

GEORGIA INSTITUTE OF TECHNOLOGY

OFFICE OF RESEARCH ADMINISTRATION

RESEARCH PROJECT TERMINATION

Date: June 15, 1970

Project Title: Computer-Assisted Translation of Books into Braille

Project No: B-1805

Principal Investigator: Dr. Vladimir Slamecka

Sponsor: Atlanta Public Schools

Effective Termination Date: June 30, 1970

Clearance of Accounting Charges: Completed by February 18, 1970

Final Report Submitted: December 30, 1969

COPIES TO:

Principal Investigator
School Director
Dean of the College
Director of Research Administration
Associate Controller (2)
Security-Reports-Property Office
Patent and Inventions Coordinator

Library, Technical Reports Section
 Rich Electronic Computer Center
 Photographic Laboratory
 Terminated Project File No. B-1805
 Other _____

B-1805

COMPUTER-ASSISTED TRANSLATION OF BOOKS INTO BRAILLE

Final Report

Prepared for the
ATLANTA PUBLIC SCHOOLS

December 30, 1969

JAN 3 1970

SCHOOL OF INFORMATION SCIENCE
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332

PREFACE

The School of Information Science, Georgia Institute of Technology, is pleased to submit to the Atlanta Public Schools its final report on a study of the proposed restoration of a library in braille.

The principal objectives of this study were threefold: to examine and evaluate existing and promising approaches to computer-assisted translation of inkprint texts into Grade 2 braille; to recommend and outline the design of a system suitable to accomplish such a translation; and to advise on the organization and management of the project.

During the course of its three-month study the School has received a considerable assistance from several organizations and individuals. We are particularly grateful to Dr. Jarvis Barnes, Assistant Superintendent for Research; Dr. Curtis Henson, Assistant Superintendent for Instruction; Mr. Boozer, Director, Exceptional Children Pupil Services; Dr. Arthur Lownes, Program Coordinator; Mrs. Patricia Richards, Braille Librarian; Mr. John Bales, Controller; and Mr. Tom McConnell, Director, Information Services, and others of the Atlanta Public Schools. Valuable assistance was also received from Mr. Lee Steele, State Board of Education Library for the Blind and his staff; and from Professor R. W. Mann and his associates at the Massachusetts Institute of Technology.

The project team of the School of Information Science included Professors A. P. Jensen, V. Slamecka, M. Valach; graduate research students E. M. Pass and Sidney Heidt; and undergraduate Dov Delatici.

Respectfully submitted,

VLADIMIR SLAMECKA, Director
School of Information Science
Georgia Institute of Technology

Atlanta, Georgia
December 31, 1969

C O N T E N T S

PREFACE	i
CONTENTS	
LIST OF FIGURES	iii
LIST OF TABLES	iii
I. SUMMARY OF RECOMMENDATIONS	1
II. THE STATE OF THE ART	3
III. SYSTEMS ANALYSIS	7
1. Educational Services for the Blind	
2. Vocabulary-Based Translation Into Braille	
3. The Algorithmic Approach to Braille Translation	
IV. SYSTEMS DESIGN	20
1. Compatibility With Environmental Considerations	
2. Project Objectives and Assumptions	
3. Quality Controls	
4. Recommended System Design	
a. System Functional Flow Diagrams	
b. Input	
c. Processing	
5. Alternate System Design	
6. Additional Options	
V. PROJECT MANAGEMENT	38
1. Organization	
2. Tasks and Structure	
3. Schedule	
4. Cost Estimate	
SELECTED REFERENCES	47

LIST OF FIGURES

Fig. 1.	Concept of An Educational Facility for the Blind	22
Fig. 2.	Book Translation Into Braille Using the APSS Educational Facility	23
Fig. 3.	Project Quality Control Flow Chart	25
Fig. 4.	Recommended System Design: Functional Flow Chart	27
Fig. 5.	Alternate System Design: Functional Flow Chart	32
Fig. 6.	Project Task Chart	44

LIST OF TABLES

Table 1.	Braille Production System Project Tasks	42
Table 2.	Project Task Schedule	45
Table 3.	Direct Costs Estimate for the period January 1 -- August 31, 1970	46

I. SUMMARY OF RECOMMENDATIONS

The School of Information and Computer Science, Georgia Institute of Technology, is pleased to offer the following conclusions resulting from the present study:

1. The objective of the Atlanta Public Schools to produce book materials in Grade 2 braille for educational purposes of the visual impaired, and the intended technical approach employing information processing technology are both sound and consistent with the state of the art.
2. The production of books in braille by computer-aided transcription from inkprint materials should be viewed as an integral element of a multi-media, broad-purpose educational facility for the visually impaired, and it should be implemented in a manner which is technologically compatible with the capabilities of such a facility.
3. In view of its conformity with the above recommendations and of its operational readiness, the DOTSYS inkprint-to-braille translator and embosser are the first choice as the computer-aided mechanism for the production of books in braille.
4. A system design is recommended which employs an alternate method of inkprint-to-braille translation (in the event the DOTSYS system cannot be made available), essentially with impunity for the objectives and other functional modules of the project.
5. A management approach is recommended which, while assuming

a full responsibility and commitment on behalf of the Atlanta Public Schools, allows for flexible implementation of project activities either in-house or under a subcontract.

II. THE STATE OF THE ART

The primary objective of this state of the art survey was to review existing or realistically potential developments in several areas which might have an impact on the technical approach to the planned translation of inkprint books into Grade 2 braille.

Specifically, this survey was intended to answer the question of whether there are other methods and/or devices which might make obsolete either the overall purpose and objectives of the planned project, or the intended technical approach of using computer-aided translation. The study has also sought to ascertain the characteristics of various devices whose application in the planned project appears potentially attractive.

With these goals in mind, the project team has given attention to the following areas:

- a. The information-communication patterns of the blind;
- b. The state of development, acceptance, and future of braille;
- c. The "linguistics" of Grade 2 braille, especially from the viewpoint of the possible degree of algorithmization of its rules;
- d. The technology of braille print production;
- e. Related research in bioengineering, the computer and information sciences, and in electronics.

The purpose of this section of the report is not to review in writing the state-of-the-art in these areas. As a result of the study, however, the project team is prepared to affirm the objectives and general technical approach adopted by the Atlanta Public Schools with

respect to the planned translation of inkprint materials into Grade 2 braille.

With respect to the above areas of the survey, the following brief comments highlight the findings.

Information and the Blind User. There is little doubt that the braille-print medium shall for some time continue to be essential to the visually impaired. This conclusion, affirmed despite an increased recent emphasis on other media (primarily audio), is supported by

- a. a growing emphasis on interfacing the blind, via braille, with the sighted;
- b. a growing emphasis on more advanced education of the visually impaired; and
- c. an increasing attention to better job opportunities for the blind, at levels requiring higher qualifications.

The present-day emphasis on the development of other communication media and techniques does not negate this conclusion.

The State of Braille. The process through which Grade 2 braille has developed is not unlike the development of natural languages: it is the result of evolution rather than planning or design. Grade 2 braille therefore exhibits characteristics which are less than optimal from a systematic viewpoint -- a feature which is recognized with varying degrees of concern by both the blind and the sighted workers in this field.

The introduction of mechanization into braille processing is beginning to exercise pressures for greater algorithmization, or at least consistency, of its rules, as witnessed by efforts in Germany, for example. It is to be expected that when placed in general use in the

blind community, mechanized systems such as DOTSYS cannot but have a standardizing effect on the learners of braille.

While the consistency of braille rules is of obvious interest to those concerned with its automation, it is our belief that the principal concern for revising the logic of braille must be from the viewpoints of its users and its learners.

Automatic Braille Processing. The position stated in 1963 by Yngve remains essentially correct. To the extent that translation from inkprint into braille is primarily a transcription (rather than a translation) process, the algorithmic approach may be expected to have a greater degree of success than it has enjoyed in the mechanical translation of natural (foreign) languages.

Because Grade 2 braille does not allow a one-to-one transcription of symbols, neither the "algorithmic" nor the "dictionary" approach to its translation are based purely on rules and vocabularies: each employs both rules and a dictionary. The difference lies in the respective uses of rules and vocabularies to handle exceptions. In the algorithmic approach, words which cannot be handled by the rules are transcribed by accessing a dictionary of exceptions; in the vocabulary-based approach it is the rules which attempt to decide the possibly ambiguous transcription of words.

The Technology of Braille Production. The systems in operation at the M.I.T. Sensory Aids Evaluation and Development Center, at the American Printing House for the Blind, and the MEDCOMP Research Corporation's BRAILLETRAN represent the state of the art. Both the IBM high-speed printer and the M.I.T. embosser offer good printing quality; the latter has the advantage of greater versatility as a system component.

Other output devices such as the electrical version of the Perkins Brailier (manufactured and currently field-tested by Howe Press) are showing promise as system components.

Related Developments. Considerable volume of research is underway in several sectors of technology; this research is likely to have a strong impact on the communication facilities of the visually impaired. None of the known efforts appear to be close to producing systems or devices for the mass market; therefore, it is premature to gauge the effect of these developments on the relative position of braille as a means of communication for the blind. Among examples of the more significant work currently underway are devices which scan information and transform it into tactile (or other sensory) signs; character and pattern recognition techniques; complex language analysis; and on-line transmission of handwritten and graphic information.

Notwithstanding the laboratory nature of much of this research, it is the concensus of the School of Information and Computer Science that a judicious design of advanced-capability, multi-media communication systems for the blind is feasible at the present time. Consequently, it is desirable and propitious that the inkprint-to-braille translation project of the Atlanta Public Schools be viewed from within such an environment.

In summary, the survey of the state of art has confirmed the soundness of the general technical approach proposed by the Atlanta Public Schools. It is therefore appropriate to devote the next section of this report to the results of a systems analysis of the alternate technical choices available within the general approach.

