

か しま えい いち  
鹿 島 英 一

学位の種類 博士(文学)  
学位記番号 文博第11号  
学位授与年月日 平成5年7月8日  
学位授与の要件 学位規則第4条第1項該当  
研究科・専攻 東北大学大学院文学研究科(博士課程後期3年の課程)  
英文学英語学言語学専攻  
学位論文題目 文字の統計学的処理  
論文審査委員 (主査)  
教授 中村 完 教授 中村 捷  
教授 加藤 正信  
助教授 千種 眞一  
助教授 後藤 斉

## 論文内容の要旨

本論は日頃、文化・文明論的な考察の対象になりがちな文字の、情報伝達符号という側面に焦点を当てたものである。目的は文字を相互に比べる方法を探ることである。かかる試みはほとんど類例を見ないために、全くの第一歩から始めなければならなかった。尚、方法論の支柱には集合論的な発想に加え、幾何学や統計学などを採用している。

全体は七つの章よりなるが、それは内容面から前半と後半に二分できる。即ち、第二章までの三つの章が前半であり、第三章以後の四つの章が後半である。ところで、「文字を比べる」と言っても自明ではない。そこで先ず、前半では広く文字と文字言語の一般を取り扱い、その中で具体的に「何」を「どう」比べるのがよいかについて考察した。そしてこれに対し、後半では前半で提示した方法を実際にラテン系文字を使用する、現行かそれに準ずる諸語文(大文字:52、小文字:82)に適用した、その結果を示した。尚、(この方法の有効性を試すための)最初の例としてラテン系文字を選んだのは、推測統計学が利用できる数的条件(一定数以上)と、幾何学的外観(字形)のかなり異なる複数種の文字(大・小文字)を持つことなどによる。

序に当る章「文字言語」では、先ず文字言語の特徴を幾つかの観点から考えた。例えば、音声言

語との違いや文字関係の技術の変遷である。そして、先ず前者に基づいて、(本論で扱う文字言語の)分類単位が現行の常識的なものと必ずしも一致しない(が、大幅な変更も要しない)旨を述べ、併せて文字体系(文字言語の構成要素)の分類の概略も示した。

次に、後者の観点を軸にした、書写材料・道具と字形との関係を未来から過去に溯って考察した。具体的には、先ず(最近始まったばかりの技術面での変化に起因する)色の将来の字形面への寄与の可能性や、音声言語の保存・運搬の実現に伴って顕在化しつつある文字・音声両言語の競争関係について述べた。そして、次に書写する際の材料・道具・方法を結果の明らかな過去にまで溯って拾い集め、同時に各文字に書法・字形・用途の三方面から簡単な分類を施した。歴史的な実用技術としての文字の特徴を探るためである。

尚、過去の実例の収集に際しては、東アジア(漢字圏)、インド文字圏、古代メソポタミア(楔形文字圏)等の各地域・文字圏を単位とした。

第1章「文字収集の構造」では、(文字比較に際して、具体的に)「何」を比べるかについて集合論的な観点から述べた。換言すれば、各語文の実際の文構成要素を数学的な意味での集合と考えること、それは幾つかの下位の部分集合(表意文字と基本文字<以下:文字>と句読点と狭義の符号<以上:符号>)に分けられること、そして本論ではその一つ(基本文字集合)を比べるつもりである旨などを述べ、併せて集合の要素の決定に関係する文字単位の採り方も示した。

少し詳しく見てみよう。始めに、(自明でない)問題点を二つ指摘した。即ち、一つはドイツ文の [ä] やフランス文の [æ] や英文の [ch] を一文字単位と見做すのか否かといったことを典型例とする文字単位の取り方であり、いま一つは(実際の文には常にある)日本語の [。] の様な句読点の類や諸言語に見られる所謂アクセント符号や算用数字や英文の [\$] (貨幣単位)等と普通の文字(基本文字)との境界の付け方である。

先ず、二番目の方から始める。ここには明らかに文字と違うものがある。一般に、句読点に代表されるこの種の符号は概して古い文字言語より現代語文に豊富であり、多い方が文が分かり易い。そこで、この辺の事情と使用者の技術水準の関連を指摘し、併せて他の水準低下効果にも触れた。そして、「アクセント符号と分ち書き」「分音符号と弁別符号」等々の項目を立てて詳しく検討し、文字と符号の境界付近にあるものを仕分けした。

