

談話室

ソクタイプによる漢字処理法

佐伯 功 介*

1. ま え が き

漢字かな文を機械で処理する方法は、近年、分速何千字何万字という、高速度のものが開発されつつあるが、こういう機械の入力はほとんどが漢字テレタイプ式のさん孔テープによっている。現在の漢テレさん孔速度は、せいぜい毎分 70 字だから釣合がとれない。

筆者は昨年、ソクタイプによる漢字入力法を発表したが、これはほんのアウトラインだったので、多数のお問い合わせをうけた。今回もページの制限があるから充分ではないが、少し詳しく述べる。

2. ソクタイプ速記の現状

ソクタイプというのは、現在全国の裁判所で使っている速記の機械である。このソクタイプは、もともと第 1 次大戦のベルサイユ会議などで使われていたフランスの発明品を、故田中館愛橘博士が日本に紹介したものであった。日本語を打つためのキー構成や記号、略語の組織を作って、今日の機械を完成させたのは、ソクタイプ研究所の川上 兎氏である。裁判所では昭和 25 年から速記官の養成を始め、20 年間で卒業生 1,000 人を越えた。

速記官の養成は 2 箇年の研修であるが、最初の半年で打ち方は全部覚える。しかし、この時点ではせいぜい毎分 70~80 語ぐらいの速さで、速記にはならない。

70 語とか 80 語とかいうのは *rômazi de kaita bunsyô* の語数である。漢字かなで書くと、統計的に 1 語が 2 字に当たる。80 語は 160 字である。生徒の力をはかるテストは、それぞれの速さで 5 分間読み上げ、誤り語数 2% 以内を合格としている。

打ち方をおぼえたあとは反復練習しかない。80 語に合格すれば 90 語、つぎは 100 語と 10 語ずつ速めて、1 年の終わりには 130~140 語に達するのが普通である。卒業資格は 170 語であるが、事実上全員 180

語以上、そのうち約半数のものが 200 語以上をとって出る。

いままでの最高は 235 語 (毎分 470 字) である。もうこの辺になると、人間がどんな速さで読めるか、試験官のほうがテストされているような形になる。

実務には 150~160 語ぐらい確実にとれば充分とされる (ラジオのニュース解説ぐらいの速さ)。瞬間的には 200 語以上もできるが、200 語で 10 秒も続くことは実際にはおこらない。

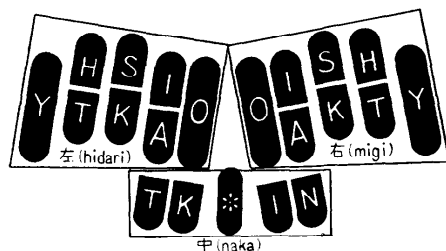


図 1 ソクタイプのキーボード

3. 機械の構造

機械は、21 の活字が一行に並び、それぞれ連結したキーを押すと、約 2 mm 突出してその場所で印字する。タイプライタとちがって、同時にいくつでもキーを押してよく、それが 1 行に並んで印字される。紙は 6 cm 幅で、60 m の長さのものをジグザグに折って内蔵してある。この 6 cm 幅の中に 21 文字が並ぶ。1 字打っても 20 字打っても、1 度打って手を離せば 1 行送る。

活字はそれぞれ個有位置で印字し、横には動かないから、つまり 21 孔のさん孔テープと同等である。ことばへの関連づけと、人間が目で見えて読むために文字になっているが、機械の中では黒丸でも孔でもよい。たとえば

* 裁判所書記官研修所講師、(社)日本ローマ字会理事
ソクタイプ研究所長

中の群の助詞	10%
両手略語の行	10%

この結果でみると、ある1つの両手略語の出現度は片手略語に比べてきわめて低い。略語は頭でおぼえただけでは役にたたない；反復練習ほとんど無意識に反射的に指に出なければならぬ。その訓練の時間と努力を *kihon* と片手略語にふりむけたほうが有利だという結論に達したので、10年ほど前に両手略語を大幅に減らしたのである。その結果は打数で2%弱の増となるが、総合的に速記者の成績はむしろ上っている。

5. 漢字入力への応用

5.1 漢字指定

以上 4. までで述べたのは、現在のソクタイプ速記のあらましを紹介したので、新しいことではない。この速記の打ち方を応用すれば普通の漢字かな文を機械に入力する速度を、現在の漢テレ式のさん孔機に比べて、5倍ないし7倍にあげることができる。その方法を紹介するのがこの論文の目的であって、以下にそのあらましを述べる。

速記記号は一種の音文字で、同音語は意味にかかわりなく同じ記号、反訳の際に前後から判断して文字を選んでいる。しかし、機械にはこの判断は困難であるから、適当な方法で漢字を区別する打ち方をすればこれを高速度漢字入力に利用することができる。幸い現在のソクタイプで中央の*印が、速記にはほとんど使われない遊びキーになっているので、これを漢字指定キーに利用できる。位置もあらかじめその目的で設計したかのように格好のところにある。

そこで、つぎのように決める。

- (a) *kihon* はすべてひらがなにす。
 (b) 略語で同音異語がなく、普通に漢字で書く語はそのまま漢字にする。たとえば

K	I	S	M	日本
TK	O	SK		問題
NK	I		T	技術

- (c) 略語で同音異語があればその1つに決める。

TK	O	AS	Y	政策 (製作)
K		S	Y	証人 (商人, 承認)
K	A		T	間接 (關節)