111. SYSTEMS ANALYSIS

This section of the report discusses the results of a careful analysis of the problem of computer-aided translation of inkprint textual materials into Grade 2 braille. The analysis has concentrated on two important parameters of the problem: the integration of the braille-translation project with related activities of the Atlanta Public Schools System; and the choice of the computerized translation mechanism for producing braille books from inkprint materials.

1. Educational Services for the Blind

The Atlanta Public Schools System currently serves as a clearing-house for several types of projects for the visually impaired students. The following major projects aimed at the blind have been identified:

- a. On a contractual basis, the Atlanta Public Schools provides itinerant teachers to its own schools and to those of neighboring counties. These special teachers (approximately 25) provide personal counselling, remedial tutoring, and special instruction in braille and typing, and they distribute special materials.
- b. Through a special State grant, a study is in progress to determine the possibility of using the IBM 360/50 computer system to generate Grade 2 braille from inkprint book materials. This study is a cooperative effort of three offices of the Atlanta Public Schools: Research and Development; Instruction; and Information Processing.

- c. A Braille Library, as a repository of materials in braille and large-print texts, is maintained by the Atlanta Public Schools.
- d. A telephone reader service is under development by an inter-county agency, devoted to promoting special education in the metropolitan area. Although the funds of this project are nominally administered by the Atlanta Public Schools, little actual control is maintained.

In general, these projects and services are funded by different organizations within and outside the Atlanta Public Schools.

Analysis of the objectives of projects for the visually impaired students, underway at the Atlanta Public Schools, indicates that they are operationally closely related, and that any one project is but an element in the attempt to provide adequate education for the blind. The relationships (and the difficulties arising from viewing the projects as unrelated) are evident from the following discussion based on interviews and an analysis of programs.

The 25 itinerant teachers serve approximately 250 visually impaired students, of whom 120 are legally blind. About 30 of these students can read braille; these students, who attend regular classes with sighted children, are scattered in grades 1 through 12 over seven different school systems. Teachers in the field report that activities associated with braille-reading children absorb a disproportionate amount of time during the school day, primarily due to the time involved in preparing Grade 11 braille materials. While the Braille Library attempts to provide the necessary braille textbooks by buying, borrowing, duplicating,

or producing them internally (through volunteer workers), such daily materials as tests, assignment sheets, and other handouts must be prepared by the itinerant teacher, often at the expense of other normally scheduled activities (such as counseling or tutoring). These teachers correctly feel that if they were relieved of the task of braille production, they could significantly improve the rate of instruction. (An important point to note is that the field teachers' requirements for instructional materials in braille are not synonymous with those of the Braille Library: the latter is tolerant of response times.)

Similar conflicts of operational conditions arise when other media are brought to bear on the instructional system for the blind, such as voice recordings, or media for reproduction of graphics (currently under development).

Clearly, a need is indicated for the coordination of the various projects managed by the Atlanta Public Schools for the education of the blind. At present, these projects appear to proceed with only minimal informal exchange of information, and under no coordinating authority. As a result, relatively autonomous activities proceed under ill-defined or overlapping authority, resulting in situations in which internal questions concerning the nature and scope of services and projects may not be easily obtainable. Under these conditions any single project -- including the inkprint-to-braille translations project -- becomes a complex management problem, and one which tends to be more expensive than necessary.

Based on our analysis of educational needs of the visually-impaired students, of the current related activities of the Atlanta Public Schools,

and of the characteristics of the inkprint-to-braille translation project, we wish to recommend as follows:

- a. That the Atlanta Public Schools establish a program for the coordinated development of an educational facility for the blind, with a consistent set of goals, methods and standards, and compatible to the greatest possible degree with educational facilities for normal children;
- b. That the inkprint-to-braille translation project be viewed as an integral capability of such a facility.

The general character of the suggested educational facility for the blind is indicated in the systems design section of this report.

2. Vocabulary-Based Translation Into Braille

The survey of the state of the art of braille production has failed to identify a clear-cut alternative to the two basic methods of computer-aided translation from inkprint materials to Grade 2 braille. As a result, we have attempted to perform an intensive analysis of the two existing approaches -- the algorithmic (grammar-based) and the vocabulary-based, with the objective of assessing their suitability for the Braille Library project of the Atlanta Public Schools. Two principal examples of these approaches (the DOTSYS project of the Massachusetts Institute of Technology, and the Braille Translation Program of the IBM Corporation, respectively) were studied in detail as to their present level of development and utility. The results of this analysis are discussed in this and the following section.

The vocabulary-based approach to translation into braille, as exemplified by the IBM program, is based on the premise that a vocabulary of natural language words can be maintained in correspondence with a collection of Grade 2 braille symbols, such that each string of natural-language symbols w_i (e.g., a word) has a corresponding string of braille symbols b_i . The process of translation then proceeds by finding and substituting, for each w_i , its equivalent b_i .

Theoretically, this approach provides a diverse basis for adapting the dictionary to the environment of use. Inter-word connections may be marked for special handling, as can hyphenation, by providing additional coding within the string of braille symbols. It is the "special handling", so frequent in Grade 2 braille due to the contractions and other context-dependent idiosyncracies of this writing system, which mars the theoretical simplicity of the vocabulary-based approach. A word marked for special handling should, optimally, be handled by the computer program which selects the correct Grade 2 braille equivalent on the basis of "rules" built into it. In the absence of any such rules (which, in effect, are rules of Grade 2 braille grammar), the vocabulary-based approach would reject each word marked for special handling, and await human intervention.

An effective vocabulary-based approach to computer translation into braille must therefore rely on the existence of at least some rules in the computer system. In addition, the suitability of this approach is a function of the complexity of the text to be translated: probably it serves adequately simple rhetoric texts having moderate vocabulary, grammar and braille-code demands. For linguistically complex texts,

comprehensive vocabularies and relatively powerful context-interpreting rules for braille transcription are desirable.

The specific analysis of the IBM braille translation program has identified deficiencies in the following areas: input preparation; vocabulary maintenance; and program efficiency. Each of these constraints are discussed briefly below.

Input Preparation. The theoretically flexible facilities of the IBM vocabulary-based translation program are rendered essentially ineffective by a very awkward user interface -- the formidable keypunch rules. The punchcard input necessitates the development of a highly skilled team of keypunchers committed to the project. The keypunching rules applied to the natural language text are so complex as to cause high error rates and, consequently, a slow keypunching output.

An alternative to keypunching is proposed and discussed in the system design section.

Vocabulary Maintenance. Entering new words into the IBM system dictionary requires them to be first interpreted into Grade 2 braille; then a cumbersome coding system is used to digitally encode the embossed position of each Braille cell. For instance, the following six cells

⠠ ⠠⠠ ⠠⠠⠠ ⠠⠠⠠⠠ ⠠⠠⠠⠠⠠ ⠠⠠⠠⠠⠠⠠

would be digitally encoded as 1, 12, 123, 1234, 12345, 123456 (the commas separating the code for each cell). Each position of a cell has thus been assigned a digit: beginning in the upper left-hand position of a six-position braille cell, the positions are numbered downward (1,2,3), the upper right-hand position is number 4, and the remaining

positions are numbered downward 5 and 6. This process is prone to transcription error in going from braille to digits.

In addition to the braille symbols, the dictionary also contains information on hyphenation and exception handling. For instance, in the previous braille cells, hyphenation between the fourth and fifth cells is indicated by a hyphen, as follows:

01, 12, 123, -1234, 12345, 123456

Zero in the first cell signals the need for exception handling.

In view of the cumbersomeness of this representation, it is desirable to consider other methods of building the dictionary. Several approaches are indicated and discussed in the system design section.

Computer Programs. The IBM program employs two dictionaries: an 800-word dictionary maintained in high-speed memory, and a large dictionary maintained in low-speed memory. Clearly, the organization of the dictionary determines the relative efficiency of the program. The program intends to exploit the frequency of occurrence of words in English text, so as to reduce the total search time; placing them in high-speed memory, inquiries to low-speed memory occur less frequently. While this concept is sound, in actual implementation several factors come to bear which offset the theoretical advantages. Among these are:

- a. The amount of high-speed memory committed to fast look-up is entirely too small to cover a spectrum of texts. While 800 frequently used words might be significant for a single text, it is doubtful that any set of 800 words will cover a collection of texts across several diverse topical areas;

- b. While the programming techniques applied to searching high-speed memory are efficient, the techniques used to search the low-speed memory are poor and unnecessarily time-consuming;
- c. Data dependencies which are built into the text scanner portion of the program are closely tied to the quality of input. Thus a violation of a keypunch rule in formatting data will cause a maximum search time to be expended in looking for a word which cannot exist in the dictionary.

These factors combine to provide a very slow running computer program which produces very little acceptable braille per unit of operation.

Several recommendations are made in the system design section of this report relative to the improvement of these programs.

The preceding discussion has enumerated several serious problem areas attendant to the existing IBM program for vocabulary-based translation to braille. These problem areas impair the utility of this program at the present time. The concomitant intent of this discussion has been to imply, however, that the vocabulary approach is not a priori inferior to the algorithmic method, and that it is a feasible one. Its apparent lack of elegance need not be a decisive drawback: there are other potential uses of a braille dictionary which make up for it under some circumstances.

The advantages of a vocabulary-based approach might be summarized as follows:

- a. It is relatively independent of the programming of Grade 2 braille rules;
- b. It is suited to handle the hyphenation problem;

- c. It can suitably provide for the transcription of foreign words (which is especially important in scientific fields) and names (personal, geographical);
- d. It assists in the proofreading stages of the translation process, in that any word correctly located in the dictionary is automatically proofread;
- e. Once a vocabulary exists, it can find additional important uses, e.g. in automatic indexing of inkprint or braille materials.

3. The Algorithmic Approach to Braille Translation

Under the "algorithmic approach" we understand a method of translation into braille which relies heavily on an application of Grade 2 braille rules, programmed into a computer. Only an ancillary use is made of a relatively small vocabulary, comprised of words which are not easily handled by the computerized rules.

