

被爆関連文献資料の
フルテキスト・データベースの構築の試み

松 尾 雅 嗣

広島大学平和科学研究センター

宇 吹 暁

広島大学原爆放射能医学研究所附属原爆被災学術資料センター

川 崎 信 文

広島大学法学部

好 村 富士彦

広島大学文学部

田 村 佳 子

広島大学平和科学研究センター

舟 橋 喜 恵

広島大学総合科学部

本研究には、平成3年度広島大学教育研究学内特別経費（「被爆関連文献資料のフルテキスト・データベースの構築と分析のための試験的研究」、代表者松尾雅嗣）の助成を受けた。

A Pilot Study on the Construction of A Full Text Data Base of Documents of Atomic Bomb Damages

Masatsugu MATSUO

Institute for Peace Science

Satoru UBUKI

Data and Specimen Center for Atomic Bomb Disaster,

Research Institute for Nuclear Medicine and Biology

Nobufumi KAWASAKI

Faculty of Law

Fujihiko KOMURA

Faculty of Literature

Keiko TAMURA

Institute for Peace Science

Yoshie FUNAHASHI

Faculty of Integrated Arts and Sciences

S U M M A R Y

There have been various efforts made to compile bibliographies and directories of documents of Atomic bomb damages in Hiroshima and Nagasaki, such as hibakusha's memoirs, testimonies, or so-called "A-bomb literature." There have been, however, few serious attempts to analyze these documents in full detail and make clear specifically human aspects of the damages caused by the Atomic bombs. The present study is an initial attempt at such an academic effort. It aims to create a full text data base of the documents and, on the basis of such a text data base, develop ways to analyze

aspects of the damages in full detail.

In this initial study, we first created machine-readable texts mainly of early documents. The machine-readable texts includes *The City of Corpses* (Shikabane no Machi) by Yoko Ota, *Atomic Bomb Poems* (Genbaku Shishu) by Sankichi Toge, *Experiences of Atomic Bomb* (Genbaku Taikenki), *Voice from Heaven* (Ten yori no Koe). Moreover, several others such as *Hiroshima Diary* (Hiroshima Nikki) by Michihiko Hachiya, Collection of Tanka by Shinoe Shoda, *Nagasaki – Records of Twenty Persons' A-bomb Experiences* (Nagasaki – 22 nin no Genbaku Taiken Kiroku) and so on are now in preparation.

Secondly, to use these texts on a microcomputer, a simple concording and word counting program was developed. It easily and speedily retrieves parts of the text where a given word or string occurs and outputs the results in KWIC or other formats. It also counts the occurrences of given words or strings and shows their distributions among parts or subtexts.

目 次

はじめに

- 1 検索プログラム
 1. 1 概要
 1. 2 検索
 1. 3 テキストの定義と入力
 1. 4 その他の機能
- 2 フルテキスト・データベース
 2. 1 テキストの選定と入力
 2. 2 峠三吉『原爆詩集』
 2. 3 大田洋子『屍の街』
 2. 4 『原爆体験記』
 2. 5 『天よりの声』
 2. 6 『広島市平和宣言集』
 2. 7 『「あの日」の証言』
- 3 プログラムとデータの公開について
 3. 1 基本的方針
 3. 2 検索プログラム
 3. 3 テキスト
- 4 今後の課題

はじめに

広島・長崎の原爆被害に関して、既に膨大な研究の蓄積があることは改めて言うまでもあるまい。しかしながら、原爆被害の解明の努力が、その緊急性重要性からしても、圧倒的に医学的側面に偏っていることも、また、言を俟つまい。また、このような「からだ」の問題だけでなく、他方で、「こころ」の問題が重要であることも否定できまい。被爆者に対する面接をはじめとする様々な社会調査、

被爆者運動の調査、被爆者の手記、証言の収集、編纂、公刊、あるいは原爆文学による表現等多様な形で「こころ」の問題を記録し、調査する試みが行なわれ、現在も継続している。しかし、資料の調査、発掘、収集、編纂を越えて、このような公表、未公表の文献資料記録を分析し、被爆の全体像をいかに解明するかという点に関しては、統計処理に馴染む調査結果や原爆文学と称される作品群の分析、批評が若干行なわれているに過ぎないと言っても過言ではない。

広島・長崎の原爆被爆関連文献資料に関しても、被爆記事目録、被爆手記目録、原爆文学目録等、幾つかの目録作成の作業は行なわれているが、学問的立場から、これら文献資料を比較、分析する試みは、遺憾ながら、きわめて立後れていると言わざるを得ない。この意味で、個々の文献資料を詳細に検討し、(比較)分析することによって、原爆被害の全体像の一側面を解明するという作業は、今日緊急の課題であると言ってよからう。

このような立場に立ち、われわれは、個々の文献資料を詳細に分析し、さらには比較分析することを目的とした。そして、二つの研究目的をを設定した。

ひとつは、個々の文献資料をコンピュータに入力し、フルテキスト(全文)・データベース化することである。被爆関連の文献資料を全文機械可読化する試みは従来まったく行なわれておらず、データベース化することにより、適当なソフトウェアがあれば、多くの研究者あるいは関心ある市民の共有財産とすることができると考えるからである。具体的には、われわれは、初期の代表的被爆手記、証言、原爆文学作品、広島市・長崎市平和宣言などをパーソナル・コンピュータに入力し、フルテキストとしてデータベース化することを当面の目的とした。

