年分の紙面から，朝刊1440，夕刊930ブロックを抽出。

B 2. 層別……G(文種別)，P(位置別)，S(署名態度別)，T(話題別)の4種の角度から記事を見分け，層別情報とする。Gは17，Pは8，Sは10，Tは12に分かれている。G≧14(小説，広告，漫画)の時は，P＝8，S＝10，T＝12(その他)とし，G以外のP，S，Tは書いても書かなくともよい。出典以降に現われる( )内で示し，変更ある層別のみ示す。

B3．データ・ゼロ……区画指定された紙面が写真・広告などでデータがない時，(Z)を出典の次に示す。

B4．センテンス……新聞上の文と計算機処理のため前編集された文とを区別するために，後者をセンテンスということにする。たとえば，新聞の見出し，株式欄などでは，センテンスが句点で終わるとは限らないので，1センテンスの終わりを#で示す。原文ブロック先頭の（出典）（層別）#をセンテンス・ナンバー00#とし，以下順に1ずつ増す。

B5．長単位……語彙調査・第1次計画(急行コース)で採った調査単位語。

C．計算機処理の立場からみたデータの構成

紙テープ(P／T)は数個の原文ブロックとE／Bブロックからなる。

C1．原文ブロック……最初が（，最後が#で，ギャップとギャップの間にあるセンテンスの集まり。

（1） 区切り符号……(，)，#，㊎の4種の区切り符号により次の(2)～(5)が弁別される。

（2） センテンス　（3） 出典　（4） 層別　（5） 長単位

（6） C/R……漢テレの自動復改機構のセットまたは復改キーの打鍵によって，H—3010コードでは＊Jの2桁がP/Tに穿孔され，印字シート上では改行復帰が動作される。後に述べる漢テレチェック・コードとして用いる。たとえ長単位の途中であっても原文ブロックの任意の所に現われてもよく，正しいデータとして入力される時はサプレス(削除)される。

C2．停止コード……紙テープの掛け換えを行なうための読み込み停止コードとしてE/B (H—3010コードでは＝E/B)を用いる。

〔注〕 P/Tのラベル……ラベルは紙テープには穿孔せず，紙テープの掛け換えの操作ミスを防ぐためには，所定の書式で書かれた紙のラベルを視覚的に確認しマニュアルでレジスターにその情報を入れる方法を採った。

§ 1.1 チェックの種類

今回の語彙調査においては，漢テレ入力データについて次の3つの観点からComputerによるチェックを行なっている。

i） 漢テレチェック……C/R (H—3010コードでは＊J)をチェックコードとして紙テープ上の機械的なパンチ・エラー〔参照§0.2の(1)〕を検

出する。

ii) フォーマット・チェック……プリ・エディット（前編集）されたデータの区切り符号を拠りどころにして，出典・層別・度数・長単位からなるレコード作成上必要なデータ様式（フォーマット）を満たしているかを主としてチェックする。〔参照§0.2の(3)〕

iii) ブロック・チェック……処理に必要な原文ブロックが過不足なく入力されたかをチェックする。〔参照§0.2の(4)〕

i)，ii)のチェックは個々の原文ブロックに関してのいわばスタティックな観点からのチェックであるが，iii）は語彙調査・システムを動かしていく時ダイナミックな観点から必要なものである。そこで，i)，ii)は§1の以下の節で，iii)は次の章§2で述べることにする。

§ 1. 2 漢テレ・チェック

A．エラーの種類

システム・デザインの当初は，それまでの漢テレ・エラーの知識より，次の2つのエラーを想定した。

（1） 漢テレ1字を表わすX，Y軸のH—3010コード2桁のうち，1方がオール・マーク（紙テープでの抹消コード）に化する。

（2） XY2桁のうち1方がオール・スペース（スプロケット）に化する。

この2種のエラーは，主として，連続したミス・タッチを訂正するために，ファンクション・キー〔松本論文附5—2(4)を参照—P．78〕の後退キー，抹消キーを連続打鍵したときに発生し易いようだ。

その後，漢テレ・パンチ操作に慣れるにつれ，シフト・ペダル〔松本論文§2，—P．62以降を参照〕と打鍵のタイミングずれによると思われるパリティエラーが発生した。（これは，漢テレに回路を追加してメカ的に改善された。但し，ミス・タッチとして穿孔印字される。）また，たとえば紙テープの端をとめるゼム・クリップの傷などで，紙テープの一部が読み込み不能となることがある。これらの機械的あるいは人為的な取り扱いミスを一括して，