This approach is said to be consistently used by the DOTSYS system. The following description of DOTSYS is excerpted from Goldish (1967):

The DOTSYS program ... is capable of taking source material in any form-compositor's media punched tape or cards, or direct input-and converting it into Grade 2 braille output - either on tapes, cards, or direct. This program is based on the Schack translation program presently in use at the American Printing House for the blind.

Program Format

The Program is actually a series of subprograms, each performing part of the production sequence.

1. Interpreter Programs: The interpreter programs - TELCON, MONOCON, and the like - provide the link between the individual compositor's media codes and the rest of the DOTSYS program. Thus, as new input source forms become available, only an additional interpreter sub-

program need be written; no changes in the rest of the program are necessary. The interpreter programs interpret the source codes and convert them into an expanded code called UNIVERSAL.

2. UNIVERSAL: - This subprogram interprets the many symbols used as inkprint format indicators which are normally recognized from context by sighted readers. The program distinguishes the many different ways that paragraphing may be presented in inkprint and converts each of these to a special Start Paragraph code. It determines whether the symbol (') represents a single right quote, part of a double right quote, or an apostrophe. It performs a scan to determine whether a (.) is a period or a decimal point. It provides information about centered headings, inserts the special codes for reference notes, indicates the beginning and end of italicized passages, and generally supplies much of the formatting information formerly inserted by a keypunch operator.

3. UNICON: The UNIVERSAL codes are converted by another program called UNICON, and the result is a stream of computer codes in the same form as a key-punch operator would have produced. This program determines the length of an italicized passage and supplies the appropriate codes whether the italicized portion be a part of a word, a single word, a stream of more than three words, or a series of paragraphs. It performs a similar scan to supply the proper capital signs, eliminates unnecessary spaces, and translates the UNIVERSAL format codes into the special codes required by the braille translation program.

4. BRAILL: The output from UNICON serves as input to the BRAILL program which does the actual translation into Grade 2 braille. This program is, essentially, the dictionary-lookup and rules-listing portion of the Printing House program described earlier.

5. FORMAT: The stream of braille codes from the BRAILL program is modified by a FORMAT program which inserts end-of-line and end-of-page codes, supplies spaces necessary to center material, determines whether a centered heading may appear near the end of the page, or whether a new page must be started.

6. Output Compatibility Programs: The formatted braille then becomes input to another program which produces appropriate codes to operate the embossing equipment. Two such programs have been written - one to produce a tape for the high-speed, tape-controlled embosser, and another to produce cards for the card-controlled stereograph equipment. Here, too, when other output is desired, only an additional compatibility sub-program need be written.

Equipment and Speed

The full DOTSYS program has been run primarily on the IBM 7094 with 8K (8,000 words) of storage. For direct reading of teletypesetter tape into the computer, the CX tape reader mentioned earlier is connected to the computer.

The 7094 processes the material through the complete program at the rate of 200 braille pages in 4 1/2 minutes. The primary reason for having used the 7094 computer has been its availability. It is likely, however, that the program could be run on a smaller machine; no studies of this have been made as yet. A smaller machine would require more processing time but at a lower cost, so that the economic aspect would be a major criterion. Also, because of the speed and characteristics of the program, it is best to run the program as a unit on one machine.

Implementation

The translation and output subprograms of the DOTSYS system have been in operation under production conditions at the Printing House with punched-card input and have proven themselves admirably. The compositor's media input program - a more recent development and not as complex as the translation program - has been successfully run under simulated production conditions and is deemed ready for production use.

At present, teletypesetter tape is the only compositors' medium being read, and further interpreter subprograms will be written as needed. Also, the DOTSYS system presently handles only literary braille, but a Nemeth-Code mathematics translation routine is under development.

It is our opinion that the DOTSYS system represents the state of the art of computer translation to Grade 2 braille. The translation program appears to respect the braille rules to a high degree, and the output is generally acceptable to the braille experts associated with the Massachusetts Institute of Technology. The DOTSYS system acknowledges certain compromises, especially where the current state of the art does not permit the programming of braille-transcription rules into the computer; the M.I.T. consensus is that the effect of these compromises on the quality of output is negligible, however.

An example of the latter attitude is the problem of hyphenation of braille. It is not easily feasible at the present time to develop a complete set of braille hyphenation rules programmed into a computer, although a satisfactory performance can be obtained by using rules based on word prefixes and suffixes. The latter rules are scheduled to be built into the DOTSYS system in early 1970; at the same time, the general attitude is that the problem of perfect hyphenation is not a critical area of braille translation (because of the natural division of text by paragraphs).

A caveat should be inserted at this time relative to the degree to which the DOTSYS translator is an "algorithmic" one. In the above quote, Goldish refers to the BRAILL translation program as being "essentially the dictionary look-up and rules-listing portion of the Pringing House program". DOTSYS clearly makes use of a dictionary which, to our knowledge, contains about 1200 words. Although no confirmation could be obtained, it would seem that the program of rules is essentially complete (to the degree that these rules lend themselves toward algorithmization), and that exceptions are handled via the dictionary.

Our evaluation, in a practical demonstration, of the M.I.T. embosser (a component device in the DOTSYS research program) has shown it to have been successfully designed to a set of specifications which observe the quality of braille output, telephone transmission rates, psychological considerations in its use by the blind, and other consideration.

Principally because of its advanced operational status, the DOTSYS system has a decided advantage as the potential translator of inkprint

books to braille. The availability and use of DOTSYS in the Atlanta Public Schools would, furthermore, have the effect of simplifying the procedures and reducing the cost of braille production by computer. Without regard to the detailed and unknown idiosyncracies of DOTSYS, it is clear that the input text to it is subject to nominal formatting constraints; this input text (which may be referred to as "braille input format files") can be created from other files, e.g. those generated via the IBM Administrative Terminal System (ATS/360).

From the information available to us, there appears to be no penalty for using DOTSYS (other than the necessity of reprogramming it for the IBM/360 computer system); it is fast, efficient, and flexible relative to input and output media. The feasibility of using DOTSYS hinges, rather, on administrative considerations; these were spelled out in a previous correspondence between the Georgia Institute of Technology and the Atlanta Public Schools.

IV. SYSTEM DESIGN

This section of the report presents a technical design of the system for the translation of inkprint text to braille.

The technical design has two important characteristics. First, it views the translation project as an activity related to and compatible with other projects, current and perceived, for the instruction of the visually impaired. Second, it allows for an alternate translation mechanism to be employed if the desired administrative arrangements with the Massachusetts Institute of Technology (under negotiation at the time of writing this report) do not materialize.

1. Compatibility With Educational Services for the Blind

The broad, encompassing commitment of the Atlanta Public Schools to serve the blind students of this region has come through very clearly at every stage of the system analysis. An equally persistent thought arising from this commitment has been the necessity to consider individual activities as being directed toward a major goal, and to view them as elements of a broad program.

In the absence of a statement describing such a broad program, the study team of the School of Information Science has attempted to formulate a concept of a comprehensive educational facility for the blind, as recommended to the Atlanta Public Schools in this report. The overall objective of this facility is to make available -- on-line in the classroom or home, and under the control of the teacher or student -- materials in spoken, braille, hand-written, and graphic form.

The design on this facility, which should and can be extended to

serve simultaneously the sighted student (Figure 1), is not the objective of this report. Its realization is sufficiently feasible, however, that a requirement was imposed on the design of the inkprint-to-braille translation project to make it compatible with such a facility. Figure 2 illustrates the compatibility of the braille translation system design with the concept of the educational facility for the blind.

By imposing this compatibility requirement, the system design of the braille translation project also serves as a guide to the design and technical approaches of the educational facility for the blind.

2. Project Objectives

The purpose of the project is to provide a system for effective translation of inkprint texts into Grade 2 braille. Toward this purpose, the following specific objectives are identified:

- a. To determine and maintain the requirements of the Atlanta Public Schools and other relevant users for braille book production;
- b. To establish and maintain the systems and procedures for selecting, scheduling, converting and certifying textual input materials to a braille-producing computer program;
- c. To produce, via cost/effective applications of information processing technology, textual materials in braille, of a quality consistent with the defined requirements of the Atlanta Public Schools;
- d. To establish the procedures for certifying, storing, cataloging and disseminating braille texts for educational uses by the Atlanta Public Schools and other authorized users.

