
 寄 書

漢字テレタイプライタの使用経験*

西 村 恕 彦**

1. はじめに

さきの情報処理 1969 年 9 月号は、漢字情報処理の特集号であったが、これを読んでみると、われわれの経験から二、三いっておいたほうがよいと思われることもあるので、ここにそれをまとめてみた。

われわれの研究室では、自然言語のいろいろな処理を容易にするために漢字テレタイプライタを導入してそれ以来 4 年間の使用経験を積んできた。漢字テレタイプライタは当時すでに相当な台数が新聞社などで実用されており、一般的な技術面ではとくに問題とすべき点はなかったが、ただ計算機にたいするオフラインの入出力装置としての使用例はほとんどないようであった。その面から既製の仕様を検討した結果、多少の手直しが必要と思われた。

ここでは手直しのためにどのような考慮を払ったかということ、それが経験からみてどのくらい妥当していたかということとを批判的に述べよう。

2. 符 号 系

穿孔符号系については、一般に符号変換を計算機のプログラムで行なえばよいから、かなり任意のものでよいと称されている。しかし符号変換の手続きを、コボルやフォートランのような言語で行なうことはやっかいであるうえに、ソーティングなどの常用ルーチンでは符号変換の機能はないのが普通である。さらに、自然言語処理においては処理待ち (process limited) になることが多く***、入出力待ちにはなりにくい。

* The Experience with a Kanji-Typewriter, by Hirohiko Nisimura (ETL, MITI)

** 電気試験所・電子計算機部・情報基礎研究室

*** 紙カード/印刷機/紙テープなどのデータ移送速度は、1ms/字ほどである。自然語の機械翻訳の実績は 10ms/字がかなり速いほうである。KWIC 索引の作成では、1~10ms/字である。フォートランのような簡単な言語の、かなり高速なコンパイラでも翻訳速度は 1ms/字に及ばないものが多い。

磁気テープの移送速度は、適度にブロック化されていれば、20 μ s/字ほどである。IOCS のオーバヘッドとして、CPU 時間が 2~10 μ s/字ほどとられる。10進2進変換やコード変換のような簡単な処理でさえ、10~20 μ s/字で実行できる金物をもった計算機システムはまれである。

したがって、遅い処理のうえに符号変換までやったのでは、まったくやりきれない。デバッグやダンプも格段にめんどうになる。

これらのことを考えて、穿孔符号系は符号変換を必ずしも用いなくてもすむように、できるだけの考慮をした。穿孔符号は 6 ビット、64 種類が可能であるが、漢字テレタイプライタでは第 1 列にたいしては 48 種類、第 2 列にたいしては 52 種類をえらんで使うようになっている。

符号変換をしないですますためには、この 64 種類中の約 50 種類を、計算機で通用している符号系と良く整合するようにえらばなければならない。符号系の基準として、アイビーエムカードの 48 字系 (H セット) をとった。

この考慮は成功したようで、約 2500 種類の漢字テレタイプライタの文字は、ほとんどが穿孔カード符号や行印刷機の標準の 2 文字の組み合わせで表現でき、デバッグやダンプ、あるいはカードからの入力、プログラム中の定数としての利用などに有利であった。

しかしアイビーエムの 48 字系 (H セット) には大きな欠点がもともとあった。それは 6 ビット符号が 2 進数的にいて連続してとられていないことである。8 進数で表現すると、01~11, 21~31, 41~51, 62~71 にたいして通常の英字と数字; 00, 13, 14, 20, 33, 34, 40, 53, …… などにたいして記号類が割り当てられていて、両者の関係が組織的・系統的でなく、さらに、符号そのものも中間に欠番がある。

アイビーエムの 48 字系 (H セット) を基準にとったことは、カードや行印刷機など、他の入出力媒体との関係では操作性に富んで有利であったが、反面、欠番があるために、表引き、判定、検査などのプログラムが不利になった。