他のひとつは、このフルテキスト・データベースを利用して、被爆の実態の一側面に接近する手法を確立することである。このように作成したフルテキスト・データベースを使用した被爆手記、原爆文学などの学術的研究、例えば内容分析手法、語彙統計の手法などによる分析は皆無と言って差し支えない現状だからである。具体的には、フルテキスト・データベースの検索と分析のためのコンピュータプログラムを開発することが目的である。

上述の研究状況に鑑みれば、被爆手記、証言、文学作品などをテキスト・データベース化し、その検索、分析の方法を確立することは、学術的にもきわめて重

要であり大きな意義を有するものといえる。

所期の目的を十分に達成したとは到底言い難いが、以下の本研究の成果を報告する。

1 検索プログラム

1. 1 概要

フルテキスト・データベースを構築するためには、プログラムの作成は不可欠と言える。既成のパソコン用プログラムの使用も不可能ではないが、文字データの検索、用例付き索引作成などの目的を迅速に達成できるものはない。プログラムの開発に当たっては、文字列の用例の迅速な検索を最優先の課題とし、操作性を第二の課題とした。また、移植性を考慮し、ソース・プログラムはC言語を用いた。このようにして開発したのが、漢字テキスト検索システムKRである。ここで言う漢字テキストには、漢字仮名交じりのテキスト、正確には2バイトコードで構成されたテキストを含む。このプログラムは、以下の特徴をもつ。詳細に関しては、本号所収の松尾論文を参照。

このプログラムは、テキストの検索を主目的とするものであり、文字、語彙の数量的分析、文法的分析は対象としない。但し、補助的機能として、以下の機能を有する。

- サブテキストの作成
- 文字の度数つきリスト作成
- 文字列の度数算出
- グループ別文字列頻度リストの出力
- テキストの外部出力
- KW I C索引の一括作成

KRは、以下の環境で実行可能である。

OS	MS-DOS (2.xx 以前に関しては保証の限りでない)
対応機種	NEC PC 9801 シリーズ
メモリ	640K

ハードディスク、プリンタは必須ではないが、あることが望ましい。また、モニターは、カラー、モノクロいずれでも可。また、その目的からして、日本語フロントエンド・プロセッサは不可欠である。画面 20 行で使用でき、画面の最下行を使用した後は、以前表示されていた行を復活できることが望ましい。

KR の実行開始には、MS-DOS のプロンプトに対し、

KRMN

というコマンドを投入する。すると、KR のメインメニューが表示されるので実行するメニューを選択する。以後、KR は原則としてすべてメニューにより実行する。メニューによらないで実行することも勿論可能である。KR のメインメニューでは、以下のいずれかが選択可能である。

- S : 検索
- D : テキスト定義
- I : テキスト入力
- T : サブテキスト
- U : ユーティリティ
- Q : 終了

以下、各実行メニューについて略述する。

1. 2 検索

KR は文字単位の処理システムである。従って単語単位での処理はない。このことは、一方で、テキストの入力に際し、単語に切り分ける必要がないこと、検

索に際し、文字、単語の区別なく任意の文字列を検索できることという利点をもつが、他方で、検索用のデータが肥大化すること、単語単位での検索、分析ができないことといった欠点ももつ。

KRでは、通常の文字列の検索のほか、二つの文字列を指定して、その対を検索することもできる。また、ワイルドカードの使用も可能であるが、文字列単位の検索を行なうので文字列の前後では意味をもたない。また、後述のサブテキスト処理を行えば、検索範囲をテキストの特定の、必ずしも連続していない、範囲に限定することもできる。

検索結果は用例集合として保存され、検索の実行中は必要に応じ、画面、ディスク、プリンタに何度でも出力可能である。しかし、検索結果の出力は、利用者がワープロソフト等で再度編集することを前提としているので、印刷機能はきわめて貧弱である。出力は、所謂KWIC形式とKWOC形式（KRではコンコード形式と呼ぶ）のいずれかが選択できる。後者の形式では、当該の文字列の出現する行の前後の行も文脈として出力できる。また、検索結果に対し、論理積と論理和の演算により、新たな用例集合を作成することもできる。検索結果に対し、対象となった文字列の前あるいは後ろの文字列により、並べ替えることもできる。

検索、結果の出力に関しては、利用者が必要に応じて変更できるパラメータが用意されている。

次の例1に、『原爆体験記』における「皮」のKWIC形式での検索出力例を掲げる。

1. 3 テキストの定義と入力

KRは、通常テキストファイルのまま検索するわけではない。KRで検索を行なうためには、対象とする文献資料等のテキストを予めMS-DOSテキストファイルとして作成しておき、これを変換して検索専用のファイルを作成しなければならない。この作業ををKRへの入力と称する。

