



Title	集合としての文字
Author(s)	樺島, 忠夫
Citation	語文. 1974, 32, p. 122-128
Version Type	VoR
URL	https://hdl.handle.net/11094/68629
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

集合としての文字

樺 島 忠 夫

日本語の「文字」[、]かな[、]漢字[、]などの語は、個々の文字を意味するとともに、その集合体をも意味する。たとえば「ヤというカタカナは」というときのカタカナは個体としての文字を意味し、「動植物名はカタカナで書く」というときのカタカナは集合体を意味しているよう。

表記論、文字論の研究においては、個体としての文字、文字の集合体、文字が集合の中から取り出され、並べられてできる文字列の三者が区別されることが必要であると思うが、集合体を意味する用語は、国語学においてまだ一般的に認められたものがない。

国語学概論の文字論の章を見ても、以上の三者、あるいは少なくとも個体としての文字と、文字の集合体との二者を区別して説いているものがないところを見ると、用語以前に、概念としてこれらの別が定着していないように思われる。

ただ数理言語学において、この三者が基礎概念として設けられており、集合体は「アルファベット」「字母系(野口宏)」「字彙(水谷静夫)」、文字列は「語」「列」「連糸(水谷)」などと呼ばれている。

文字を個体としてでなく、集合体としてとらえることは、日常生活においても行われている。そして、このとき集合体を、ただ単に文字の集まりとしてではなく、ある特性になったものとしてとらえている場合があるように思われる。

たとえば「ローマ字で日本語を表記すると、長たらしくなり読みにくい」などということがある。これを分析すると、

(1) 「ローマ字」という語によって漠然とながら、文字の集合体を考えている。

(2) その集合体を、日本語に対する表記機能の面からとらえている。ということが抽出されよう。

一人の人間を見るとき、われわれは、頭、手足、胴、衣服などの集合体としてみるだけではない。その人間の人格という面をも総合してみている。このように、文字の集合体をとらえるときにも、ただ単に文字が集まったもの——たとえば、ひらがなであれば四十六(八)文字から成るまとまり——としてではなく、表記機能の面から特性づけて見る観点があり得る。

実際には、かなもじ論者、漢字論者などというときの、かなもじ、漢字には、人間の人格に当たるものが与えられているとみてよいの

ではないかと思う。

＊

私は、文字の集合体を、ある言語に対する表記機能から特性づけてみると、「文字体系」と呼んだらどうかと考えている。

では、文字の集合体のある言語に対する表記機能から特性づけてみるとはどういうことか。

これを説明するためにまず抽象的モデルを構成してみよう。文字論の出発にあたって、現実存在する、具体的な文字の分析から始めず、抽象的モデルの分析から始めるには、それだけの意味がある。文字の集合というためには、何を文字とすることが明らかになっていなければならない。要素を確定することが必要である。たとえばカタカナ表記における長音符号「ー」は文字か、またひらがな表記における反復記号「いろく」の「く」は文字かなどが問題になる。

何を文字と認めるかをきめる規程を設けること自体は簡単である。たとえば、

(1) 音に対応する。

(2) 他の文字と組み合わせられて音列に対応する文字列を作ることができる。

この二つの条件をみたす記号を文字と定義する、などである。

しかし、「音に対応する」とは、単独で、一つの音に対応するという意味か、あるいは文字列中において音と対応する部分を構成するという意味かを考えると、問題は複雑になる。結局、何が文字かをきめるには、単純な判定規程を設けるだけでは不完全で、文字の表記機能あるいは、その文字による表記体系をとらえる中で行われ

なければならないことがわかる。

つまり、何が文字かを決定するには、文字論、表記論の中で体系的に行われなければならない。これは文法論における品詞の決定と同じであろう。

ところが、文字論、表記論を行うためには、文字あるいは文字体系が決定されていなければならない。そこで手続き上で矛盾が生じる。

これを解決するには、文字、文字体系など基本となる概念を仮に設けて、文字・表記の抽象的モデルを構成し、ここから理論を導き出し、その後これによって現実を解釈する方法をとることが必要になるのである。

