

「ソクタイプを入力手段に用いるワープロ」作りその後

一、はじめに

「録音機」219号と、1985年6月、速記部同窓会本部委員会発行の「反訳機器研究資料」に、ソクタイプを入力手段に用いるワープロの研究をしており、実用化も可能だから今後も研究を続ける、と書いた。今回、本部からの薦めもあって、その後の研究経過を書くことにする。

前回使用したコンピュータは、当時マイコンといわれたMZ80-Bで漢字もまともに出力できなかったため、主な研究テーマは「速記符号がマイコンといわれる超小型コンピュータで処理できることの証明」であった。その後、マイコンがパソコンといわれるようになり、性能も格段に上がり、一昔前の中型コンピュータ並みの性能を持つようになった。したがって、今回の研究テーマは「実用化を前提として、ソクタイプを入力手段に用いるワープロが通常のワープロより優秀であることの証明」ということになろうか。

二、システムの概要

1、ハード関係

- ① パソコン……NEC PC9801RA（昨年まではPC9801VM2）
- ② ソクタイプ……通常のソクタイプから印字部分を取り外し、キーの下にマイクロスイッチを取り付け
- ③ ①と②をフラットケーブルで接続（キーのタッチは、印字部分が取り外してあるため軽すぎて、決していいとは言えない。）

2、ソフト関係

（1）登録辞書

速記符号を解析し、漢字仮名交じり文に変換するため表1の辞書が設けてあり、それぞれの辞書には、見出し語に当たる速記符号、約200種類に及ぶ品詞、反訳文に当たる日本語が登録してある。必要な語はすべて登録すべきだが、プログラマーとデータ入力者を1人で兼ねるという貧弱な研究体制のため、部分的に辞書を作り、徐々に登録語を増加させながら実験を続けてきた。現在の登録状況は表2のとおりで、登録語は約1万7000語。通常のワープロでは三、四万語必要だといわれており、ソクタイプ符号の特殊性から実際には五、六万は必要だと思われるので、今後辞書データの入力課題として残る。

<表 1 >

左C略語辞書	中のむれを含む左片手略語
右C略語辞書	中のむれを含む右片手略語
左B略語辞書	中のむれを含まない左片手略語
右B略語辞書	中のむれを含まない右片手略語
左キホン辞書	左のむれのキホン
右キホン辞書	右のムレのキホン
中キホン辞書	中のムレのキホン
わのも辞書	中のムレの助詞
両手略語辞書	T K及びKの略語、その他の両手略語
付属語辞書	助数詞（円、万、匹、羽、軒、件）等 中のむれあるいは右のむれから始まり 二つのむれ以上にわたる付属語
一般辞書	左のむれから始まるキホンあるいは片 手略語を含む1ストローク以上の自立 語及び付属語

<表 2 >

動詞、助動詞	すべて登録済み
数詞	基本的なものは登録済み
動詞	ほとんど登録済み
略語に関連する自立語 例 裁判、会社、に関する、 私、あなた、鳥、取り	すべて登録済み
左、中、右のむれ、それぞれ 一つのむれのみで打てる語 例 手、歯、今日、蟻、回、 会 員	ほとんど登録済み
一般辞書	最初の読み仮名が五十音の アで始まる語から、サで始まる 語まで登録済み

(2) 入力と変換の仕方

ソクタイプで速記符号を入力すると、ディスプレイの右端部分にソクタイプ符号が表示され、変換符号あるいは句読点符号を入力すると、ディスプレイの左側部分に漢字仮名交じり文が表示される。(ディスプレイでの表示自体は図1、図2を参照。また、新設符号については図3を参照)

<図1>

共和平成三年五月二十一日、

Y K ASK T
Y H A O O AS Y
SAI TK O AS H
AIO
HK O A KH
H S I OI S TY
I T I I S H
T I
NO A T
NO A H

Y協和 T今日

<図2>

今日は平成三年五月二十一日、火曜日、曇。

Y K ASK T
Y H A O O AS Y
SAI TK O AS H
AIO
HK O A KH
H S I OI S TY
I T I I S H
T I
NO A T
NO A H
T A O
NO A K Y
K A Y
T KSA O
NO A T