括弧内の語は別の打ち方をする。

この (b), (c) によって *ryakugo* からすぐ漢字でできる語が約 250 語ある。

(d) *印を利用して漢字1字ずつのコードをつくる。このコードの作り方は、つぎの 5.2 で別に述べる。

(e) 頻度の高い熟語には、略語のない場合にも、熟語としてのコードをつくり、一挙に2字以上をまとめて指定する。最初この種類 245 語を選んだ。

5.2 漢字コード

ソクタイプの1打 (21 ビット) は 200 万以上の变化を持つから、1打で漢字1字を指定することはたやすいともいえるが、それを人間がおぼえて、とっさに出すことが必要だから、文字とコードとの連想、同一性をつくるのにくふうがいる。ただ、機械的に対応させたのでは、人間の側に通用しない、人間が発信するコードをつくるために、つぎの3つの方法をとる。

(a) 訓の強い文字は、その訓をキーワードとして、それをソクタイプの普通の打ち方で打つと同時に*を打つ。

岸	K	I	*	I	S
月	T		K*!		

この種類、当用漢字 1,850 のうち 660 字約 1/3 にのぼる。

例：花米母苗泥田土道丸岩岸顔月波板坂……

訓の中には後娘屈柳卸戻葉量など2打語もあるが、コードとしてはこのキーワードの第1打だけをとる。

(b) 字音からコードをつくるものも 250 字ほどある。

例：盆晩案電度液芸非不会……なるべく訓のない字、音でつかうことの多い字、他の語に補足的につくことの多い字などを選んだ。

(c) 熟語をキーワードとするもの。これが約半数である。キーワードはなるべく音訓のどちらか、1字の内部で処理するほうが、オペレータの心理的負担は軽いのであるが、多くの同音漢字を区別するために、その字を含む熟語をキーとすることはやむをえない。1つだけ実例をあげると、*do* という音の字は「土度怒努」であるが、いちばん流動性の大きい「度」に音の *do* をあて、「土」は訓の *tuti* がよく固まっている。「怒」には「怒号」「努」には「努力」をキーとする。

土	tuti	T	T*!	
度	do	TKKS	*	
怒	怒号	TKKS	*	IA KH
努	努力		TK*	SKHT (ryakugo)

6. 熟語本位への切り換え

6.1 文字よりも単語

以上ざっと説明した漢字指定の方法で少し実験してみたところ、分速 250 字 (120~130 語) ぐらいは無理なくこなせることがわかった。

しかし、1つの漢字語を1字1字打つのと、1語としてとらえるのでは非常に能率がちがう。ただ、打数など機械的の面だけでなく、心理的にも1字ずつの漢字を意識の表面に取り出すことはオペレータには大きな負担になる。

言語学的にみても、読んだり、書いたり、タイプしたり (しゃべったり、聞いたりはいわずもがな)、すべて国語を駆使する場合、文字は意識にのぼらないのが理想であり当然でもある。文字を意識するのは、国語として実は異物であり、病的現象なのである。漢語でも文字の意識の強いうちは、未熟語イとみななければならない。

そこで昨年 (1969年) 初め、漢字指定を単独漢字主義から熟語本位に切り換えた。切り換え前にも頻度の高い語、約 500 語は文字に分解せず、一塊として処理した (5.1 (b)(c)(e))。今度はこれを約 10,000 語に拡張した。10,000 語あればたいがいの安定した漢語ははいるから、オペレータの疲労度と速度に大きく影響すると思われる。

この 10,000 語は筆者が辞書から選んだだけであるから、今後各方面の検討を期待する。語数も 10,000 語と限ったことはない。20,000 でも 30,000 でも機械の記憶能力は充分余裕がある。そんなに覚えるのはたいへんだらうという人もあったが、たいがいまともな漢語はみなあるとなればかえって覚えなくてもすむ。

6.2 同音語の処理

漢語でいつも問題になるのは、同音語の取扱いである。漢字ソクタイプでは、言語音以外の方法で機械に漢字を指示するために、5種類の手段を用意している。音に関連を持ちながら、それぞれ打ち方を変えるのであるが、紙数がないから、詳しい説明は割愛し他日にゆずる。

7. むすび

漢字情報をふくむ 21 ビットのさん孔テープができたら、機械がこれを受けて普通文書をつくる道程については、将来もっと近道があると思うが、さしあたりは 12 ビットの漢テレテープにまず反訳するのである。

それによって、現在の端末機械 (たとえば新聞社の漢テレやモノタイプ) に連結する。この反訳の字引と、プログラムは相当膨大ではあるが、今日の電算機の能力からは、むしろ軽い仕事かと思う。

この方法によれば、日本語の文書作成速度がのろいというなやみだけは解消する。現在の漢テレによる入力に比べて 5 倍ないし 7 倍、さらには速記、すなわち、口でしゃべる速さまで延びる見込みがある。

将来、日本語をローマ字で書く時代になっても、漢字を必要としなくなったソクタイプは、ローマ字タイプの 2 倍ぐらいの速度になると思われる。

Y IOTK INOIASKHTY (ということになって)
T S IOTK INO KH (つづけるのでございます)
T KSAIOTK IN IAS H (ことわりなければならぬ)

など、かなの部分で速度がかせげるからである。

(昭和 45 年 1 月 17 日受付, 同年 5 月 1 日再受付)