＊

そこで、まず文字体系の特性についてモデルを設けることによって観察しよう。

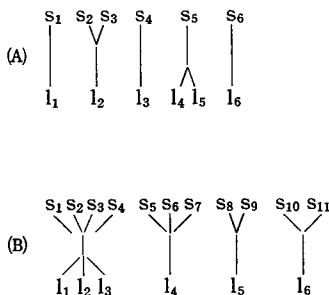
有限な数の文字と、それを要素とする集合 A を考え、 A を文字体系とよぶ。

一方、ある言語を考え、この言語を構成する最小の音の要素（たとえば音素）の集合を考える。

文字と音との間には対応関係があるが、一つの文字が必ずしも一つの音に対応するとは限らない。たとえば英語の th 、日本語の「きゃ」は、二字で一つの音素、音節に対応する。

そこでその言語の、意味のある音列 $s_1 s_2 s_3 \dots$ と、それに対応する文字列 $l_1 l_2 l_3 \dots$ とを考えると、たとえば図のようなものがあるだろう。(A)では、音列 $s_2 s_3$ が一つの文字 l_2 に対応し、 s_5 が文字列 $l_4 l_5$ に対応している。(B)では、音列 $s_1 s_2 s_3 s_4$ の全体が文字列 $l_1 l_2 l_3$ の全体

に対応している。



たとえば

/kjo/ = 「きょう」

がこの例である。

まず音に対応する文字列の最小の要素を求める。その要素は、

- (a) 音列に対応する最小の要素であり、その要素の数も最小であり、
- (b) 音列との対応関係を記述するとき、たとえば「ウ段拗音表記の文字列に続くウはオと読む」などのような条件つき対応規則を要しない。

という条件をみたすものとする。

意味のある音列 $s_1 s_2 \dots s_i \dots$ と、これに対応する文字列 $l_1 l_2 \dots l_j \dots$ とがあるとき、次の条件をみたすように文字列を分割する。

- (1) 部分文字列 L_k に対応する音列を $S(L_k)$ 、 L_{k+1} に対応する音列を $S(L_{k+1})$ とすると、

$$S(L_k L_{k+1}) = S(L_k) S(L_{k+1})$$

すなわち、文字列 $L_k L_{k+1}$ に対応する音列 $S(L_k L_{k+1})$ は、 L_k に対応する音列 $S(L_k)$ の次に L_{k+1} に対応する音列 $S(L_{k+1})$ をつけたものに等しい。

- (2) こうして求められた部分文字列 $L_1 L_2 L_3 \dots L_k L_{k+1} \dots$ のすべては (1) の条件をみたす最小の文字列である。

この最小の文字列を、表記要素、と名づけることにする。この表記要素が先にあげた条件 a、b をみたすものである。

たとえば、かなであれば

/kjoo toqkjnu hikari /
きょう とっきゅう ひかり

のように「きょう」「と」「きゅう」「ひ」「か」「り」が表記要素となる。ここに

/kjnu / = /kin / + /u /
きゅう きん う

だから「きゅう」は「きん」と「う」との二つの表記要素に分けられるように考えられるが、

/kjnu / ≠ /kinu /

つまり、長音と、同一母音を二つつけたものとは異なると考える

と、
 /kin / + /u / ≠ /kjnu /
きん う きん

となると考えられるので、拗長音を表す文字列は一つの表記要素と解釈するのである。

表記要素に対応する音列は、たとえばかなの場合「あ、きや、きょう」のように音節であったり「ん、っ」のように音素であったり

して、音の単位としては一定しない。これは文字（列）中心の記述であるから、当然そういう場合も生じるわけであり、表記要素という概念の有用性をそこなうものではない。

*

この表記要素が、どのような形の音列に対応するかによって、文字体系を特性づけることができる。

- ・表記要素が主として音素に対応するとき——音素文字（体系）
- ・表記要素が主として一つの音節に対応するとき——音節文字（体系）

・表記要素が n 個 ($n=1, 2, 3, \dots$) の音節に対応するとき——多音節文字（体系）

日本語におけるローマ字、英仏独におけるラテン文字は音素文字、日本語におけるかなと中国における漢字は音節文字、日本語における漢字は多音節文字である。

たとえば、

$a=/a/, b=/i/, c=/u/$,
 $aa=/ka/, ab=/ki/, ac=/ku/$,
 $ba=/sa/, bb=/si/, bc=/su/$,
 $ca=/ta/, ca=/ti/, cc=/tu/$,
 $aaa=/na/, aab=/ni/, aac=/nu/$,
.....