KRでは、テキストは、一般に複数種のテキスト構成要素（これをテキスト・ユニットと称する）から成る。例えば、行、頁、段落、巻などがそれである。被

] 顔に異状を覚えた。／ぬぐった顔の皮膚がズルッとほがれた感じにハッとした。／ 落下傘 13
 ら指の先までズルズルにむけて、その皮膚は無気味にたれ下っている。／左手は手首か 落下傘 13
 手は手首から先、五本の指がやっぱり皮膚がむけてしまってズルズルになっている。 落下傘 13
 ってくるような鈍痛だった。／両手の皮膚のむけたところは黄色な分泌物がにじみ出て 落下傘 16
 たトマトを突きくずしたようになって皮膚ができなかった。／床にやっとな身を起せるよ 落下傘 19
 途中釘が服に掛った。／皮膚をカスった。／痛さは全く感じなかったが、 汚水の味 39
 くぶくに腫れ上がっていた。／両手の皮がハゲてたれかかり、見る影もな | く | い | あ 汚水の味 41
 の児の背中、いたんで黒ずんだびわの皮をグルリとむい | た | だ | ように、皮膚はペロ ケロイド 47
 その度に薄桃色の肉のようなものや、皮膚はペロリとたれ下がり、私は思わず顔をそむ ケロイド 47
 びとの姿は凄惨を極めていた。／背の皮などが | ついて | 相ついて | はがれ、下から少 ケロイド 48
 ぶらさげているかのように。／そして皮は全部はがれ、腰からぶらさがっている。／あ 硝子破片 55
 そして大抵の人たちには、黒く焦げた皮をはがれた背中からは鮮血が幾すじも流れてい 硝子破片 55
 っておるぜ」／川べりの所には顔の皮膚の破れ目から、赤い血潮が流れていた。／も 原爆に日 74
 になった肩からひじ、手の甲へかけて皮が「ずべり」とむけて、下顎のあたりから垂れ 消えぬ悔 93
 いは傷をうけ、むごたらしくむくみ、皮膚がすっかりまくれ、その端はぶらんと垂れ下 親切 98
 爛れ、身体は衰弱 | のため | に | 骨と 皮は垂れ下がり、物凄い出血は埃にどす黒くな | 親切 98
 めのうち、やっとな密着したばかりの皮ばかりと | なった | なり | 。／自他ともに永く 親切 102
 異様に大きくふくれあがり、露出した皮膚が割けはせぬかと、幼児のように障子につか 親切 新 103
 | 違 | 等 | は激しい痛みを訴え、裸の皮膚は夕立に洗われた空からの、まぶしい正午の 還らぬ魂 110
 ひどい火傷だ。／皮膚はそれに触れるたびに、ぬるぬると潰えて | 還らぬ魂 112
 で | 白血球数千以下となり | 発疹様の 皮膚を | ぶらりとさげ | ぶらさげて | 倒れている 二日間 新 128
 を訴える。【】背部と腹部には多数の皮下出血が現れたら、ほとんど例外なしに死の転 奇蹟妻 152
 来連続し、強心剤も怠らなかつた。／皮下出血が現れた。／ | 彼女 | 妻 | は前膊のこの 奇蹟妻 153
 小さくなるように思われるが、夏期の皮下出血の症状が絶えてからも止血剤の注射を続 奇蹟妻 154
 いた。／ほとんどの者が焼けただれた皮膚充血期には発赤が濃くなり、やや | 膨隆 | 膨 奇蹟妻 155
 /指さすその左肩を見ると、さけた皮膚を | ぶらりとさげ | ぶらさげて | 倒れている 重傷婦人 158
 に会う。／薬一つ塗っておらず真黒い皮膚の上に肉 | が盛り上がって | を盛って | いる 重傷婦人 161
 のかと思つたが、それは爆風に裂けた皮膚をたらし「隊長異常なし」とく敬礼をする 重傷婦人 162
 マーク【】ロを塗って貰ったから、皮膚が雨にさらされて真白くなって、ブラブラと 浩ねむれ 170
 手をあげたシャツの破れからダラリと皮を剥がれた兎のようで、どんなに見ても助から 浩ねむれ 171
 /顔といわず足といわず泥まみれの皮膚が垂れ下がり、髪はザンバラで、特に目を引く 黒雨 新 183
 してその上に涙がついたのか、異様な皮膚をあらわにしながらふるえている親子がいる 黒雨 新 183
 となっています。／剥けて十センチも皮のブラ下った両手を膝の上にして、入口の階段 愛児捜し 新 221
 わり流れ来る。／右手の甲を見れば、皮膚はやけどしたかの如く皮が剥がれている。ま ビンタ 新 233
 を見れば、皮膚はやけどしたかの如く皮が剥がれている。／また凄くそこが熱走る。 ビンタ 新 233
 んのことか全然見当がつかぬ。／手の皮膚はますます激しく痛み、頭の亀裂はずきずき ビンタ 新 233
 中に背負うために身体を差上げると、皮膚はズルズルとすべり、その下から赤く脹れた ビンタ 新 234
 物も洋服もボロボロに焼落ち、身体の皮膚は大きく脹らしている。／口も鼻も目もどこ ビンタ 新 234
 だけで、丸裸というてよいだろう。／皮はやぶれて、ところどころ、ぶらぶら下り、髪 妻よ許せ 新 241