紙テープ上の穿孔符号は、2 列が 1 組になってはじめて意味がある。紙テープを 1 列ずつ読んでゆくときに、けたずれがおこらないようにしなければならない。これをどういうやりかたで保証するかは、むづか

しい問題である。JICST では、紙テープ符号の第 7、第 8 単位目を利用して好成績をおさめた (大場⁸⁾ 336 ページ)。

われわれは特定の文字符号についてこれを検出することにした。この符号として、復改符号と間隔符号とを考えて、けっきょく間隔符号についてだけ、この検査を行なうことにした。これはたいていの場合に有効であるが、まれに間隔 (空白) をまったくふくまない文章があると、お手上げである。

なお、紙テープの第 1 列目に、送り符号 (オールスペース) または抹消符号 (オールマーク) が現われると、紙テープは 1 列ずつ送られ、有効な符号がきたところで 2 列ずつ読まれるようになる。

3. 文字の組

文字としてどういうものをふくめるかということは総字数が約 2500 字とおさえられると、ほとんど決まってしまうだろう。われわれの漢字テレタイプライタでは、最初の四半分のブロックに、機械符号、記号、ギリシャ文字 (大小)、数字、英字 (大小)、ロシア文字 (大小)、ひらがな (大小清濁)、かたかな (大小清濁)、図形用記号などをまとめ、残りの四分の三に漢字 1834 字をおさめた。

この 1834 字のなかには、伏せ字記号 (ゲタ)、踊り字 (々)、丸印の零 (〇) が入っている。

漢字の字種の選定にあたっては、全面的に国語研報告¹⁾を参照した。実際の文字の採択にあたっては、もちろん非常にこまかい考慮を払ったが、基本的な態度は高橋⁷⁾の考察 (308 ページ) とほとんど同様である。ただし使用目的は、(1) 科学技術文献の機械翻訳、(2) 言語情報処理の研究、(3) 一般的な言語処理、などにおいたので、えらばれた字そのものはいくらか出入りがある。

多くの漢字テレタイプライタでは、標準の漢字の組に入っていない漢字を、外字、あるいは盤外字などと呼んで、特別な符号を与えたり、2 字以上の符号を組み合わせて表現したりする方法をとっているようである (たとえば、長谷川⁶⁾ 282 ページ)。

われわれの漢字テレタイプライタでも、符号に多少の空きがあるので、そこへ標準以外の漢字を一時的におさめることも考えてみた。しかし、漢字の配列性 (次項) を重視すると、とうていそのような配慮を認める余地がなく、けっきょく標準以外の漢字は伏せ字記号にするか、あるいはかななどに書き替えるしかなくな

った。

アルファベットとして、英字、ロシア文字、ギリシャ文字を入れた。これらのなかには、A, B, M, O など、同じ字形で表現してかまわない文字がふくまれている³⁾。しかしそれらの文字の符号を同じもので代用すると、植字機としてはさしつかえないが、計算機の入出力としては、たいへんめんどうになることが、計算機ヤマトでの経験からわかっていった。

それで、これらの文字もそれぞれ独立した文字として、それぞれのアルファベットのなかの、しかるべき位置におさめた。字形の類似した文字が紛れないように、字形のデザインに格段の考慮を払った。

すべての文字は一樣なピッチでしか印字されないの、ある種の記号 (括弧や点の類) は内側に寄せて位置づけるようにデザインした。拗音や撥音のための小さいかなは、左下 (左横書き) に寄せてデザインした。しかし実際に文章を印刷させてみると、それらの記号や文字の右側 (記号によっては左側) があきすぎて見づらい。中央に位置づけてもよかったのではないかと考えている。