（3） リード・エラー　として，計算機のオペレーションにおいて迅速に対処できるようにプログラムに組み入れる必要がある。

B．漢テレ・チェック項目とその方法

1. 偶数チェック……漢テレ1字はH—3010コード2桁で表わされるから，漢テレによるデータが偶数桁読み込まれたかをチェックする。

2. XYチェック……漢テレ・読み取り部のXYチェック(松本論文附6—3及び§ 2，注3—P．67参照) と同じ考えに立つもので，X軸にJ符号があればエラーとする。(X軸＝第1列のコードとしてJ＝$(41)_8$が使われていない。)このため，自動復改をセットして漢テレ穿孔を行ない，C/R(＊J)によって，漢テレシート1行(漢テレ35字／行)以内にJ符号が少なくとも1回(正常ならばY軸に)現われるようにしている。次のJ符号が現われるまでに1桁脱落のエラーが偶数個発生しないかぎり，この方法で，漢テレ・エラーの種類(1)，(2)をチェックできる。

3. LAST＃……スプロケットが，データ・テープに飛び込んで，2ブロックに分かれて入力されるのを防ぐために，原文ブロックの最後が＃(但しその後にC/Rは何個あってもよい)であるかをチェックする。

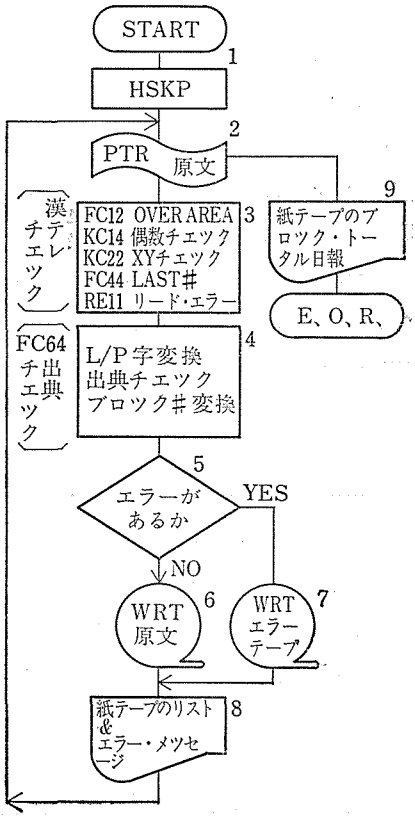
4. リード・エラー……穿孔テープのパリティ・エラーあるいはテープのいたみのため，読み込み不能でリード・エラー(REのランプが計算機の操作卓について計算機が停止する) が発生した時，紙テープを1桁進めるか，読み込み可能な所をセットするかして特定のアドレスから再スタートする。そしてそのルーチンでは，それまで読み込まれたセンテンス内の偶・奇数を調べ，データが偶数桁になるように，その後に3桁のH—3010コードのP＊＊あるいは＊＊Pを埋め，偶数チェックとは独立の状態にして，読み込みを続行する。H—3010コードの＊＊は，漢テレ字で㊊と印字され，その位置を見つけ易い。

なお，漢テレ・チェックのプログラムでは，上記の外に，データの長さが読み込みエリアの範囲内(2000漢テレ字)か，出典がブロックの先頭にあってフォーマットを満たしているかをチェックする。(〔図 2〕を参照)

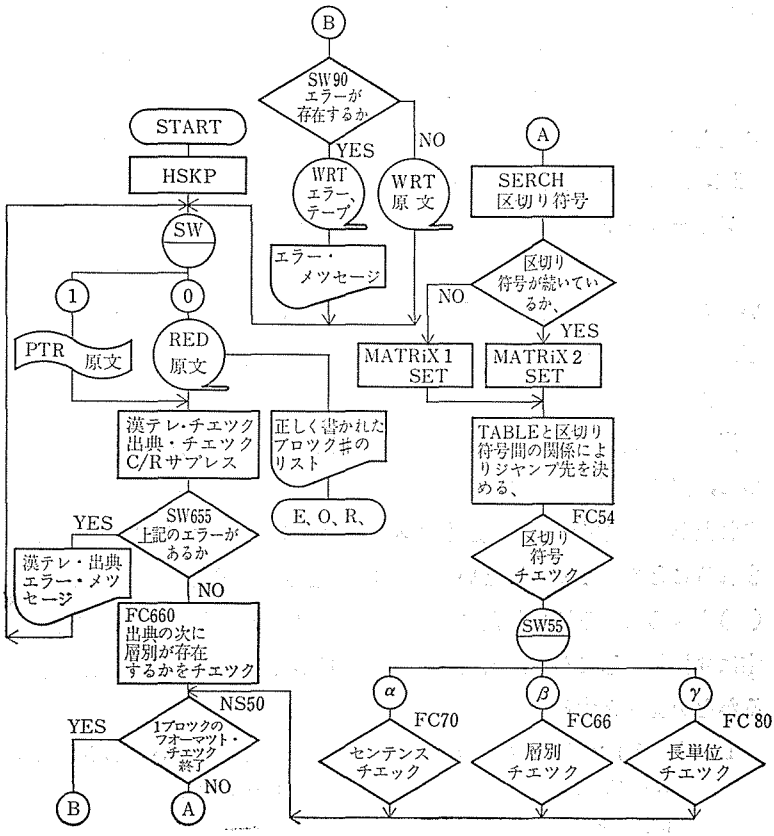
リード・エリア，LAST＃，出典に関しては，ブロック全体として；偶数チェック，XYチェック，リード・エラーは，センテンス毎にチェックする。