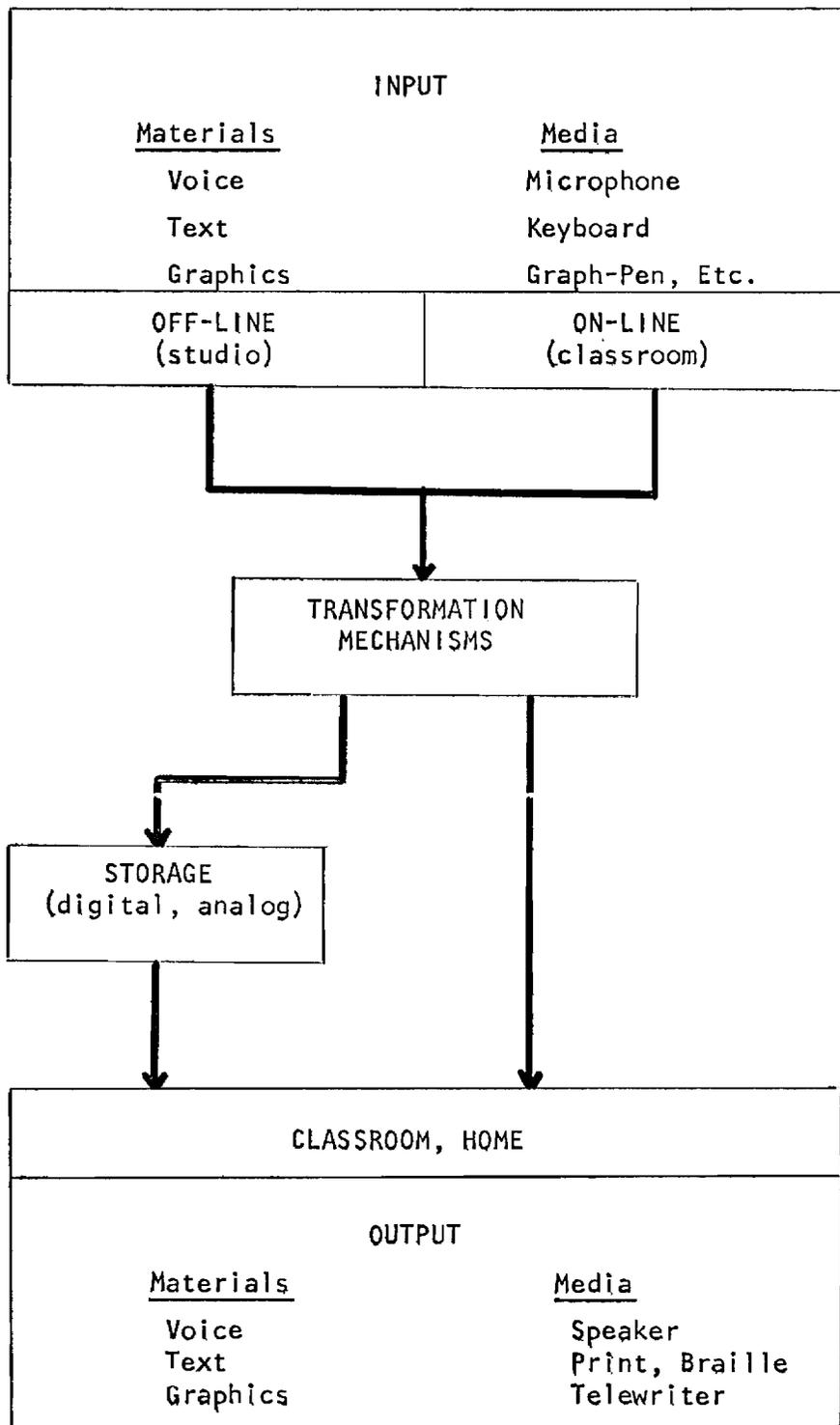


Fig. 1. Concept of An Educational Facility for the Blind

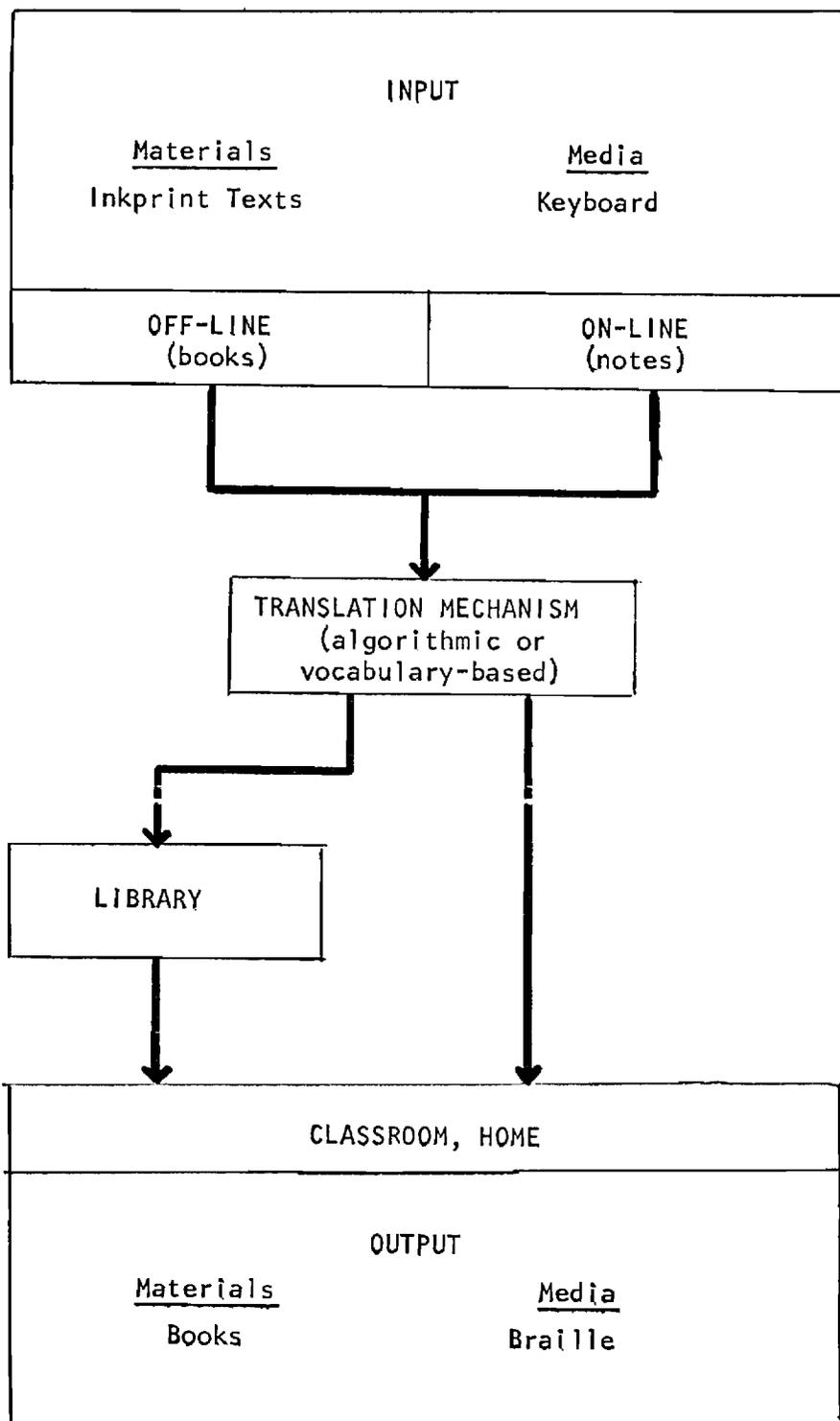


Fig. 2. Book Translation Into Braille Using the Proposed APSS Educational Facility for the Blind

These objectives are based on the premise that the Atlanta Public Schools System is committed to promote improvement of education of the visually impaired. With respect to book production in braille, this commitment may be extended to serve the needs of the entire State.

On the other hand, the stated objectives do not specify at this time that the project be financially self-supporting through the sale of braille materials. In our opinion, such a design requirement would impose an unrealistic initial constraint on the quality of the product.

3. Quality Control

Regardless of the method used for the translation of inkprint text to braille, the system adopted must attain a prescribed level of quality of product. This report will not attempt to establish such a level, but will identify the instances and outline possible procedures for testing and insuring a given level of quality. It is axiomatic that the level of quality of braille, like that of any other product, is a function of the resources expended toward its control.

In the proposed system design, quality controls should be applied at both the input and the output stages. No matter how perfect the translation process, the quality of the output shall be a function of the quality of input; hence input quality control is mandatory. To the extent that neither the algorithmic nor the vocabulary-based translation programs are error-proof or completely automatic, output quality control of braille text is equally mandatory.

Figure 3 is a diagram showing the sequence of quality control stages. Since both the algorithmic and the vocabulary-based translation approaches employ dictionaries (to a different extent and to different ends),