また、使用文字の別による語彙の分類表を使って、[\$] を一例とする表意文字を集合としても考えた。その表は、先ず現行の日本語文を主な範として、比較的タイプの異なるマレー文、ヒンディー文、英文等によって補った。

最後に、語中での位置に応じた字体の交替する文字対や語頭母音の表記法やセム系語文の「母音符号」等も考慮し、文字単位の取り方を決定した。

第2章「文字集合の統計的指標」では、(文字比較に際しての)物理的な「何」とそれを「どう」比べるかについて述べた。

具体的には、(序章末での書法・字形・用途の三方面からの分類の結果を受けて)冒頭で、これ

ら分類相互間の関係を分析した。その結果、本論では（絵的文字を除く）直線的な字体を対象とする旨を述べた。つまり、これが物理的な「何」である。そして、この字体だけを対象としても幾つかの理由のために、実際にはかなり多数の文字言語を扱えるのである。

以後は、それを「どう」比べるか、即ち比較のための指標の選択と基準の設定についてである。結論としては、複雑さと曖昧さという概念を採用し、これを①「画数」、②「類似文字数」、③「文字の型」という指標と（字体の性質も考慮した）「基準字形」の導入によって表現した。そして、字形の異同も含めて各種の基準も用心深く立てた。そこには、記述の際と判読の際で基準設定にずれが出る際には前者によること、基本的には数学的な方法、特に「平面ユークリッド幾何学」によること、文字体の類似関係の設定には異体字も認めることなどが含まれている。

また、指標に関係する数学的なものには図形の回転・反転（合同変換群）とそれを表す八個の直行行列（基本の回転角は $90^\circ$ ）がある。そして、それは②（類似の認定）と③（型分類）に関係するが、密接さは相互に異なる。即ち、前者では加筆（例：[F]→[E]）や弁別点（例：[ハ]→[バ]）と並ぶ部分的なものだが、後者では専らそれに依存し全面的である。

第三章「ラテン系文字の類似と型分類」では、始めに各26個の標準的な大・小二種類のラテン文字を対象にした三指標の一覧表を示した。また、②ではやや複雑な類似認定操作の必要な文字対に限り補助の表も付した。

次に第二章の②（類似の認定）の諸基準に照らした詳細な議論を要する文字対の類似関係の説明が続く。尚、異体字は、大文字ではギリシャ文字の時代にまで溯って歴史的な字形にも考慮が払われているのに対し、小文字では概ね幾何学的なものが考慮の中心となっている。

最後に、派生文字を中心とする実際の標準文字の型の一覧表を表した。大半は、弁別点の付加関係（20種類余、平均10文字弱）である。他に、主に旧式のリズ文字に見られる回転・反転法によって派生した文字や関連字形の無い無地もあるが、数的には多くない。

第四章「外来語用文字」では、各集合の要素（数）を確定する必要から、専ら「基本文字集合」の協会の確定に関する問題を扱った。要するに、当該語文の「（基本）文字は幾つか」ということを考えたのである。

前半では、まずこの問題に対する外来用文字の重要さを広く指摘した。流用した他言語の語彙やそれに近いものに使用されている文字まで基本文字の中に含めるまは難しいからである。ただ一口に外来語と言っても範疇が曖昧である。そこで、始めに第一章で言及した語彙の分類表を用いて対象を調べた。その際、「廃止されたはずの文字」（例：日本文 [ゐ]）等の境界確定に関わる他の要素にも触れた。（ただし、アクセント符号が関係するものは一括して第五章で扱った。）

また、外来語専用文字の構成法（造字法）の分類も行なった。造字法はそれ自体の傾向把握が重要である外に、文字数の確定にも影響する例があるからである。

後半では、ラテン系文字から作った外来語専用文字の一覧表を示した。（実際には、第五章（5-5）で確定したものの一部を表にした。）また、一覧表から簡単な統計を取って造字法の分布も調

べた。例えば、欧州と中国では対照的な分布を示す。そして、この傾向自体は地域毎の文字の導入経過の相違からして予想されないことではないが、やはり定量的に数値化されたことは意義がある。尚、本論の第五章（5－5）や別冊諸表でも同じ分類・項立てにしてある。