爆手記集であれば、個々の手記もテキストユニットの例である。テキストユニットの個々の現われ（トークン）には、それを識別するための番号あるいは文字型の値を与える。この値が、検索された用例に付される索引項目となる。

すべてのテキストユニットは、さらに、行から成る。この行は、後述の入力ファイルの1行に相当する。

KRに入力するテキストは、MS-DOS テキストファイルとして作成するが、このようなKRの要請するテキストの構造を満足するものでなくてはならない。ただし、テキストの長さに制限はない。テキスト作成に当たっては、利用者が処理単位としたい文、段落などをテキストファイルの1行とする。また、索引に使用するテキストユニット、例えば、頁、章、節、巻、号等々の区切を示しておく必要がある。さらに、テキスト・ユニットのトークンを識別するために、頁番号などの数値あるいは文字型の値を与えておく必要があるが、これはすべての行に与える必要はないし、連続した数値の場合、初期値だけ与えればよい。この数値あるいは値が用例の出現個所を示す索引項目として使用される。

また、このようなテキストユニットの番号や値を利用して、後にテキストの特定の部分だけを対象とした処理を行なうことができる。このようなテキストユニットは最大6個まで指定できる。上述のように、KRは文字単位の処理を行なうので、入力データで単語、分節等を区切る必要は一切ない。

この説明からも明らかなように、実際には、入力データを作成する前に、何をテキストユニットとするかななどを定めておく必要がある。そして、さらに、KRのメインメニューの「テキスト定義」で、KRにテキストの名称、索引項目の名称などを知らせておく。

実際のKRへのテキストの入力は、メインメニューの「テキスト入力」により行なう。この場合は、テキストの名称と入力ファイル名を与えればよい。追加入力は何度でも可能であり、従って、入力されるテキストは、複数のファイルに分割されていてもよい。

1. 4 その他の機能

KRは用例の検索を主たる機能とするが、いくつかの補助的機能を備えている。

また、この種の補助的機能は、容易に追加、拡張することができる。以下、KRの補助的機能について概略を述べる。

① サブテキスト

このメニューは、フルテキストから、テキストの特定の部分であるサブテキストを設定する。サブテキストは、テキスト中で連続している必要はない。一つのテキストについて、最大20までのサブテキストを設定することができる。サブテキストは、相互に重複して差し支えない。

サブテキストの設定は、テキストユニット、あるいはテキスト行を単位として行なう。

設定したサブテキストについては、検索および他の処理を、特定のサブテキストに限定して行なうことができる。

② テキスト情報

サブテキストの作成がメインメニューで行なわれるのに対し、以下の処理は、メインメニューの「ユーティリティ」のサブメニューとして実行される。「テキスト情報」は、対象とするテキストの文字数、テキストユニットの名称などの情報を画面、ディスク、あるいはプリンタに出力する。

③ テキスト出力

テキストあるいはサブテキスト全体をディスクに出力する。出力には任意の索引項目を付すことができる。また、索引項目の位置を指定することもできる。これにより、例えば、巻、頁などの値を付したテキストを作成することができる。

④ 文字リスト

テキストあるいはサブテキストについて、使用されているすべての文字のリストをディスクに出力する。コード順、度数順（昇順、降順）の選択が可能であり、また出現度数をつけることを省略することもできる。

⑤ 文字列リスト

テキストあるいはサブテキストについて、利用者の与えた文字列（文字であっても単語であってもよい）の出現度数を、ディスクに出力する。文字列の配列は、利用者の与えた順あるいは度数順（昇順，降順）のいずれかである。文字列の入力はディスクファイルから行なう。

⑥ グループ別文字列リスト

任意のテキストユニットあるいはサブテキストごとに、利用者の指定した文字列（文字であっても単語であってもよい）の出現度数を、ディスクに出力する。文字列の配列は、利用者の与えた順である。出力は、テキストユニットと与えられた文字列とのクロス表となる。

⑦ KWIC索引一括出力

テキストあるいはサブテキストについて、利用者の与えた文字列（文字であっても単語であってもよい）の用例のKWIC索引をディスクに出力する。文字列の配列は、利用者の与えた順である。出力の形式，例えば，出力1行の長さなど，はパラメータを変更することにより，かなりの程度制御できる。

2 フルテキスト・データベース

2.1 テキストの選定と入力

既に述べたように、対象とするテキスト全文を機械可読化することが本プロジェクトの基本方針であり、入力もこの方針に従った。

入力の対象とするテキストの選定に関しては、定評ある代表的なものを第一の規準とした。また、できるかぎり初期のもの、特に昭和二十年代のものを優先することとした。このような規準により、選定し、入力したテキストは、以下のとおりである。これらのテキストの利用に関しては、次節を参照されたい。

『原爆詩集』（峠三吉）

『屍の街』(大田洋子)