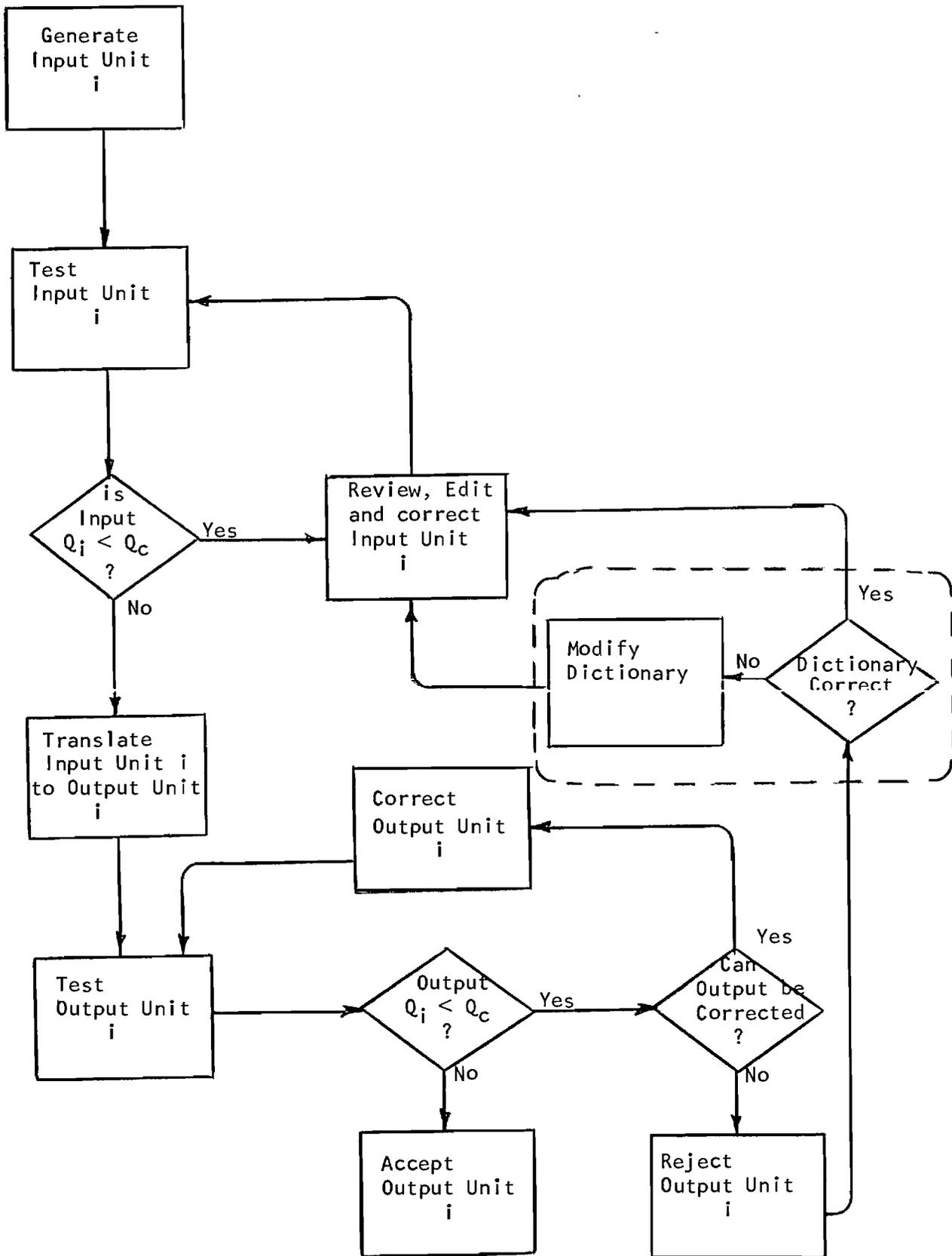


Fig. 3. Project Quality Control Flow Chart

vocabulary modification and maintenance is an additional dimension in the quality control procedure.

The quality control procedure basically employs a control measure Q_c , to determine the quality index Q_1 of an input or output unit. The definition of the acceptable confidence level Q_c for the input and output stages of the braille translation program is an important early task of the project.

As regards the quality testing procedure, it is probable that simple proof-reading is inadequate, and that the employment of a sampling technique is indicated. One approach might be to use a pseudo-random number generator to produce a sequence of pseudo-random line numbers. For a text containing N lines, there is some number M of random lines such that the probability of error determined by careful character-by-character examination of the set of M lines is the probability of error in the full set of N lines (at some level of confidence, e.g., 95%). This probability of error in the sample M can be construed to be an index of quality. If $Q_c = 10\%$ and the confidence level is 95%, M must be of a sufficient size relative to N . In other words, we insist on being 95% certain that the error of a given unit i is less than 10%.

4. Recommended System Design

Figure 4 shows the functional flow diagram of the recommended system for braille production from inkprint texts, employing the DOTSYS translator of the Massachusetts Institute of Technology.

The selection of the DOTSYS translator is given by its being representative of the state of the art. The use of DOTSYS is contingent on its availability under terms and conditions complementary to the objectives

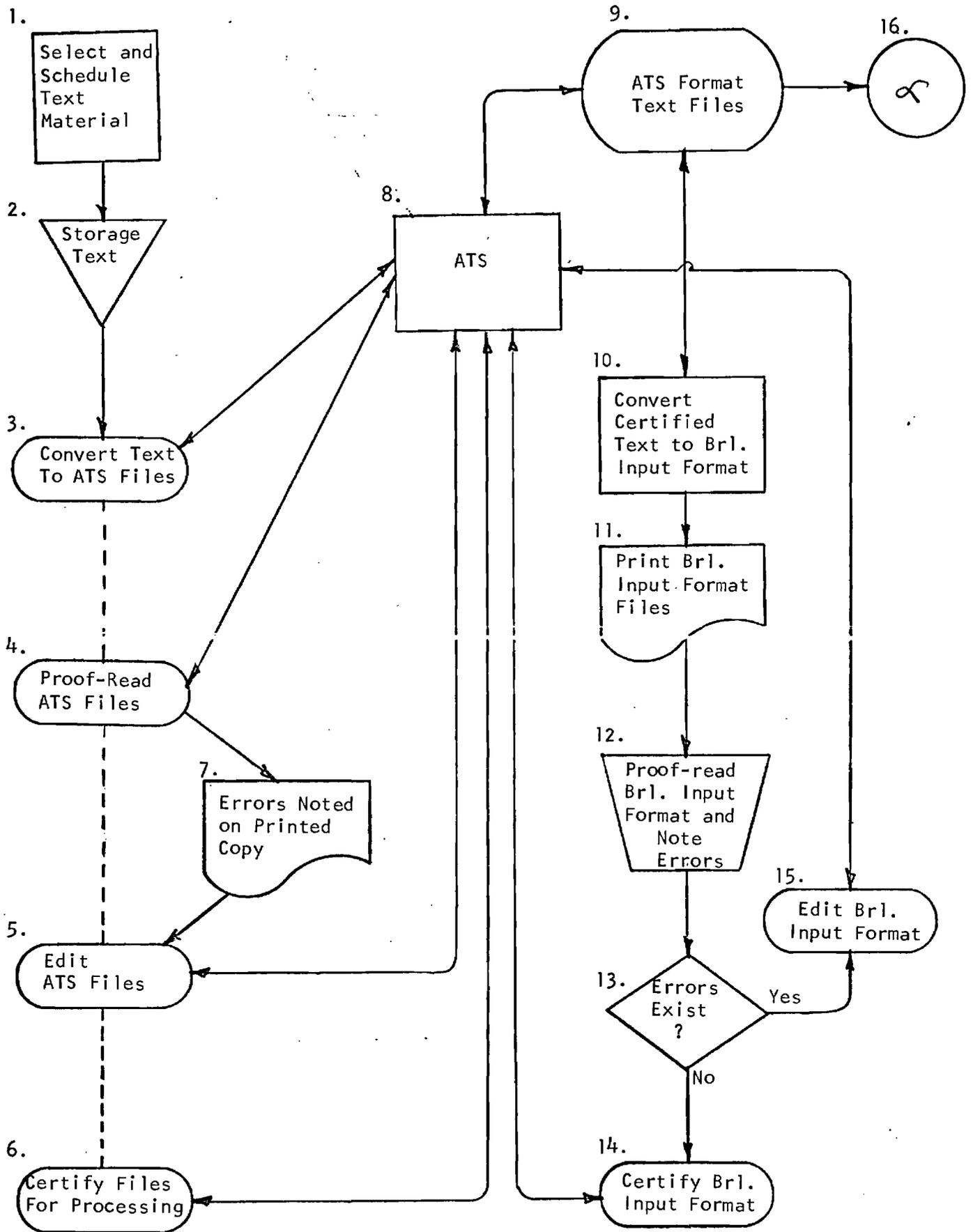


Fig. 4. Recommended System Design: Functional Flow Chart

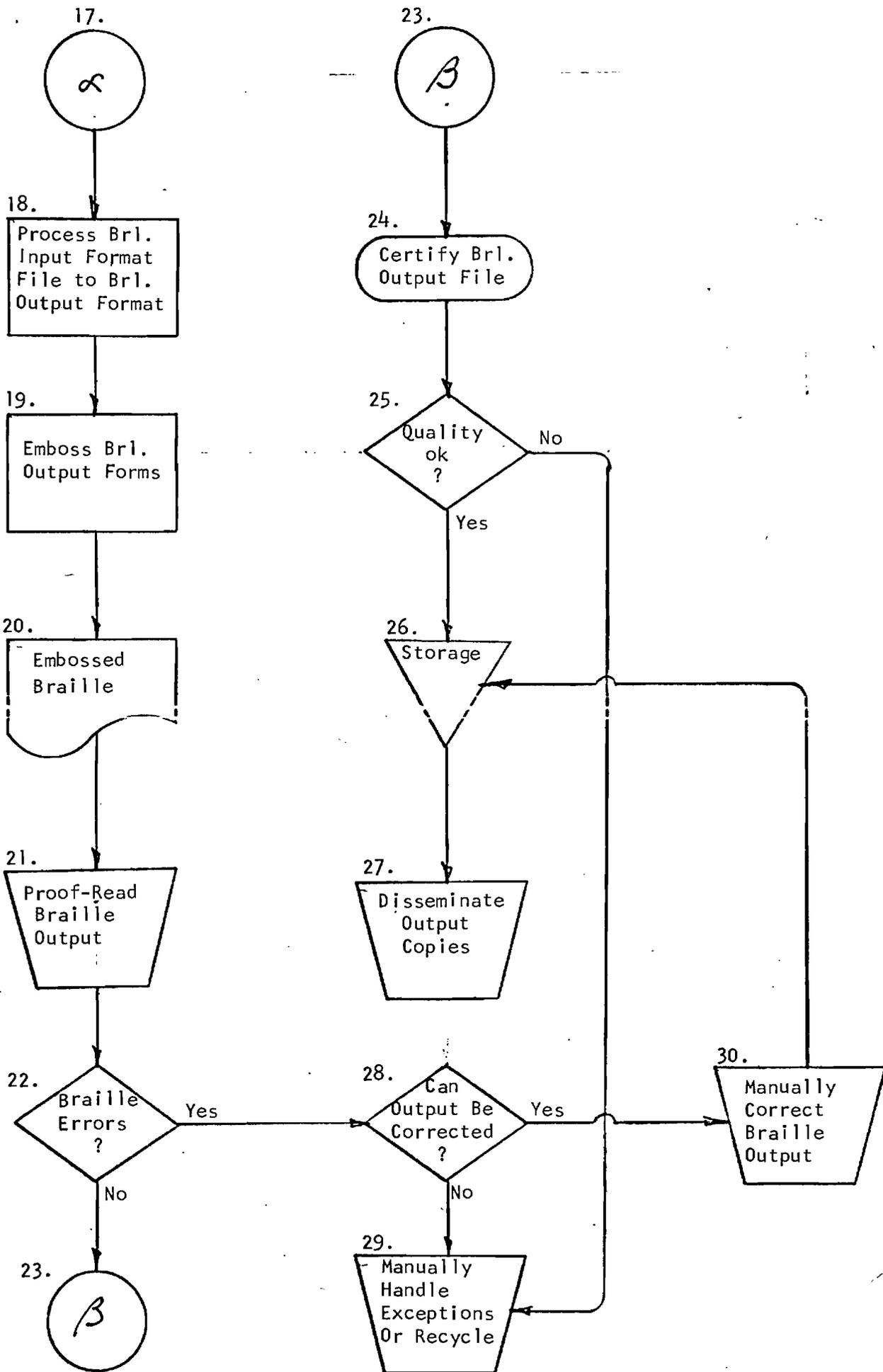


Fig. 4. (Continued)

of the Atlanta Public Schools.* In the event that DOTSYS may not be available, an alternate design employing the modified IBM Braille Translator is presented in the subsequent section of this report. There are certain by-products of the vocabulary-based approach which may be of interest to the Atlanta Public Schools; these are discussed in the last section of this chapter.