第五章「ラテン系文字諸言語の文字集合」は二つの部分から成る。5－5が後半で、他は前半である。量的には圧倒的に後半の方が多い。本論の核心でもある。そのため、文字表や統計データ表に中心にして、かなり部分を「諸表」として別冊に編んだ。

最初に、先ず大文字と小文字が別文字として扱えることを示した。両者の字系間には、例えばキリール文字大文字とラテン大文字くらいの差が認められ、それはキリール大文字とギリシャ大文字との相違より大きいほどだからである。

次に、所謂アクセント符号について考えた。実際の役割は様々で、一部は基本文字数の確定に影響するからである。ここまでが前半である。

今度は、統計標本の対象となる言語別の基本文字集合の確定作業である。始めに、近年に於いてラテン系文字を使用した経験があっても、本論の統計標本として採用しない語文の一覧表を示した。それと共に、採用する場合の基準も示した。

次に各語毎の確定作業の過程を、大文字と小文字の順序で、示した。語文数が多いので、量的には必要最小限に絞った。また、結果の詳細は全て別冊の『諸表』を見ればわかるようになっている。

最後に、確定した基本文字集合を標本として取った三つの指標の統計値の全データを別冊『諸表』に記載した。指標の順序は次の通りである。

- ①「画数」      ②「類似文字数」      ③「文字の型」

第六章「結論」では、第5章（別冊『諸表』）で得られたデータを推測統計学の手法を使って解析し、基本文字の集合について何が言えるか考えた。

前半は、主にここで利用する推計学の説明である。先ず、（必要な範囲内で）簡単に理論の要点を述べ、次にそれを基本文字数の統計データを用いて実際の推定過程の中で説明した。続いて、本論で文字集合の比較用に提案している三指標の推定に適用し、同時にその結果からラテン系文字の全体像を組み立ててみた。

後半では、前半で求めた指標間の相関関係を扱った。中心は基本文字数と三指標間の関係で、始めに相関図を画し、相関係数を計算した。また、相関係数の高い場合には、回帰直線を求めて、図中に各いた。（そして、細かい技術的なことにも触れた。例えば、個別言語の統計値の分布計上の変化具合や良否、相関図で回帰直線から大きく外れている語文を手掛かりに、造字方や基本文字集合の構成方と指標相関の変化の様子などである。）

その主な結果はこうである。即ち、

- (1)「類似文字数」②は大小両文字共に、文字数に高い相関関係を示す。
- (2)「文字の型」③は大文字では高い関係を示すが、小文字は無相関に近い。
- (3)「画数」①には弱い相関関係が認められるに留まる。

また、大小両文字の指標値間の関係は次の様である。即ち、(4)指標③は(一部の傾向<sup>(\*)</sup>を除き)かなり異なるにも拘らず、指標①と②では平均値がほぼ一致し、指標①では分布も極めて近い。

結局、この三指標は文字集合を比較・評価するに有効であると結論できる。

以上が、文字の情報伝達符号という側面に焦点を当て、文字の相互比較方法を探った『文字の統計学的処理』と題する試論の要旨である。第一歩は充分踏み出せたと思う。尚、三指標の内、最も直感的に捉えにくい「文字の型」の有効性と(4)の<sup>(\*)</sup>については拙論『諸文字の幾何学特徴』(参考論文)を参照されたい。そこには、(20字から60字という文字数の制約付きだが)世界中の直線的な字体の諸文字の型分布に、類似の傾向が見られることが指摘されている。

## 論文審査結果の要旨

本論文は、文字を情報通信用の符号として取り扱う立場から、字形面に即して比較しうる方法と理論の構築を目指したものである。六章本論と諸表別冊から成る。

序章では、文字言語の特徴をいくつかの観点から考える。まず音声言語との差異を八箇条にわたって独立性を明らかにする。論者の経験に基く状況証拠を補う。つぎに両極の言語の差異を歴史的に考察し文字言語の分類単位を取り上げ、文字の呼称に触れる。そして文字言語の構成要素を考え文字体系に迫る。書字材料・道具として字形の観点から考察が展開する。文字の未来、現在、過去を概観したのち、書法、字形、用途の面から分類が試みられ、各地域・文字圏ごとに書写材料・道具をあてはめて終る。穏当な記述である。