のような対応をもつ文字体系 $\Delta=(a, b, c)$ があれば、この文字体系は音節文字体系である。

このことから、音素文字、音節文字などをきめるのは、文字ではなく表記要素であることがわかる。

*

文字体系を特性づける第二の面は、

「同一音列に対応する二つ以上の表記要素があつて、視覚的に表現性の分担を行うかどうか」

である。

この場合にまずあげられるのは、表意文字体系である。漢字ではカンという音節に対応する表記要素に観、看、感、間、乾など（ただし現代共通語音）があつて意味の別を分担している。英語では、*night, knight* の *n* と *kn* とがこれにあたるだろう。

英語では文頭、ドイツ語ではさらに名詞の頭に大文字を使うから、大文字、小文字の別も視覚的に表現性の分担を行う、異なる表記要素を構成する文字と考えてよいだろう。かなの中に異体字を置いて何らかの表現性の分担をさせるとすれば、これもここに入れてよい。

*

ある文字体系によって、同一言語を表記するとき、今あげた、

・その文字体系が音素文字体系か、音節文字体系かなど、表記要素に対応する音列の長さ

・同一音列に対応する二つ以上の表記要素があつて、視覚的に表現性の分担を行うかどうか

によって、その文字体系の他の特性が導き出される。

ただし、表記要素に対応する音列の長さといつても、*a, b, c, a a, a b, a c, …* が音節に対応する場合を考えたように、表記要素自体の長さも考えなければならない。

そこで、表記要素をすべて集めた集合を作り、表記要素を構成する文字の数の和で、表記要素に対応する音列の音の和を割った値を

求め、これを、一文字あたりの平均音数、とよぶことにする。

たとえば、ひらがなの場合、

a、一音素に対応する一文字の表記要素(2)

/N/—ん, /q/—っ

b、二音素に対応する一文字の表記要素(65)

/a/—あ, /i/—い, /u/—う, /e/—え, /o/—お,

/ka/—か, /ki/—き, /ro/—ろ

c、二音素に対応する二文字の表記要素(3)

/ca/—っあ, /ce/—っえ, /co/—っお

d、三音素に対応する一文字の表記要素(4)

/wa/—わ, /ja/—や, /ju/—ゆ, /jo/—よ

e、三音素に対応する二文字の表記要素(98)

/kia/—きゃ, /kju/—きゅ, /kjo/—きょ, /tjo/—りょ

/'aa/—ああ, /'ii/—いい, /'uu/—うう

/kaa/—かあ, /roo/—ろっ

f、三音素に対応する三文字の表記要素(3)

/caa/—っあ, /cee/—っえ, /coo/—っお

g、四音素に対応する二文字の表記要素(4)

/'waa/—わあ, /'jaa/—やあ, /'juu/—ゆう, /'joo/—よう

h、四音素に対応する三文字の表記要素(33)

/kja/—きゃあ, /kju/—きゅあ, ...

/'ja/—りゃあ, /'juu/—りゅあ, /'joo/—りょあ

とすれば、一文字あたり、平均一・五音素となる。

これを使えば次のことがいえる。

I 同一言語を表記するとき、一文字あたりの平均音数が大である

ほど、また同一音列に対応し表現性の分担を視覚的に行う表記要素が多いほど、

a その文字体系に属する文字数(異なり文字数)が大になる。

b その文字体系に属する文字の字形(示差的特徴の組み合わせ)が複雑であり得る。

bはaの結果として生じるが、文字体系によっては余分な示差的特徴をそなえる——必要以上に複雑な形を持っていることもあり得るので、aからbが必ず生じるとはいえない。

II 同一言語を表記するとき、一文字あたりの平均音数が大であるほど、その文字体系によって表記された文字列の長さ(延べ文字数)は小になる。

たとえば、日本語をローマ字で書く方より、かなで書く方が短く、漢字かなまじりで書く方がさらに短い。

*

その言語の音のちがいをどの程度に区別するかという、音の面の区別をも考えておく必要があるだろう。たとえばローマ字で

ta, ti, tu, te, to

と書くか

ta, chi, tsu, te, to

と書くかといったことである。

このとき、次が いえる。

III 同一言語を表記するとき、一文字あたりの平均音数が等しければ、音の区別を精密に表現する文字体系ほど、それに属する文字の数(異なり文字数)が大になる。

IV 同一言語を表記するとき、異なり文字数が等しければ、音の区

別を精密に表現する文字体系ほど、一文字あたりの平均音数が小になる（あるいは表記要素の長さの平均値が大になる）。

表記要素の数を音の増加に応じて大にするには、異なり文字数を増すか、または文字を組み合わせて一つの表記要素を形作らなければならぬ。前者の場合がⅢで、後者の場合がⅣである。