The use of either of these translator programs guarantees that a significant number of full-text files will come under machine management.

The remainder of this section on the recommended system design discusses possible or preferable approaches and techniques pertinent to the various process steps.

Input Text Conversion. As an alternative to keypunching, the facilities of the computer program ATS/360, available on the Computer system of the Atlanta Public Schools, are preferable. The availability of this program, designed for the purpose of entering, developing, revising and outputting text, provides an opportunity to capture natural-language text, by typists copying inkprint materials in essentially textbook format, onto machine processable files. The same program facilitates editing.

The ATS/360 character set encompasses the entire 256 characters contained in the EBCDIC set. Text may be entered from a remote terminal (a modified IBM SELECTRIC typewriter), a local card reader, or from magnetic tape. It may be developed and revised from a remote terminal. It may be formatted and outputted to a remote terminal, a local high-

*See letter of Dr. Jarvis Barnes to Dr. T. Mann (December 2, 1969).

speed printer, a local high-speed card punch, or on magnetic tape.

In terms of the requirements of the braille project, ATS allows a ready transcription and editing of source text as well as of dictionary entries. It permits to interface its internally-formatted files with externally-formatted files; thus COBOL programs may be employed to edit and manipulate text after it has been initially entered and corrected.

The conversion of input text (System Block 3) may then be considered a typist-like function of copying inkprint text into machine processable form called ATS file format. (ATS file format refers to natural language text files stored in computer processable media and produced through the Administrative Terminal System. The information in these media is encoded in such a way as to preserve the content of natural language text: upper and lower case, punctuation, paragraphing, chapterization, and pagination.) A minimum number of additional rules is imposed on the typist.

The functions of proof-reading and editing (System Blocks 4, 5) are also carried out through standard facilities of the ATS/360 system.

Conversion to Braille Input Format. This function (System Block 11) presumes the existence of a COBOL program which accepts certified ATS files and converts them to an intermediate input format for the Braille translation program. For the DOTSYS translator, the requirements of this COBOL program are not known. (It is possible that subsequent functions of System Blocks 12, 13, 14, 15, & 16 could be eliminated if the Braille input file to DOTSYS needs no refinement.)

Translation to Braille Output. System Block 18 intends to use the DOTSYS translator.

The remainder of the diagram prescribes a continuing level of quality

control. Assuming that for given textual materials there will be units of text which cannot be machine translated, both manual correction and manual exception handling are anticipated. These latter functions should be minimized through input monitoring.

5. Alternate System Design

In the event that the DOTSYS translator cannot be made available for the braille book translation project of the Atlanta Public Schools, an alternate translator, based on the vocabulary look-up approach, may be employed.

Figure 5 shows the functional block diagram of this alternate design. Compared with the preferred design (Fig. 4), the principal operational difference of the alternate system lies in the extended area of vocabulary management. (System block numbers associated with only the alternate design are marked with an asterisk.) Both system designs conform to the concepts formulated and expressed in Figures 1 and 2.

Specifically, the necessity to employ a large vocabulary in the alternate system calls for imaginative approaches to the dictionary production, and for a modification of the existing IBM braille translator program. Possible approaches to the former, and aspects of the latter are discussed below.

Methods of Dictionary Development and Maintenance . System Blocks *10 through *19 suggest that no text should be computer-processed on a production basis until a proper dictionary exists for that body of text. It is therefore desirable to develop additional software to preprocess each text, so as to produce a word list with frequencies of word occurrence (System Block *10) which can be checked against the existing dictionary.

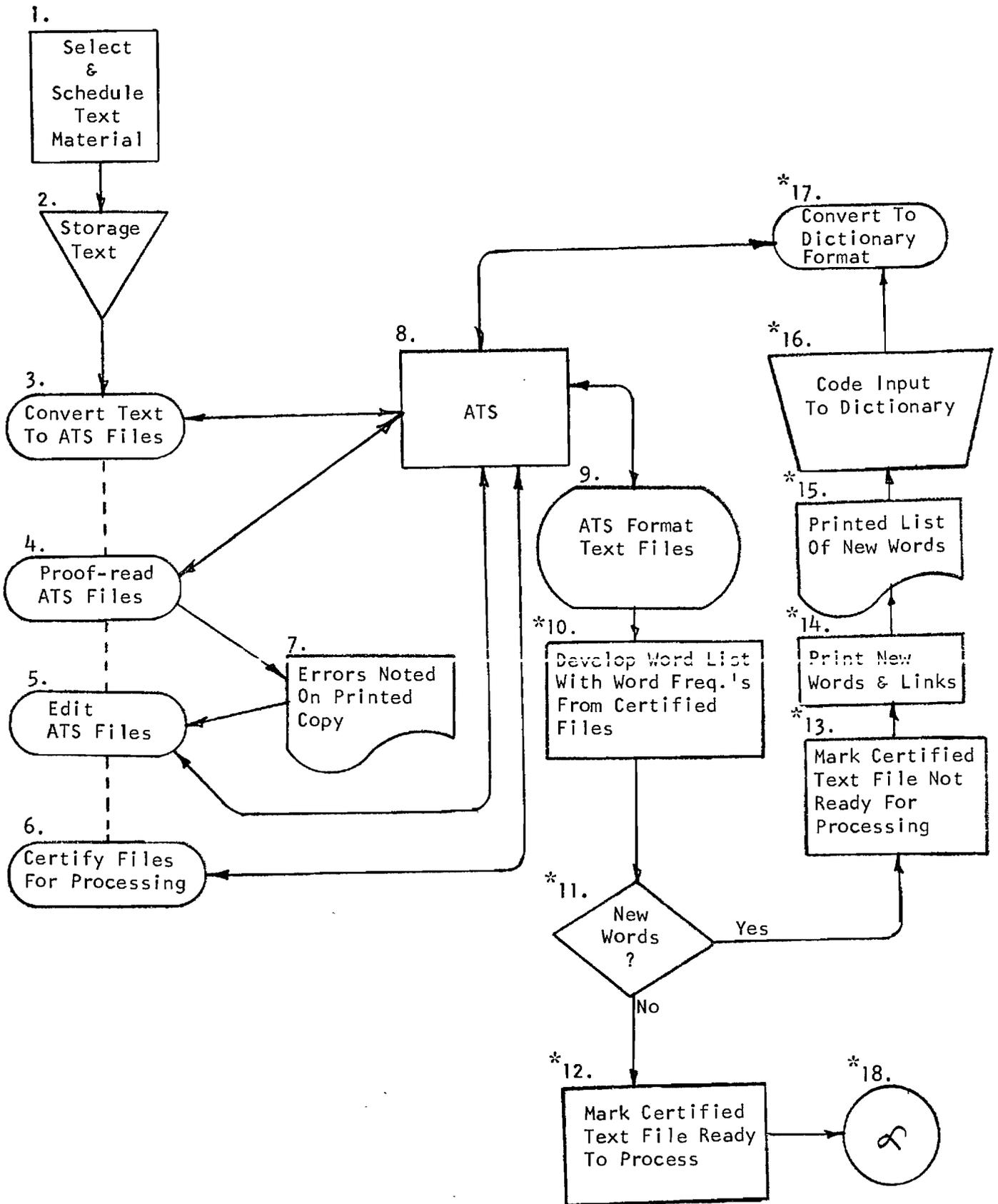
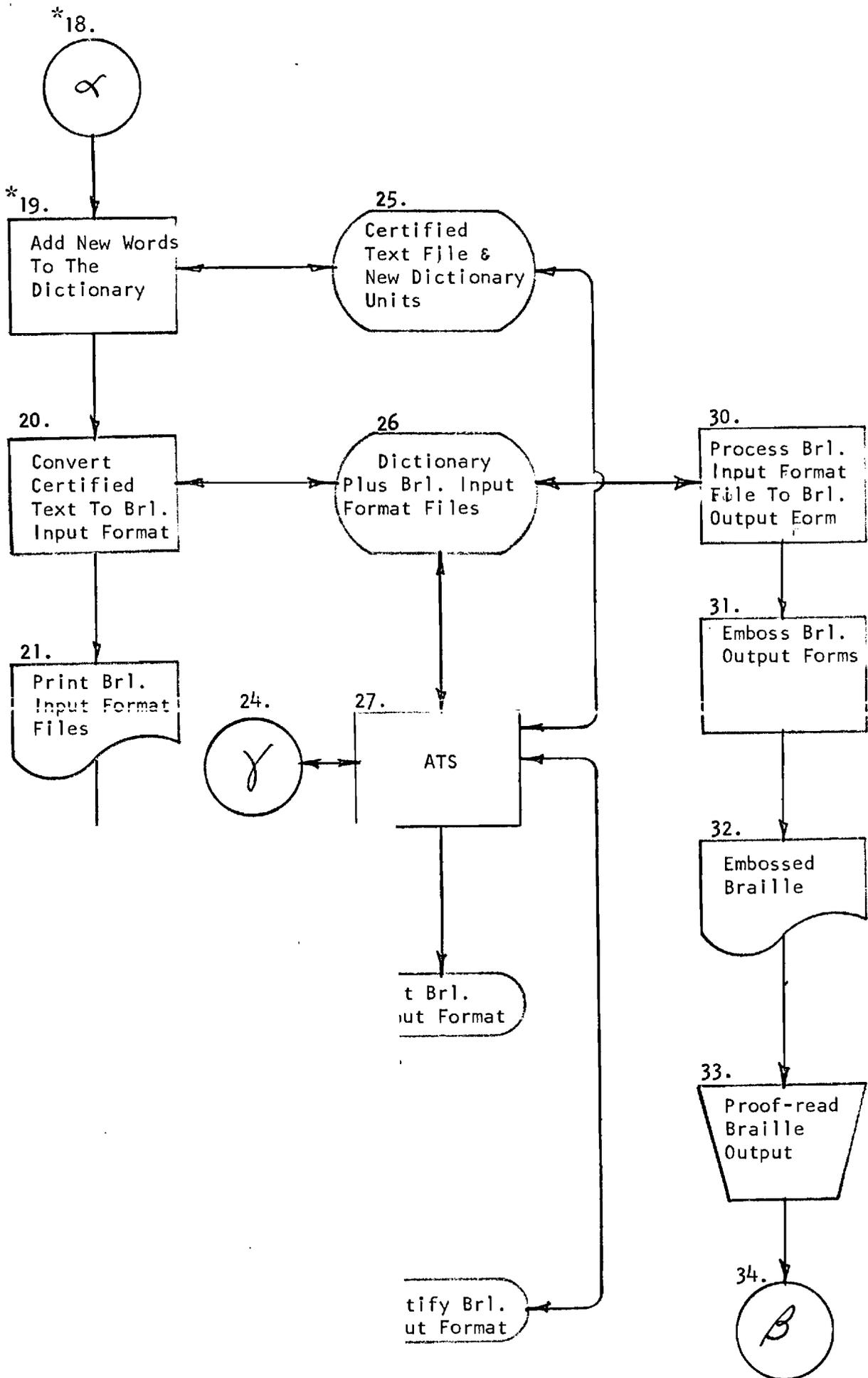