第一章「文字集合の構造」では、文字単位、文字と符号についての問題点の所在を明らかにし、文字言語の古今に触れ、文字集合の内部構造として句読点、符号を取り上げて分析する。文法標識を示す符号、アクセント符号と分ち書き、分音符号と弁別符号、そして表意文字と基本文字を考察してから文字単位の取り方を問題にする。

第二章「文字集合の統計的標識」では、対象とする字形を、絵的文字を除いた直線的なものに限定すると述べる。歴史上で重要な役割を果たした文字と現在使用中の文字の多くが実際これで扱えるからである。文字の書法、字形、用途の三方面から分類し、後で相互の項目を横糸で繋ぐ。複雑と曖昧の面を明らかにした、それぞれ画数と類似文字数という指標を立て、加えて文字の型と基準文字を導入することによって解決を計ろうとする。それらは何れも、基本的には「平面ユークリッド幾何学」に依っているが、中でも文字の型分類と累次関係の認定にはそれが顕著である。論者は視力検査表をヒントにしているが、そこで使う諸々の(変形)字形と、ここで言う文字の型分類の間にはこれまで確固たる理論上の繋がりが示されたことは多分なかった、ことを明言する。増してや、その型分類が実際の諸文字体系において意味があるなどということが、数字によって定量的に

示されたことはまったくなかった。論者は、平面ユークリッド幾何学とそれを表す合同変換や八個の直交行列など駆使して、実際にそれらの関係を明らかにして見せたのである。こうして、扱う対象は専ら現行かそれに準ずるラテン系文字の諸文に絞られる。もちろん、いわゆる大文字と小文字の両方である。

第三章「ラテン系文字の類似関係と型分類」では、始めに26個の標準的なラテン文字を対象にした三つの指標の一覧表を出す。次に、詳細な議論の要する文字対の類似関係の説明が続き、最後に派生文字を中心とする標準外文字の型の一覧表が付される。

第四章「外来語文字」では、専ら「基本文字集合」の境界の確定に関する問題を扱う。要するに、当該語文の基本文字はいくつかということを考える。先ず、この問題に対する外来語（専用）文字の重要性を指摘し、つぎに諸語文から作った外来語専用文字の一覧表を示す。その他、前半部分では境界確定に関する他の要素にも触れ、後半部分では一覧表から簡単な統計を取って造字法の分布などを調べている。

第5章「ラテン系文字諸言語の文字集合」から核心部に入る。最初に大文字と小文字が別文字として扱えることを示す。例えばキリール大文字とラテン大文字の違いと同程度だということである。次にいわゆるアクセント符号について考える。全てが基本文字に分類されるわけではないからである。そして、最後に統計標本の対象とする言語別の基本文字集合の確定とその簡単な過程を示す。併せて、標本として採用する場合の基準と採用しない語文例の簡単な表を付す。なお別冊『諸表』は論者の独創で、これを見ればラテン系文字語文の文字集合に関することが一目で分かる。

第六章「結論」では、始めに推測統計学の要点を、基本文字数の統計データを用いた実際の推定過程の中で説明する。次にそれを三指標の推定に適用する。そして、それらの結果を基にラテン系文字の全体像を組み立て、併せて文字一般の特性についても考える。以上が前半。後半では、前半で求めた指標間の相関関係を扱う。中心は基本文字数と三指標間の関係で、始めに相関図を画き、相関係数を計算する。また、相関係数の高い場合には、回帰直線を求めて、図中に画く。そのさい、リス文字に特別な配慮をするのは、これが文字系統の異動や文字の一般的特徴を考えるための試金石と観たからという。次に、個別言語の統計データの分布形状の変化具合や良否などを扱って、そして、最後に指標間の相関図での回帰直線から大きく外れている語文を手掛かりに、造字法や基本文字集合の構成法と指標相関の変化の関係を追ってみせる。

以上、文字比較において論者の示した努力と知見は、先行の研究の乏しい現在にあっては、評価されてよい。ラテン系文字を例として、その組合せの中に文字の普遍的特徴を見出そうとする論者の方向は今や定められたと見られる。

よって、本論文の提出者は、博士（文学）の学位を授与されるに十分な資格を有するものと認められる。