4. 文字と符号の対応

符号系が定まり、そこにふくめるべき文字の組が決まったら、次には、両者のあいだの関係をつけなければならない。そこで考慮した事項は、次のとおりである²⁾。

- (1) 打鍵の容易さ (とくに非熟練者)
- (2) ソーティングにおける整合性
- (3) 文字の類別の識別
- (4) 活用などの語形変化の処理

文字の鍵盤配列については、熟練した専任操作員にあってはどのような配列でも大差はないようである。しかし非熟練者がたまたま打鍵するときには、ある文字が鍵盤中にふくまれているか否か、もしふくまれているとしたらどこに位置しているかの二点を常に考えながら打鍵しなければならない。

一般に推奨されている、頻度別に文字をブロック分けし、各ブロックのなかでさらに音訓別などに配列する方法は、このような観点からは、はなはだ好ましくないものである。すなわちある漢字を次々に何回も各ブロック中で位置づけて探さなければならず、見落としの不安から解放されない。

そこで鍵盤配列は厳密に一意の基準に従ってならべること重点をおいた。すなわち、音標文字や記号類

は一つのブロックにまとめ、他の三ブロックには漢字を「音」によって配列した。この方針はいいして成功したようであるが、「音」がポピュラーではない漢字が少数個あること（たとえば、筈、匂、石、峠、杉など）、および傍（つくり）の同じ漢字が同音であるためにある種の誤りが発生しやすいことなどが気の付いた短所である。

漢字テレタイプライタにおいては、鍵盤配列と符号系とは密接に関係している。それで、上記のような鍵盤配列で得られた文字符号系は、そのままソーティングしてもきれいな配列が得られる（植村⁵⁾ 275ページ）。

ソーティングに強い考慮を払って純粋に単一の原則によって漢字の配列をきめたのはわれわれの機械と、それをそのまま踏襲したと考えられる東京電機大学のもの³⁾とだけのものであるが、計算機処理のことを考えると、もっと広く採用されてよい。

JICST の漢字テレタイプライタも、相当に配列性を考慮してあるようだが、下位 6 ビットが大分類、上位 6 ビットが小分類になるような符号系であって⁸⁾、なぜそのような符号系を採用したのか、まったく理解に苦しむ。

日本語のかなについては、活用語尾の生成や解析（機械翻訳用）を簡単に行なえるようにしたかったので配列性の原則は大きくくずれている。かなは種類が少ないから、符号変換を計算機でやっても大したことはなかろう、という意見もありうるが、基準となる符号系は、使用目的にあうようにしておくべきであろう。

われわれのところでは機械翻訳のことを考えたので 1 つの行に属する 5 文字（たとえば、カキケコ）が一つながりの符号になるようにし、ある文字の属する行や段が簡単に計算機で識別できるようにした。

筆者は以前の論文⁴⁾で、国語辞典式の配列を計算機で実現するための具体的な方法を示した。しかし、ひるがえって考えてみると、国語辞典式の配列の歴史はわずかに数十年をさかのぼるにすぎないし、いまここで、計算機操作が容易で、しかも合理的な点多々ある、計算機式の配列を実用化すれば、案外にそれで受け入れられるのかもしれない、と思わないでもない。

たとえば高橋⁷⁾ (310 ページ) は、「漢字と発音とコードを一意的に固定し、機械的にソートする……字順を一意的に定め、すべてその字順に配列する」ことを提案している。

このような目的にわれわれの漢字配列がよく適合していることは、植村⁵⁾の例 (275, 276 ページ) に見ら

れるとおりである。情報検索などのことを考えると、かなの配列もまた、しかるべき順序になっていたほうがよかった。

5. 機能符号

紙テープを読み取って複写穿孔するときには、あらゆる穿孔符号がそのまま複製される。しかし読み取って印刷させるときに、上記のように割り当てられた以外の符号が入ってくるとどうなるかは、仕様書には記載してなかった。納入後に試験してみたところ、ある種の符号については印刷機が機械的にロックされてしまい、そのリリースがかなりめんどろなことがわかった。またある種の符号については、それが他の有効な符号に変換されて、その有効な符号に相当する、かってな文字が印刷される。