Fig. 5. Alternate System Design: Functional Flow Chart



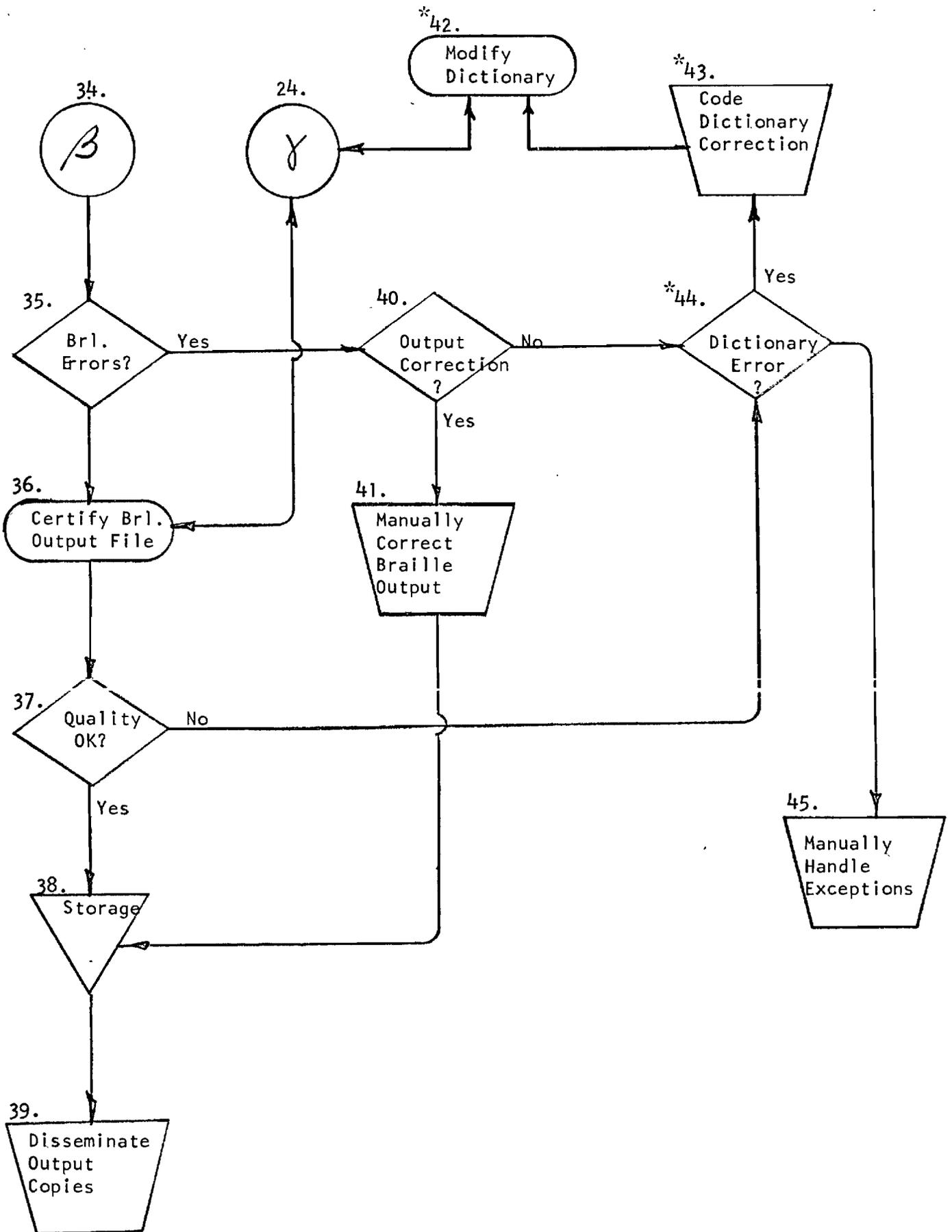


Fig. 5. (Continued)

Words not found in the dictionary must be given their braille equivalents and posted to the dictionary in the fast or the slow memory, depending on their frequency of occurrence.

In view of the cumbersome method employed by the existing IBM program to build the dictionary, serious consideration should be given to one of the three following alternate techniques for vocabulary development:

a. An elegant, highly automated approach to the development of vocabulary is possible by using the DOTSYS translator, under a special arrangement with M.I.T. A computer tape (or a portion of it containing words of low frequency) resulting from the preprocessing (System Block *10) would be processed by the DOTSYS program, yielding braille equivalents.

b. A dictionary-building process employing the page scanning or test scoring equipment of the Atlanta Public Schools may be considered. To be used, this technique requires the development of a form capable of representing braille cells. After marking by a brailist, the form would be processed by the page scanner to establish an entry in the dictionary, using a COBOL program.

In order to use the scanner, it is necessary to link the inkprint representation of a given word in a COBOL file with the braille pattern with which it is to be associated. If this link is some six-digit decimal number, this number and the desired braille symbols can be coded into a scanner record with the following characteristics:

0	1	2	3	4	5	6	7	8	9	a	b	c	d	e	f
								/		/	/	/	/	/	/
							/	/		/	/	/	/	/	/
						/					/	/	/	/	/
					/									/	/
						/				/	/	/	/	/	/
							/			/	/	/	/	/	/

Through the scanner, this record would permit an assignment of the braille information associated with cells a, b, c, d, e and f ... etc., to the word associated with the link numbered 987678. The first marking position (upper-left position) of each cell is available for hyphenation information. The second position of cell a is available to indicate exception handling. The remaining six positions of each cell correspond to the six embossing positions of a standard braille cell. While this form may look awkward at first, the approach avoids multiple transcriptions of data.

c. Another possible approach is to establish the dictionary input having a brailist use the ATS/360 system. He would create, at an ATS terminal, an ATS file consisting of inkprint links followed by the appropriate sequence of digits. As in the previous technique, the files created through the ATS terminal would be processed into dictionary input form by a COBOL program.

The details of the suggested alternative approaches to vocabulary building are outside the scope of this report; they are introduced here to indicate that different techniques are available to the project manager.

Dictionary Look-up and Translation Programs. As already pointed out, the existing IBM translator program is deficient in its handling of

the low-speed memory search. Jointly with the improved quality of input text, the procedures for searching and accessing the disc-resident dictionary must be modified if the program is to be efficient.

6. Additional Options

It would be possible, of course, to undertake the design and development of an entirely new computer program for the translation of inkprint into braille. The magnitude of commitment and resources to such an alternative would significantly exceed those necessary to accomplish either of the two previously described technical approaches, and hence it should be considered only in the event of either the DOTSYS or the IBM programs being unavailable to the Atlanta Public Schools.

On the premise that the Atlanta Public Schools shall embark on the development of an integrated learning facility for the blind, the School of Information Science recommends that a serious thought be given to the development (perhaps at a later time) to a comprehensive dictionary in braille, even in the event that the book translation program will use the DOTSYS system. (As pointed out earlier, the DOTSYS system can be employed to produce efficiently a vocabulary in braille.) It is possible to implement, as part of such an interactive multi-media learning facility, advanced information retrieval functions in which such a vocabulary serves the blind student as an access device by means of which he can communicate with and manipulate stores of electronic information. Versions of such a vocabulary can serve as (computer-produced) indexes to the contents of learning materials, as encyclopedias in braille, etc.

V. PROJECT MANAGEMENT

1. Organization:

The project design and its implementation are based on the following assumptions:

- a. The Atlanta Public Schools is fully committed to a continuing project for the production of braille by computer in support of education for the blind.
- b. The Atlanta Public Schools will devote a significant portion of its computer resource to the creation and maintenance of textual materials on a continuing, long-term basis.
- c. The operation of the Atlanta Public Schools computer system is sufficiently stable to ensure the availability and use of ATS from 8:00 A.M. until 5:00 P.M. each work day of each week.
- d. The Atlanta Public Schools is prepared to provide an administrative framework conducive to the implementation of the project on a consistent, cost effective basis.