『原爆体験記』

『天よりの声』

『広島市平和宣言集』

『「あの日」の証言』

なお、複数の版がある場合には、必ずしも初版に拠らず、テキストとして最適と判断したものを使用した。詳細については、それぞれのテキストに関する項で述べる。

具体的入力に際しては、原則として以下の方針に従った。

- ① 目次、献辞は除外する
- ② 手記、詩の表題、章、節の名称、(小)見出し等は、テキスト本文として扱わず、索引項目とする。
- ③ ルビ、傍点は無視する。
- ④ アルファベット、漢数字はすべて全角文字とする。
- ⑤ 字体、仮名づかい等は原則としてテキスト本文に従う。
- ⑥ 明らかに誤植と思われる部分は訂正した。これには、仮名づかいの混用、旧字体と新字体の混用も含む。詳細は、各テキストに関して述べる。
- ⑦ 改行は、原則として文単位に行なったが、小見出し、項目の羅列のように、本文で改行されている場合にはそれに従った。1文が頁をまたがる場合、与えられる頁番号はその文が始まる頁を示す。

以下、既に入力済のテキストに関して、底本など特記すべき事項を略述する。

2. 2 峠三吉『原爆詩集』

1952年6月15日青木書店刊青木文庫版『原爆詩集』を底本とした。

本文と「あとがき」のみを収録し、冒頭の献辞、目次、中野重治による「解説」

は省略した。

また、詩が複数の連から成るとき、本文では2行改行により連の区切が示されているが、これは無視した。連に番号が付けられている場合（「その日はいつか」）これも無視した。

索引項目としては、詩の題名と頁が参照できる。以下の詩については、略称を用いた。

仮縋帯所にて	→仮縋帯所
倉庫の記録	→倉庫記録
ととったお母さん	→お母さん
ちいさい子	→ちいさい子
河のある風景	→河の風景
一九五〇年の八月六日	→一九五〇
ある婦人へ	→婦人へ
その日はいつか	→その日は

2. 3 大田洋子『屍の街』

昭和25年5月30日冬芽書房刊の同名の作品集所収の『屍の街』（同書17頁－218頁）を底本とした。以下これを冬芽版と称する。本作品の初版は、昭和23年11月10日同名の単行本として中央公論社より刊行されている。この初版と冬芽版とは大幅な異同があり、これを含めたKRテキストは近く作成予定である。

明らかに誤植と思われる部分は訂正した。これには、歴史的仮名遣いととの混用をも含む。一覧は本節末尾に掲げる。また、僅かではあるが、漢字が新字体になっている個所もある。これはそのままとした。

索引項目としては、章の名称、節番号、頁番号を与えた。章の名称については、以下の略称を用いた。

鬼哭啾啾の秋 → 鬼哭啾啾
 運命の街・広島 → 運命の街
 街は死體の襤褸筵 → 襤褸筵

冬芽版の誤植の一覧を以下に掲げる。

頁	原文 (誤)	正 (修正後)
4 2	黄硫島	→ 硫黄島
8 1	火の不始末よ火。くらい消して出ればいいのに。	→ 火の不始末よ。火くらい消して出ればいいのに。
8 6	砂原へ りる途中の	→ 脱字
8 8	支えられてね	→ 支えられてね
1 0 1	よろび合う	→ よろこび合う
1 0 9	いるのかもれない	→ いるのかも知れない
1 5 4	焼けてよますよ。	→ 焼けていますよ。
1 5 5	出會はない	→ 出會わない
1 5 6	やうに	→ ように
1 6 5	ゐる	→ いる
1 7 8	九日中旬	→ 九月中旬
1 8 3	いろんな人ちた	→ いろんな人たち
1 8 9	市長代理の出迎へ	→ 市長代理の出迎え
1 9 0	威力のあとを見に來え	→ 威力のあとを見に來た
	W. Hローレンス	→ W. H. ローレンス
2 0 2.	黄金の波を打っていた。一面の稲田も	→ 黄金の波を打っていた、一面の稲田も
2 0 9	[表題] 秋晩の琴	→ 晩秋の琴
	水久に	→ 永久に
2 1 4	琴瓜	→ 琴爪

2. 4 「原爆体験記」

昭和25年8月6日広島民生局社会教育課編による『原爆体験記』が広島平和協会より刊行された。この手記集（以下旧版と称する）は、市の求めに応じて応募された手記164編（「刊行のことば」）より、18編を選んで収録したものである。その後、被爆20周年に当たる昭和40年7月20日、この手記集は同名の表題で広島市原爆体験記刊行会（代表広島市長浜井信三）編として朝日新聞社より再刊された。これを新版を称することにする。

この「再版」に当たり、応募手記のうち旧25年版に収録されていなかった11編が新たに収録されるとともに（新版「刊行のことば」への追記）、大江健三郎による巻末エッセイが追加されている。

この新版は、昭和50年に、朝日選書の1冊としてさらに再版されている。これが現行版である。この版と新版の異同は、厳密に点検したわけではないが、基本的にはないものと考えてよい。なお、現行版を参照するときには、新版の頁数に2を加えたものが現行版の頁数となる。

入力はこの新版を底本とし、献辞、目次、「はじめに」（浜井信三による新版への序文）、「刊行のことば」（旧版の再録）、大江健三郎による巻末エッセイ、「刊行のいきさつ」とそれに対する付記を除くすべての手記を忠実に再現した。ただし、手記中、執筆者の「被爆場所、職業、被爆状況」に関する略述とその後の消息に関する編者の付記は除いた。

これに加えて、新版と旧版の間には、新たな手記の追加にとどまらず、多くの異同があるので、本テキストでは、二つの版の異同を以下の二つの形で示した。異同の詳細は本節末尾で述べる。