*

以上のように、音の集合と文字の集合、さらに両者の間の対応関係を定めることを行うことによって、文字体系の特性が定まってくる。

現実に見られる文字の集合体は、

- (1) それに対応する音の区別の精密さ
 - (2) 音列と文字列との対応関係
 - (3) 同一音列に対応する表記要素の表現性の分担のいかんをそなえており、それに応じて
 - (4) 文字の数（異なり文字数）
 - (5) 一文字あたりの平均音数が定まり
 - (6) 文字の字形の複雑さ
- もある程度定まっている。またこの文字の集合を使って表記したときの
- (7) 表記された文字列の長さ
- が定まる。

つまり、ある言語を表記するために、ある文字の集合をとって考えたとき、その文字の集合は、以上の七つの特性の束によって特徴づけられたものとなる。

文字論、表記論においては、文字を表記機能の面からなめることは当然のことである。人間を、ただ手足や頭などの部分品からなる物体としてみるだけでなく、ある働き、経歴、人格を持つ存在としてみるように、文字の集合を考えるときにも、表記機能に関する特性の束になうものとして考えるのが当然であろう。文字の集合をこのようなものとしてながめるときに、文字体系とよんだらどうかと思うのである。

一つの文字体系に属する文字の間にも何らかの関係が認められるようである。たとえば文字間の視覚的な差は、体系内の差よりも体系間の差が大きいとか、ひらがなにおいては、使用率の大きさと字形の複雑さとの間に相関関係がみられ、よく使われる文字ほど字形が簡単である（樺島忠夫・佐竹秀雄「ひらがなの字形に順序を与えらる」計量国語学66）などである。

*

文字体系あるいは文字の集合というためには、どの範囲のものを文字と認めるかが問題になる。

たとえば、かな表記の文章を観察すると、文章（表記された記号列）は次の三種に分割できる。

- a 表記要素
- b 表記要素に対する反復記号（「 」）
- c aおよびbを文章から取り除いたときに残る記号（句読点、かっこ、ダッシュなど）

厳密に考えるためには、句読点がなぜ表記要素に入らないかをきめる手続きが必要であるが、今はこれに触れないことにする。

以上の三つに記号列が分割できたとしよう。表記要素を構成する

記号を文字と定義することが、文字定義の一つの方法であるが、表記要素を構成する記号を分類すると、

イ、単独で表記要素を作り得る記号

ロ、単独では表記要素を作り得ず、イの記号に付属して表記要素を作る記号（小文字の「や、ゆ、よ」、カタカナ表記における長音符号「ー」）

ここで、イを自立文字、ロを附属文字とよぶことにすれば、文字の範囲が確定する。

なお同じ文字が異なる文字かの判別も文字体系を確定するために必要である。同じ文字とは、同一文字体系に属するものであることが必要だが、これは、使用状況や歴史的過程によってきめる以外に方法はない。

そこで、同一文字体系内において同じ文字が異なる文字かを判定する問題が残る。同じ文字とは、同一の文字体系に属し、同一の機能をもつものであると考えることができる。

イ、ある表記要素中の文字 l_1 を l_2 に取りかえても、その表記要素と音列との対応関係は変わらない。

ロ、ある表記要素中の文字 l_1 を l_2 に取りかえても、文章中におけるその表記要素の表現性は変わらない。

の二つの条件をみたすとき、 l_1 と l_2 とは同じ文字だといってよいだろう。

ロの「表現性が変わる・変わらない」はあいまいに見えるが、表記要素の意味が変わるとか、大文字で書くべきところを小文字で書くというように習慣で認められているところをそこなうとかを「表現性が変わる」と考えてよいだろう。

こう考えると、文字体系というときには、英語の場合には、その要素として大文字と小文字とを別のものと認めることになる。

*

以上いろいろと述べてきたが、文字論、表記論を行うためには、文字、文字体系などの基礎的な概念をはっきり立てること、またそのためには、ただ文字・表記の現象を観察・整理するだけでなく、モデルを作って、表記機能の面から操作・整理を行って文字・表記を観察する視点を設けることが必要ではないかと考えるのである。

（京都府立大学教授）