K	O SK
T K I	
	N ASK
	NO A K Y

<図 3>

NO A T	①読点「、」
N ASK	②句点「。」
NO A H	③変換符号
N K	④部分確定符号
NO A K Y	⑤全文確定符号

同音異義語があれば最初の同音異義語にカーソルが移動するので、①変換符号②部分確定符号③全文確定符号のいずれかを入力する。

変換符号を入力すると、選択候補が最下段に表示され、そのうち一つを選択することができる。

部分確定符号を入力すると、次の同音異義語にカーソルが移動する。次の同音異義語がなければ全文確定する。

全文が表示どおりであれば全文確定符号を入力する。

図 1、2 が変換状態を示しており、図 1 はソクタイプで

キョーワ　ヘーセー　サンネ　ン
ゴガ　ニジュー　イチニ　チ

及び読点符号を入力し、カーソルが「共和」にある段階で変換符号を入力した状態。

図 2 はその後「今日」を選択するために「T」を入力し、「カヨー　ビ」及び読点符号、「クモリ」及び句点符号、全文確定符号を入力した状態である。

(3) 変換の仕組み

変換符号あるいは句読点符号が入力されると、コンピュータは辞書に登録してある語の中から、入力された符号に合致する最も長い語を探し、接続可能な品詞であるかどうかを検定する。接続不能であれば、次の語、さらにその次の次の語という具合に接続可能な語が見つかるまで検定を繰り返す。

したがって、二通りに読める速記符号であっても前後の接続関係から適正な漢字の選択が可能であるし、「お」と「を」の区別も可能である。

(4) 変換できる文章

「登録辞書」の項で説明したように、当システムは辞書が完成していないため、いかなる文章でも変換できるというわけにはいかない。最初の読み仮名が五十音のシで始まる語以降の語は変換できない。しかし、ソクタイプ入力されたデータを日本語の漢字仮名交じり文に変換する基本システムは出来上がっており、基本的な語は登録してあるので、入力実験はできる。入力実験をしてみたい文章の中に登録していない単語が出てきたら、同じ品詞の登録している単語で言い換えることによって変換はできる。図3は変換可能な文章の一例であり、口語文法が採り入れてあることに注目してほしい。

三、ソクタイプを使用する利点

1、文字データの入力方式

文字データをコンピュータに入力するには、キーボードを用いなければならない。音声入力あるいは手書き入力という手段も研究されてはいるが、実用には程遠い。

キーボードを用いて入力する方式には、

- ① 目的文章の文字一字一字を人間が指定する方式
- ② 目的文章を読み仮名で入力し、漢字へのあてはめはコンピュータに任せる方式がある。

①については、漢字と文字コードが一対一に対応し、コンピュータにとっては入力されたコードに対応する文字を出力すればよく、その負担は軽い。一方、人間にとっては使用する漢字すべてについて文字コードを覚えなければならない、大変な努力が必要である。手書き文字の場合、あの複雑な漢字を辞書もなしで書いていることからすれば、文字コードを覚えることなど、さほど困難なことではないように思えるが、人間の頭脳は不思議な構造になっており、記憶しやすいものと記憶しにくいものがある。数字の羅列などは最も記憶しにくい部類に入り、JISコードの記憶などできるものではない。

そこで、①の方式は、人間が記憶しやすく、かつ高速に入力できる文字コード体系を如何に作り上げるかが課題となる。そのうまく作り上げられた代表的な例がラインプットであり、そのほかに豊橋技術科学大学で考案されたタッチタイプ、日立の2ストローク方式等もある。また、全文字を一枚の板状に配置し、目で確認して入力するネオライターのワープロ版であるペンタッチ方式、あるいは文字盤を多段に配置した多段シフト方式もある。

②が現在主流の仮名漢字変換方式である。この方式は、人間にとっては五十音の入力方法を覚えればよく、その負担は軽い。一方、コンピュータにとっては膨大な漢字の中から適正な漢字を選択しなければならず、難しい作業となる。

2、仮名漢字変換の方式

ワープロの仮名漢字変換方式には次の種類がある。

- ① 単漢字変換……漢字の音訓を入力して、一文字の漢字に変換
- ② 単語変換……単語の読みを入力して一語の単語に変換
- ③ 単文節変換……自立語に続く付属語も一度に入力して、一文節の漢字と仮名に変換
- ④ 複文節変換……複数の文節をべた書きで読み仮名を入力して、漢字と仮名に変換

ユーザーが求める変換方式は当然べた書き複文節変換である。すなわち、人間がコンピュータに期待するのは、文章が長かろうが短かろうか、頭に浮かんだ文章をそのまま入力し、それが正しい漢字仮名交じり文に変換されることであって、単語ごとに区切るとか、文節ごとに変換キーを押すなんてことはしたくないはずである。そして、現在はほとんどのワープロはこの方式が可能となっている。

ところが、現実のワープロはこのべた書き複文節変換が可能ということであって、適正な漢字仮名交じり文に変換されることが保証されているわけではない。現実には、多文節の長い文章を変換させると、ワープロは文節の区切り間違いを頻発し、ユーザーはワープロの区切り間違いを訂正するのに時間を取られる結果となる。したがって、文節単位あるいは単語単位で変換したほうが結果的には早い、というのが現状である。