多量のデータ処理におけるエラの表示方法としては，紙テープ1巻毎にP/Tネーム付きの全ブロック・リストを作成し，エラーのあるブロックには

〔図 2〕 漢テレ・チェックの概略フロー



〔図 3〕 フォーマット・チェックの概略フロー



その項目と所在としてのセンテンス#をプリントする。(たとえば，XY エラーは KANTEL 01#，リード・エラーはRE 05R，LAST #のエラーはLAST #11L) なお，漢テレ・出典チェックに関して正しい原文ブロックは，マニュアルで入力したラベル (P/T ラベルの項・参照) をファイルネームとしてM/T (磁気テープ)に書かれる。

§1. 3 フォーマット・チェック

フォーマット・チェックは前編集で区切り符号を用いて附加された出典・層別情報，単位切りがデータの作成の仕方あるいは構成のルールに合致しているか否かをチェックするものである。まず，〔図3〕のフローに従い処理の手がかりを主として述べ，次にチェック項目の概略を記す。〔§1. 0参照〕

A. 処理の手がかり

出典は原文ブロックの先頭にある( )内の7漢テレ字で固定的に示されるので，出典に関してのチェックは紙テープの原文ブロックが先頭から読み込まれたか(原文ブロックの途中にスプロケットが入って2ブロックに読み込まれることもあるので) という漢テレ・チェックの意味も含まれる。そこで，( )も含めて読み込まれたブロック先頭の9漢テレ字(18桁)を9桁のL/P字に変換(注1) してから，出典情報の各項目がその値のとりえる存在範囲にあるかをチェックする。

〔注1〕 ラインプリンター (L/P)で印字可能な48字種については，X軸に'の弁別コード，Y軸には3010コードと一致するコードを当てて漢テレ字をL/P字に交換可能になるよう漢テレが設計されている。〔松本論文§2の基本的要求(5)—P. 63参照〕

次に，層別・センテス・長単位に関しては，原文ブロックの出典情報以降において，各情報の種類を弁別している区切り符号を前から順にさがし，前後2つの区切り符号の組み合せからその時のチェックに必要な処理を決める。その，符号の組み合とせ処理とを示した表を"処置行列"(注2) という。

〔注2〕 参照：「ALGOL入門」 森口繁一編
第3部 ALGOL処理のプログラム

語彙調査における原文のフォーマット・チェックでは，区切り符号間に盤内コードがある(長単位 or 層別が2つの区切り符号の間に存在する。〔図3〕ではMATRIX 1をSETしている) 時と，区切り符号が続いている(〔図3〕

ではMATRIX 2をSET)時との，2種類の処理行列がある。

〔MATRIX 1〕 区切り符号間に盤内コードがある時

SIGN 2 / SIGN 1	(	)	＃	㊊
(		FC 66		
)	FC741		FC 74	
＃	FC701		FC 70	
㊊	FC801		FC 80	

：区切り符号エラー（この符号の組み合せはありえない）

NS 50：㊊＃の時は次の区切り符号をさがす。

〔MATRIX 2〕区切り符号が続いている時

	(	)	＃	㊊
(	ED 58			
)		ED 58	FC742	FC743
＃	FC702		ED 58	FC703
㊊	FC744		NS 50	ED 58

ED 58 ： 同じ種類の符号が2つ続いて穿孔された時，1つをサプレス(削除)する。

FC66：層別に関するチェックを行なう。

FC70～FC703，FC74～FC743：センテンスに関するチェックを行なう。

FC80～FC801：長単位に関するチェックを行なう。

なお，上記のセンテンス及び長単位に関しての処理内容が区切り符号の組み合せによって異なるのは，次の理由による。次の，B．フォーマット・チェック項目 で述べる＊FC65，＊FC72，＊FC74の判別においては，単に2つの区切符号の組み合せだけでなく，それ以前もしくはそれ以降の区切り符号間の処理に影響されたり影響を及ぼすからである。