ひとつは、新版に新たに収録され、旧版には収録されていない部分であり、これは索引では、「新」の表記で示す。これには、新版で新たに収録された手記の全体と、旧版に収録された手記のうち、同版で削除、省略され、新版に収録された部分を含む。

逆に、旧版にあり、新版では、理由は詳らかでないが、削除されている部分も僅かだが存在する。これは、「旧」として示す。

その他の、句読点、送り仮名、仮名遣い、新旧字体の相違を除く、部分的な異

同は、次の方法により示す。

| x | y | x : 新版
 y : 旧版
< > 新版にのみあり
[] 旧版にのみあり

原稿そのものあるいは印刷段階に由来すると思われる明らかな間違いも原則としてそのままにした。変更箇所は以下のとおり。

新版 87頁 来て →着て
 197頁 お母さのん→お母さんの

索引項目としては、手記表題、筆者、新版頁、版（新旧で示す）、旧版頁が使用できる。筆者と手記表題に関しては、以下の略称を使用した。

手記表題	筆者	手記略称	筆者略称
あっ、落下傘だ	北山二葉	落下傘	北山二葉
爆心に生き残る	野村英三	爆心生残	野村英三
友に助けられつつ	土井貞子	友に助け	土井貞子
汚水の味	奥村昌司	汚水の味	奥村昌司
ケロイドを残して	金谷満佐子	ケロイド	金谷(満)
ガラスの破片に想う	喜多輝子	硝子破片	喜多輝子
めしいとなりて	三木正	めしい	三木正
兄も見違えた顔	前田正弘	見違えた顔	前田正弘
原爆に遭った日	杉本直治郎	原爆に日	杉本(直)
「失明」の悲しみから	沖土居春子	失明	沖土居
消え去らぬ悔	木島克巳	消えぬ悔	木島克巳

忘れ得ぬ親切	橋本くに恵	親切	橋本(く)
還らぬ魂	陸勝利	還らぬ魂	陸勝利
父の死	高木輝彦	父の死	高木輝彦
恐怖の二日間	新保英夫	二日間	新保英夫
十六キロの遠い道	重道昇	遠い道	重道昇
奇蹟に生きる妻	宇都信	奇蹟妻	宇都信
重傷の婦人を負う	門田武	重傷婦人	門田武
奇怪な火炎	手塚良道	奇怪火炎	手塚良道
浩よ、ねむれ	桧垣兵市	浩ねむれ	桧垣兵市
ああ、父と母	長野ふみ子	父と母	長野(ふ)
黒雨をついて	益信之	黒雨	益信之
師とともに泳ぐ	中前妙子	師と泳ぐ	中前妙子
腕に釘が……	山下智万	腕に釘が	山下智万
子の屍を焼く	岩本伯三	子の屍	岩本伯三
愛児を捜して	堤実三	愛児捜し	堤実三
狂った運命	山科光枝	狂た運命	山科光枝
ピンクのあとで	森脇昭幸	ピンク	森脇昭幸
妻よ、許せ	伊藤勘六	妻よ許せ	伊藤勘六

上述のように、新版と旧版の間には多くの異同があるので、以下にまとめておく。

新版と旧版の間には、仮名遣い、送り仮名、漢字の字体、句読点の変更、ルビの使用、広島外の読者のための注記、傍点など編集上の変更点のほか、以下の異同がある。

- ① 新版には、旧版に収録されていない11編の手記が新たに収録されている。
(再録された「刊行のことば」への新版における付記参照)
- ② 手記掲載の順序も二つの版では異なる。次項③を参照

③ 双方に収録されている手記の表題、筆者名には異なるものがある。

以下に異なる個所の対応表を示す。新版の丸抜き数字は、新たに加えられた手記を除く、新版での掲載順を示す。

旧版		新版
爆心に浴びる	野村英三	②爆心に生き残る
汚水の味	奥村昌司	④
落下傘だよ	北山二葉	①あッ、落下傘だ
思い出のケロイド	金行満子	⑤ケロイドを残して 金谷満佐子
硝子の破片からの回想	喜多輝子	⑥ガラスの破片に想う
原爆にあった日	杉本直治郎	⑨原爆に逢った日
帰らぬ魂	伊達克影	⑫ 陸勝利
忘れ得ぬ親切	橋本くに恵	⑪
めしいとなりて	三木正	⑦
山中の一夜	土井貞子	③友に助けられつつ
奇怪な火災	手塚良道	⑬
胸に秘めた悲哀	桧垣兵市	⑬浩よ、ねむれ
奇蹟に生きる妻	宇都信	⑭
兄も見違えた顔	前田正弘	⑧
降り注いだ本	佐々木光憲	⑬父の死 高木輝彦
消えさらぬ悔	不鳥勝文	⑩ 木鳥克巳
電柱に伝言を	門田武	⑮重傷の婦人を負う
黒雨について	益信之	⑯

筆者名の変更については、匿名ではなく実名使用を原則としたためと考えられる（新版付記，259頁）。表題の変更については新版に特に説明はない。可能性としては、新版が何らかの理由で元の表題を変更した、あるいは旧版が表題を