具体例を挙げると、「上のクラスに行く」という文章を入力する場合、せめて「上のクラス」ぐらいは一気に入力するのが人間の発想からいって自然だと思われるが、私が実務に使用している最高裁の私物ワープロ持込み許可条件に合致するワープロは、「上野暮らす」と変換してくれる。仕方ないからブツブツ文句言いながら、「上野」の文節を短く区切り直して、「上」に変換させて部分確定し、「の」を無変換し、「暮らす」を片仮名変換させるという、二重三重に手間のかかる変換をすることになる。隣でオアシスを使用し、高速入力を誇るAさんは多分、次のような入力をしているものと思われる。「うえ」と入力して「上」と変換し、「の」を入力して無変換キーを押し、片仮名モードにして「クラス」と入力。

3、当システムの変換方式

ソクタイプを入力手段に用いるワープロもべた書き複文節の仮名漢字変換方式の一種であるが、通常キーボードの仮名漢字変換とは異なり、文節の区切り間違いをすることはほとんどない。理由は以下の3点。

- ① 中のむれを助詞に用いる

②略語を多用する

③両手略語を使用しない場合でも、2音以上を1ストロークで入力できる速記官であれば①と②はだれしも理解できることであるが、③についてはほとんど知られていない。

ワープロが文節の区切り間違いをする有名な例として

クルマデハコヲハコブ

という文章がある。この文章の解釈として A「クルマデ」の次の「ハ」を助詞と解釈すると、後半が「子を運ぶ」となり、前半は①「車で」②「来るまで」③「繰るまで」の3文が考えられる。

通常キーボードの仮名漢字変換では、AとBのどちらを意図して入力されたかは全く分からない。ソクタイプ入力では区別できることは説明するまでもない。

次に、「クルマデ」の3解釈についてはソクタイプ入力でも区別して入力することができないではないか、と言う人があるかもしれない。しかし、一定のルールを設けることにより①と②③との区別は可能になる。ただし、②と③は同じ品詞の同音異義語であり、ソクタイプ入力でも区別はできない。

一定のルールとは、“自立語は左のむれから順次入力し、追い込みはしない”というソクタイプの通常の入力方法で、この通常の入力方法を規則に格上げし、例外は許さないとするによって、通常キーボードでは不可能な日本語分の解析ができるようになる。

具体的にいえば、「箱」「車」など1ストロークで入力された図4⑤の「クルマデ」は、「車で」と解析することはなく、「来るまで」あるいは「繰るまで」と解釈するし、1ストロークで入力された図5①の「ハコ」は「箱」と解釈し、助詞の「は」と名詞の「子」に分解することはなくなる。

ここに挙げた区切り間違いの例文は特殊な例ではなく、ワープロを使用していると毎日頻繁に経験することである。次に、私の経験で、ワープロがあまりにも突飛な区切り方をするのでびっくりした例を挙げておくので、ソクタイプ入力では起こりえないことを各自、ソクタイプ入力を頭に浮かべて納得してもらいたい。

<図4>

H A	O K	①箱
K TK	ASK	②車
H A		③ハ
K O		コ
K TK		④クル

KSA	マ
K TK OIASK	⑤来るまで

彼は……枯葉
 乗り込んでおった……の離婚でおった
 私が医者です……私会社です
 歯と歯ブラシ……鳩はブラシ
 見ないように……皆異様に
 他事業所……他事行書
 再鑑定……再刊亭
 再登録……斎藤六
 一定量……言って衣料
 その話……園は梨
 全経過……全景か
 やったのはいつですか……やったのハイツですか

4、品詞検定の仕組み

当システムにおける品詞検定の概略を前項の「再鑑定」を例にとって説明する。

図5の①のように1ストローク目で「サイ」、2ストローク目で「カンテー」と入力すると、コンピュータは最初に一般辞書を調べ、一般辞書に登録してある語の中で合致する語を探す。（「再鑑定」という語が頻繁に使用されるなら辞書に登録すべきであるが、接頭語まで含めた語を登録すると辞書が膨大になるため、通常は接頭語、接尾語を含めた辞書作りはしない。また、図5②のように追いついで入力すると「再刊」と接尾語の「亭」を探し出す）

<図5>

Y	SAI				①再
	K A	NO A	TY		鑑定
Y	SAI		A K		②再刊
	AIO	O A	TY		亭

一般辞書を探して、2ストロークとも合致する語はない。次に、1ストローク目の「サイ」を探すが、これも合致する語はない。

次に、ソクタイプ符号の規則に則って合致する語を探すと、左キホン辞書に「サイ」があり、そこには次の語が登録されている。

- ①名詞……差異、犀
- ②接頭語……再
- ③助数詞……歳
- ④接尾語……祭

次に、2ストローク目の「カンテー」について一般辞書を探すと、「鑑定」「艦艇」「官邸」という漢語名詞があり、「サイ」と「カンテー」について接続関係を品詞検定表によって調べることになる。