B．フォーマット・チェックの処理項目

FC64 出典：常にブロックの先頭の（ ）内にあって，しかも字数がFix で順序が決まっている。

FC 640 左かっこ （ がブロックの先頭か？…… スプロケットによ

る原文ブロックの2分割を防ぐ。

FC 641　新聞名のコード(A，B，C，J，K，L)か？

FC 642　前・後半を示す0or1か？

FC 643　00000≦ブロック#≦89279か？

FC 644　ブロック#の内に英字などの数字以外の字種が混ってないか？

FC 645　右かっこ　)　か？

FC66　層別：出典の次には層別が必ず存在し，それ以降では変更ある毎に，字数 VARIABLE で表わす。ただし，文の途中で層別が変わることはない。

FC 660　出典の次に層別が存在するか？

＊FC 65　1つの(層別)情報の後に少なくとも1センテンスが存在したか？

FC 661　層別種を示す英字がG，P，S，Tか？　なお，データ・ゼロ(Z)の時は，それ以降に長単位が存在しないとみなされそのブロックに関してのチェックを終える。そこで，有効な原文の後に(Z)を付け，それ以降に覚え書きを書いてもよい。ただし，漢テレ・チェック項目　LAST#　の要請により，原文ブロックの終りは#でなければならない。

FC 665，FC 67　層別の数字に関してのチェック

FC 68　層別4種相互間でとりえる数字の範囲に関してのチェック

FC70　センテンス

＊FC 72　長単位の存在：#と#の間(すなわち1センテンス内)に長単位が少なくとも1つ存在したか？　ただし，2千漢テレ字を越える原文ブロックは，2ブロックに分割するため，文頭から連続したセンテンス#のから送りは認める。

＊FC 74　層別と#：(層別)の前後いずれか1方に#が存在するか？

FC80　長単位：長単位はレコードのアイテム(項目)としては，20漢テレ字以内では後に㊎をうめて20漢テレ字固定長，20漢

テレ字を越える場合は可変長(max 40 漢テレ字) で表わす。

FC 84　長単位の字数が 40 漢テレ字以内かをチェック。21～40 漢テレ字の時は，L／P で REMARK をプリントする。

§2　データーの流れに関するチェック

§2.0　ブロック管理

〔図1〕のフローで示したように，原文が計算機に入力される以前にも幾つかの工程＝人手と日程＝を経ている。また，計算機による入力チェックでも，漢テレチェック，フォーマットチェックの2段階で，エラー・データと正しいデータとに別けて書かれる。エラー・データは修正され，結局は正しいデータとして，1紙半年分で朝刊1440，夕刊930の原文ブロックが書き込まれなければならない。

ここに，単に個々のデータの良し悪しのチェックだけではなく，トータル的なデータの流れという観点からの，いわば管理的チェックの必要性がある。各工程毎の員数チェック，次の工程への受け渡し，作業進行にともなって生じる各記録媒体の保管管理の問題などこれに属する。

その方法として，作業内容を数量的に的確に表現する“指標”で示す計数的管理が考えられる。たとえば，語彙調査では，原文の入力に関してはラン毎の処理ブロック数，長単位に関しては延べ語数・異なり語数などその指標となろう。このうち，延べ語数と異なり語数との間には関数的な相互関係が過去の語彙調査によって得られているので(注)，その相互関係から大きく逸脱した場合などには処理上にエラーがないか調べなおしてみる必要があるだろう。“指標”は数量的にエラーを検出するばかりではなく，作業の進行度，ラン毎のデータの受け渡し，適正な処理量，処理の標準時間の算定及び日程計画，L／P 用紙・漢テレ用紙・磁気テープ・紙テープ・キャビネットなどの消耗品及び備品の必要量，などの管理資料としても意味をもつ。

注　実は正確にいうと，従来のは同語異語の判別されたものについて関数的相互関係が得られているが，今回の急行コースではこの点での操作をへていない点で問題がある。また単位の長さの点でも問題がある。

§2.1 おわりに

システム全体としてエラーは考える時,§1.0で述べたように,(1)エラーの検出 (2)エラーの修正 (3)フィード・バッグ体制 にふれなければならない。この報告では,(1)を主として述べたが,入力エラーを考える時この3つの項が組織的になされなければならないことを最後に強調しておく。