変更したのを新版で元に戻したのいずれかであろうか。例えば、旧版の「降り注いだ本」は、新版では、表題、筆者名とも変更されている。新版の注記（122頁）に筆者の消息不祥とあることからして、表題の変更が筆者の意志によるものでないことは明らかである。

④ 旧版は、個々の手記について全文を収録したものではない。このことは、旧版の「ぬきがき」に「(同人)」とか、収録された手記の筆者名が載せてある（旧版、93頁、106頁、122頁）一事からも明らかである。また、同一手記に関し旧版になく、新版にある行、文、段落が多々見られる。

ただし、旧版の上述の3つの「ぬきがき」に関して言えば、桧垣兵市分の「ぬきがき」（旧版、93頁）は、新版にも全く収録されていない。宇都信分の「ぬきがき」（旧版、106頁）は、新版に全文収録。門田武分の「ぬきがき」（旧版、122頁）は一部分のみ収録。この理由は不明である。

また、双方に収録されている部分についても、多くの場合編集上の理由と思われる理由から、細かな異同は無数に見られる。

2. 5 「天よりの声」

底本は、日本YMCA同盟出版部（YMCA出版）、1983年6月10日刊、末包敏夫編『天よりの声 — ヒロシマ・被爆二年目の手記』である。本書は、1949年4月10日に『天よりの大いなる声』の書名で東京トリビューン社より初版が刊行されたが（83年版、172頁）、同年ただちに紙型を改めた改定版が出版された（同、176頁）。テキストとした新版は、この版を元に、表題、字体などの若干の変更を加えたものである（同、186頁）。

入力に際しては、以下の個所を省略した。いずれも手記そのものではないことと、手記の執筆された1947年（同上169頁参照）とは、執筆時期が異なることによる。ただし、「萌えいづる青草一跋に代えて—」は、テキストファイルとして入力済であるので、必要ならば、検索用テキストに容易に追加できる。

萌えいづる青草一跋に代えて— (末包敏夫)

改定版によせて (末包敏夫)

「天よりの声」によせて (植村環)

ひとりでも多くの方に (相原和光)

新版あとがき (末包敏夫)

手記本文に関する変更等は以下のとおり。

- ① 引用はそのまま本文と同じ扱いとした。例えば、「契約の虹」冒頭の聖書からの引用。
- ② 括弧 () 内に小活字で与えられた部分も本文とまったく同様に扱った。
- ③ 次の個所では旧字体は新字体に改めた。
68頁 坐って番をしていた → 座って番をしていた
- ④ 次の個所は、誤植とも思われるが、原文のまま残した。
27頁 私たちとあとに残された者たちは、ひたすらに先生のお帰りを待って待ちつづけた。(傍点筆者)
正しくは、「私たちあとに残された…」か?
- ⑤ 文中の小見出しはそのまま本文として残したが、前に☆印を付けて示した。

検索の際の索引項目は、手記表題、筆者、頁である。手記表題と筆者の一部については、以下の略称を用いた。

生ましめんかな	生ましめ
ぼくは泣きながら走った	走った
ぼくの行くところがなくなった	行く

火の海の中で	火の海
万斛の涙をのんで	万斛の涙
重い土壁の中に 河元きくの	重い土壁 河元き
あの子は死んでない	あの子
妹を返して	妹
虚しきを仰がず	虚しき
孤独，恐ろしい孤独	孤独
悲しい治療	治療
家族六人の墓標	墓標
死の陰を歩む	死の陰
心の焦土から	心の焦土
残された者の誓い	誓い
つぐないの道 木村文太郎	つぐない 木村文

2. 6 『広島市平和宣言集』

底本は、広島平和文化センター編集発行『平和宣言集』である。但し、近年については、広島市が毎年配付している「平和宣言」によった。この『宣言集』が、特に初期の平和宣言の底本として、資料としてどの程度の信頼性をもちうるかに関しては確認していない。

索引項目は、元号、西暦年号、市長名である。

2. 7 「『あの日』の証言」

底本は、日本原水爆被害者団体協議会刊『日本被団協原爆被害者調査資料集 I 「あの日」の証言（その1）』（1988.11.28），『同資料集 II 「あの日」の証言（その2）』（1989.3.28）』である。

上記資料集は、日本被団協が1985年11月から1986年3月に実施した「原爆被害者調査」の間4に対する自由記述回答の中から各500件合計1000名分を選んで公刊したものである。内容等詳細に関しては、上記資料集の「発行にあたって」、「凡例」を参照されたい。

入力テキストは、「＜原爆と人間＞研究会（一橋大学）」の御好意により MS-DOS テキストファイルの形で提供を受けたものである。提供にあたり、便宜を図っていただいた一橋大学教授浜谷正晴、助手井上由紀、事務職員沼崎保宏各氏に厚くお礼申上げる次第である。

KRへの入力に際しては、以下の変更を行なった。なお、今回提供を受けた MS-DOS テキストファイルは上記資料集に含まれた誤植等を修正したものであり、資料集そのものとは若干の異同がある。