品詞検定表には、品詞ごとに、接続する品詞及び接続のしやすさを表す点数付けがされており、「カンテイ」という漢語名詞の前に接続する語としては「再」の接頭語が最も接続しやすいことが分かる。

「鑑定」「艦艇」「官邸」については品詞を同じくする同音異義であり、学習機能により直近に選択された語が最初に表示される。

四、まとめ

登録辞書の項で説明したように、このシステムは完成しているわけではない。辞書の登録語も少ないし、編集機能、印字機能も作っていない。しかし、ここまでシステムを作ってきた経験から、ソクタイプ入力の日本語処理について思うことは数多い。思いつくことを以降に書き並べる。

▼ アメリカでは、ソクタイプで入力したものを自動反訳し、即座に表示するCATシステムが完成し、実際に稼働していると聞く。しかし、日本語の場合、同じ品詞の同音異義語が非常に多く、コンピュータに適正な漢字を判断させるには無理がある。速記官が入力しながら同音異義語を選択する時間的余裕はないし、間違った漢字を表示するより平仮名表示が良いとなると、ほとんどが平仮名表示になってしまい、とても実用に耐える速記録にはならない。

したがって、日本語CATシステムが実現するとは思えない。

▼ ソクタイプ符号には、キマリと称する、速記速度を上げるために設けられた規則が多数ある。また、略語の前の特定の語は省いてもいいというカッコつきの略語も多数ある。

当システムにおいても当初はすべてのキマリ、すべてのカッコに対応できるよう努めてきた。しかし、対応しやすいものと、対応しにくいものがあり、無理に対応すると、いたずらに同音異義語を増やして、変換の際の選択に時間がかかることが分かった。具体的にいえば、図1、2に示した数詞の入力がその一例である。図1、2においては「三年」を「サンネン」と「ン」まで入力し、「五月」を「ゴガ」とのみ入力して「ツ」を省いているように、当システムにおいては、「～年（ネン）」の「ン」及び「～月（ガツ）」の「ツ」を省いても省かな

くても正しく表示できるようにはしてある。しかし、ときとして「一月（イチガツ）」のつもりで入力した「イチガ」が「位置が」と表示されたり、「一元」のつもりで入力した「イチエ」が「位置へ」と表示されることがある。また逆に「位置が」のつもりで入力したものが「一月」と表示されることがある。

ソクタイプを入力手段に用いる日本語処理システムを裁判所の速記官が使用する場合、二通りの使い方が考えられる。

① 当システムを入力システムと変換システムとに分離し、法廷では入力機器としてフロッピーディスクに入力し、書記官室では当システムを反訳機器として使用し、持ち帰ったフロッピーディスクから漢字仮名交じり文に変換させつつ、適正な漢字を選択する方式

② 法廷の立ち会いは現状どおりとし、当システムを通常の日本語ワープロと同じく反訳専用機器として使用する方式

当システムを①として使用する場合、速記速度を低下させることは許されないため、現状の打ち方を変更することはできない。②として使用する場合には、トータル的に速く速記録を完成させるのが目的であるから、そのためにはカッコの中もすべて打ち、キマリの使用についても制限したほうが良いと思われる。

▼ ソクタイプを入力手段に用いる当システムにおいて品詞の区切り間違いをすることはほとんどない、ということの正確な意味は次のとおりである。

① 一つの単語を途中で区切って別の単語と解釈することはあり得ないという趣旨であり、一つの単語と解釈するか、自立語プラス付属語と解釈するかの区別がつかない場合はある。（自立語は常に左のむれから始まるため、1ストロークで入力された「ハコ」の「コ」を自立語である名詞の「子」と解釈することはないが、「今日は」と「共和」はどちらも左のむれから始まる自立語であるため、区別できないことはある。）

② 一つの単語か、付属語を含む文節かの区別がつかない場合においても、三の3で述べたルールを設けることによって検討対象の語が少なくなるため、厳格な品詞検定によって適正な判断ができる。③ 品詞検定によっても判断ができない場合に、当システムでは同音異義語扱いとする。

発言者が固定され、話題が限定される法廷の発言においては、立ち会う事件が変わるたびに、ひとわたり同音異義語を選択することは必要であっても、以後は学習機能が有効に働くように思う。したかつて、当システムを、前に述べた反訳専用機器として使用した場合、日本語文の解析に関しては通常のワープロよりはるかに優秀で、かつ高速入力のワープロになると考える。