いろいろな誤りや事故にたいする操作性の点からはこのような任意性を残すことなく、あらゆる符号における動作をすべて仕様書に定めておくべきであった。誤り符号を印刷機が検出して、読取り動作を停止することは、いちがいに悪いとはきめつけられないが、少なくとも、そのリリースの手当てはしておくべきだった。

なお、これに関連して、電源を切って再投入したときに、前の状態がそのまま保存されることは、利点よりも欠点のほうが大きかった。操作性のことを考えると、電源投入によって、最後に読み取った符号、誤り符号のロック、信号断停止、ページ送りなどがすべて初期化されるようにしておくべきである。

復帰と改行とを別々の機能にすることは、計算機やマトの経験では、ほとんど利点がなかったので、復改符号 (NL) として一つにまとめた。

特別の印刷のことを考えて、印字後退（バックスペース）という機能符号を設けた。JEM-3800 漢字プリンタ^{6), 9)}の合成機能とは異なり、いつでも、何回でも後退できるようになっている。しかしわれわれのデータでは、印字のピッチが一樣なこともあって、ほとんど使われていないようである。

ただ、字ならびの悪化の一部が、重いプラテンの復帰 (1 行は 50 字) によっておこるので、復帰時に前進、後退を二、三回反復すると、これがある程度改善されることが経験され⁵⁾、ときどき使われている。

紙テープの読取りは、一部を訂正して複写穿孔するとか、印刷の体裁を多少変えるとかのために、一時停止する機能がのぞましい。たとえば以下のような条件

で停止させられるとよい。

- (1) 抹消符号の読取り
- (2) 復改符号の読取り
- (3) 一定字数の読取りまたは送出し
- (4) ある種の符号の読取り

いずれの場合も、その符号が穿孔部あるいは印字部に送り出される前に、読み取った時点で停止しなければ意味がない。現在、復改停止という切替えスイッチがあるが、これは符号を送り出して、穿孔や紙送りをしたあとで停止するようになっているので、使いがってがよい。

印刷や複写穿孔をしているときに、上記のような条件で一時停止させて、鍵盤からの符号の挿入やその他の手操作によって、紙テープを訂正したり、印刷体裁をととのえたりしたい。

なお、以下のような手操作の機能がある。

- 鍵盤穿孔時の紙テープもどしと紙テープ送り
- 読取り時の1字読取りと1字の空送り
- 印刷用紙の手動連続復改

6. おわりに

漢字テレタイプライタは、計算機の出力印刷装置としては、速度、性能価格比、印刷品質（字ずれもふく

めて）などの点で、まったく不満足なものであるが、オフライン機器としては、漢字を視認して直接に打鍵できる、漢字がそのまま印字されるという、比類のない利点をもっている。

計算機による言語情報処理に関連して、漢字テレタイプライタがさらに広く使われ、批判され、改良されることを希望する。

参考文献

- 1) 現代雑誌九十種の用語用字(第二分冊)漢字表, 国語研報告 22, 1963-8.
- 2) 西村恕彦: 電試漢テレ仕様案, 機械翻訳研究委員会, 1965-9.
- 3) 中野道夫, 大井尚一: 東京電機大学における日本文字入出力装置について, CL 研資 67-10, 1967-9.
- 4) 西村恕彦: 文字列の配列順序についての問題, 情報処理 10-1, 1969-1, 21-25.
- 5) 植村俊亮: 漢字かなまじり文KWIC索引, 情報処理 10-5, 1969-9, 270-278.
- 6) 長谷川実郎: 高速漢字プリンタ, 同上, 279-284.
- 7) 高橋達郎, 広田広三郎: 計算機による日本語の処理, 同上, 304-313.
- 8) 大場賢一: JICSTにおける日本語抄録の電算機処理, 同上, 333-339.