- ① 原テキストでは、「証言」は、被爆地、被爆状況、被爆時の年齢、性別に従って分類配列されているが、KRテキストでは、すべての「証言」を整理番号順に配列した。
- ② 原テキストでは、各証言に続けて、被爆地、被爆状況、性別、被爆時年齢、整理番号が付されているが、KRテキストでは、これらはすべて索引項目とした。従って、KRによる検索結果の表示の際には、それぞれの証言についてこれらの項目をすべて表示することができる。
- ③ 原テキストの半角文字はすべて全角文字に変更した。
- ④ 原テキストでは、特定される人名はすべて記号◇△等で表記されているが、提供を受けたファイル中には若干特定可能な個人名が残っている。これは、原テキストの作成方針に従いすべて記号で表記した。
- ⑤ KRテキストの1行は、原テキストの1段落分である。

検索時の索引項目は以下のとおりである。

被爆地　：広島または長崎。「広島長崎」もあり。

被爆状況：直爆の場合、爆心からの距離を0.5kmきざみで示す。

3 kmを越えるものは、「3.0-」と表記。

その他の場合は、入市、救護、特例地域及び不明のいずれか

性別

被爆時年齢

整理番号

3 プログラムとデータの公開について

3. 1 基本の方針

本研究で開発したプログラムおよび入力したテキスト、今後入力するテキストは、以下の条件が満たされる限り、原則としてこれを貸与し、研究者の利用に供する。

- ① 営利目的には一切使用しないこと
- ② 利用に関わる一切の責任は、利用者が負い、本プロジェクトは一切の責任を免れること
- ③ 利用のサポートは一切行わないこと
- ④ 利用の成果を公表する場合には、本プロジェクトより提供されたプログラムもしくはテキストを使用し旨を銘記すること
- ⑤ 第三者に、貸与、譲渡しないこと

3. 2 検索プログラム

漢字テキスト検索プログラムについては、本号に別稿として掲載されるマニュアルとともに、無条件で公開する。但し、マニュアルは、MS-DOS テキスト・ファイルとしてのみ提供し、必要に応じ利用者が印刷するものとする。また、要

求があれば、ソースプログラムも提供する。ソースプログラムは NEC PC-9801 用の Lattice C, Version 3 の所謂ラージ・モデルで書かれているので、他のコンパイラ、他のバージョンに移植するには変更が必要である。移植に当たっては、64Kバイトを越えるメモリーを操作する点に注意が必要である。

3. 3 テクスト

本研究で作成したテキストは、上述KRで検索可能な形（KRテキスト）、KR入力用のMS-DOSテキストファイル（ソーステキスト）、いずれの形でも提供可能である。いずれの形であれ、著作権、版権等に関わる問題は、利用者の責任において処理するものとする。

KRテキストの場合、用例索引の一括出力機能は使用できないことを除き、他の制限はない。

ソーステキストの場合、KRに入力するための様々の記号が含まれているので、利用者は必要に応じ、これらを変更あるいは削除して使用しなければならない。

4 今後の課題

研究の期間、予算等を考慮に入れれば、被爆関連文献資料のフルテキスト・データベース構築のパイロット・スタディという本研究の所期の目的は果せたと言える。しかしながら、これは飽くまでも、研究が端緒に着いたことを意味するだけであって、今後多くの課題が残されていることは言うまでもない。

第一に、被爆関連の文献資料と言う以上、本研究でデータベース化された資料はあまりにも少ない。少なくとも昭和20年代に公表されたものは、網羅する必要がある。また、公表されてない、私的な重要資料のデータベース化も、当然課題となろう。

第二に、これはデータベース化された資料の利用可能性とも関連するが、昭和20年代の文献資料を扱う場合、新旧字体、仮名遣いの問題もある。入力を外注するとき、旧字体、旧仮名遣いのテキストを外注することは問題がある。また、特に文学作品の場合、原文の字体、仮名遣いを無視することは、ときに原文の趣

を大幅に損ねかねない。

これは、三番目の課題である語彙あるいは表記の管理（情報検索の用語で言えば語彙の統制）の問題である。例えば、ごく単純な例でも、被爆者が水を求める用例を検索し、分析するとき、「水」、「みず」、「ミズ」のすべてを考慮に入れなければならない。このような単純な例ならまだしも、一般には、同義語、類義語、活用形、あるいはすべての表記のヴァリエントを考慮に入れる必要がある。とはいえ、入力テキストを何らかの形で標準化したのでは、原文に忠実であるという、フルテキスト・データベースの意義は失われてしまうこと必定である。

ひとつの解決法としては、かつてテキストの内容分析プログラムとして開発された「ジェネラル・インクワイアラー (General Inquirer)」のごとき、被爆に関わるシソーラスないしは用語辞書を作成し、テキストの入力、検索、分析をすべてこれを介して行なうことができるようにすることであろう。このために、多くの文献資料を入力し、分析するという過程を積み重ねる必要がある。

第四に、本研究で開発したコンピュータ・プログラムは、被爆関連資料を分析するための基礎的ツールにすぎない。次の段階では、より進んだ分析のツールあるいは、既存のSAS、SPSSなどの統計処理プログラムとのインターフェイスを開発する必要がある。

第五に、本研究の第二の目的の中心を成す、フルテキスト、就中被爆関連資料のフルテキストの、分析手法の開発は、ほとんど進捗を見なかった。これが今後の最大の課題である。