In order to assure consistent, responsible and effective management, it is recommended:

- a. that the entire mission be organized, planned, funded and implemented on a project basis;
- b. that a full-time project manager be appointed and given the technical and administrative responsibility for the implementation of the project tasks. Among the responsibilities of the manager should be: approval of technical design; authorization of funds; approval or recommendation for appointment of project personnel; and periodic, formal reporting on the technical and administrative

status of the project;

- c. that a project team be established, with individuals committed to the project on a flexible but responsible basis.

The personnel categories required by the project are shown in the budget estimate in a later section of this report.

The recommended project organization and its management may be implemented either in-house by the Atlanta Public Schools, or as a sub-contract to an outside firm. If carried out in-house, the nature of the project is such that it spans several administrative divisions of the Atlanta Public Schools; it would appear that the project staff should be composed of individuals assigned to it from these respective divisions (see Figure 6). It is essential, however, that the project manager have a temporary line relationship with the individuals assigned to the project team. The project manager may report to the chairman of an "executive committee" consisting of the heads of the three affected divisions; since the predominance of the project activities lies in the Instruction Division, the Assistant Superintendent of this Division may logically serve as the committee chairman. (In an alternate arrangement, the project manager can report to one of the division heads who should assume a responsibility for an informal coordination with his peers in the other divisions.)

If a decision is reached to subcontract the project to an outside firm, a part-time "project officer" (at the level of the project manager above) should be designated by the Atlanta Public Schools to supervise and monitor the performance of the contractor. The project officer would report as above.

2. Structure

The entire technical project can be conveniently defined and structured in terms of "tasks"; each task represents a logical unit of the project activities. The project tasks are defined in Table 1, and their structure is shown in Fig. 6 (Project Task Chart).

The division of the project into tasks suggest a convenient method of managing the entire scope of its technical activities.

3. Schedule

The schedule (Table 2) is contingent on decisions resolving the relationship of the Atlanta Public Schools and the M.I.T., and on the availability of DOTSYS as a translator. A subset of these tasks can and should be initiated independently of these decisions; these are Tasks 1-7, 8, 12, 16, and 20.

In the event that the alternate system is implemented, the scope of Task 8 must be expanded, advancing its completion date. Task 15 can be effected within the time frame anticipated; at the same time, the programmers' activity will be changed from that of a liaison with M.I.T. to the modification of the IBM program.

In either case, preliminary braille production should be achieved in June, full scale production in September 1970.

4. Cost Estimate

Table 3 shows the estimated direct cost of the activities previously scheduled for the period of January-August 1970.

Clearly, the recommended system can operate at various levels of productivity. The rate of production of braille is directly controlled by the rate of production of machine processable ATS files of natural

language text. In the budget estimate, the assumed level of ATS file production is 300 natural language inkprint text pages per week. This estimate is based on a typing rate of 40 words per minute, six hours text production per day for four terminals, five days per week. Four terminals will be engaged for an equal amount of time in proofreading, editing and certification. This rate of text production should produce about 1200 equivalent braille pages per week.

Table 1
Braille Production System Project Tasks

TASK NO.	TASK NAME
1	Select textual materials for translation into braille and assign relative priorities for processing
2	Develop procedures and rules for transcription from ink-print to machine processable form via ATS, including editing procedures
3	Convert textual materials to ATS files
4	Proofread ATS files noting errors
5	Edit ATS files
6	Develop sampling and testing procedures for certifying the integrity of textual materials in ATS file format at a specified level of confidence
7	Test and certify ATS files for computer processing to braille input format.
8	Develop specifications and COBOL programs for converting ATS files into braille input format
9	Translate certified ATS files to braille input format
10	Proofread braille input format files noting errors
11	Edit braille input format file errors
12	Develop sampling and testing procedures for certifying the integrity of textual materials in the braille input format, at a specified level of confidence.
13	Test and certify braille input files as ready for processing.

Table 1 (Continued)

Braille Production System Project Tasks

TASK NO.	TASK NAME
14	Store certified braille input format files
15	Evaluate, select and if necessary adapt the computer method for translating input format files to output printed in braille
16	Develop sampling and testing procedures for certifying the integrity of textual materials printed out in braille, at a specified level of confidence
17	Translate certified braille input format files to braille printouts.
18	Test and certify translated braille printout
19	Catalog, index and store braille materials
20	Develop a policy for disposition and use of braille library materials
21	Disseminate braille materials to points of use.

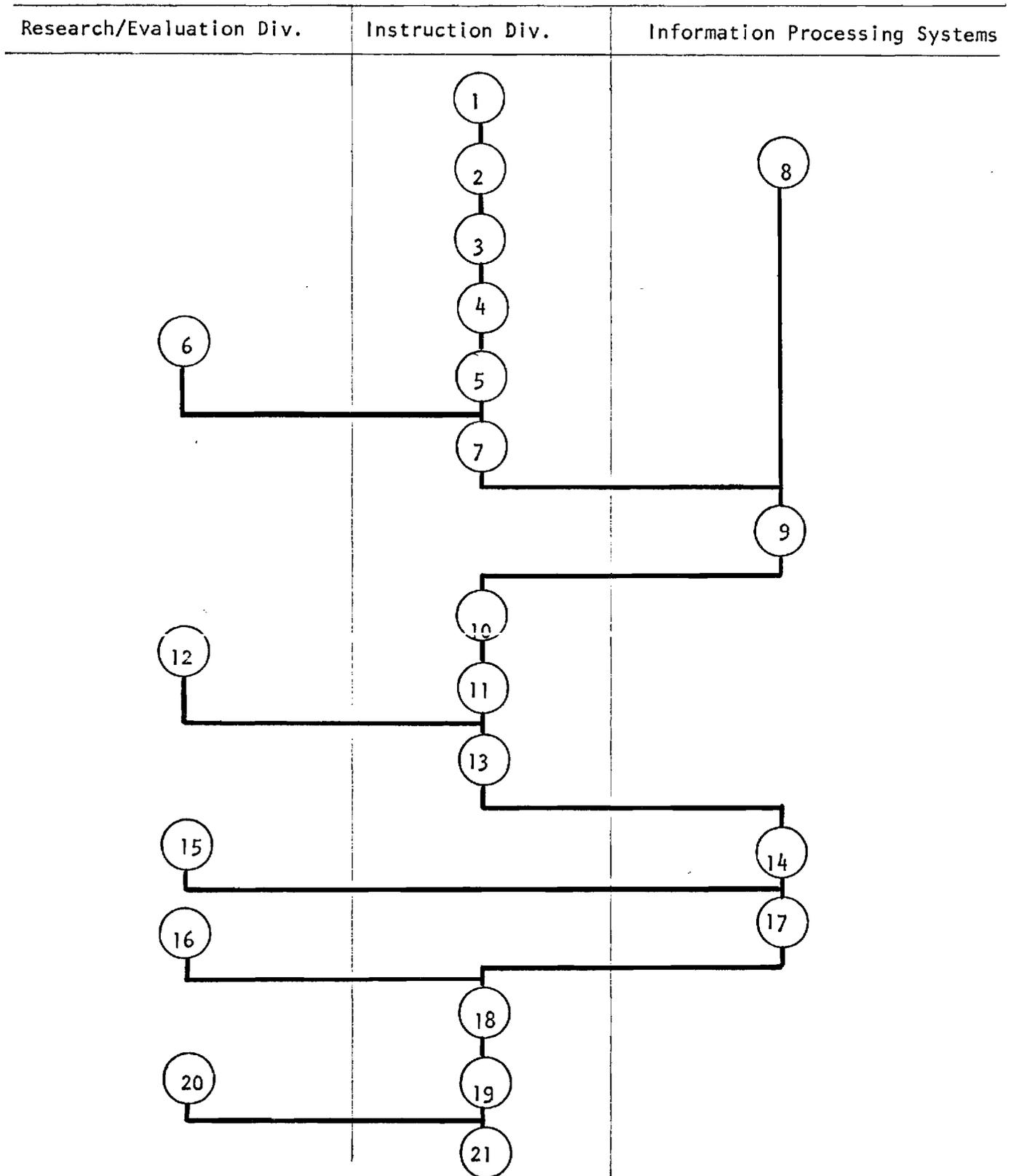


Fig. 6 Project Task Chart

Table 2 Project Task Schedule

TASK NO.	TASK NAME	EARLIEST START	PROBABLE START	EXPECTED COMPLETION
1	Select text	January	January	Continuing
2	Develop ATS Rules	"	"	"
3	Convert text to ATS	"	February	"
4	Proofread ATS files	"	"	"
5	Edit ATS files	"	"	"
6	Develop text certification procedure	"	January	February
7	Certify ATS files	February	March	Continuing
8	Develop & test COBOL programs for converting ATS to b.i.f.*	January	February	March
9	Translate ATS to b.i.f.	March	March	Continuing
10	Proofread b.i.f.	March	April	"
11	Edit b.i.f.	"	"	"
12	Develop b.i.f. certification procedure	February	February	April
13	Certify b.i.f.	April	April	Continuing
14	Store b.i.f.	"	"	"
15	Evaluate, select & adapt b.i.f. translator	January	January	August
16	Develop b.i.f. certification procedure	April	April	March
17	Translate b.i.f. to braille	May	June	Continuing
18	Certify braille	June	June	"
19	Catalog, index & store braille	June	June	"
20	Develop braille dist. policy	January	April	June
21	Disseminate braille	May	June	Continuing

* b.i.f. means braille input format

Table 3

Direct Costs Estimate for the Period January 1 -- August 31, 1970

<u>Personal Services</u>		
Professional		28,000
Project Manager	10,000	
Research Analysts	6,000	
Programmers	12,000	
Non-Professional		42,000
ATS Operators	32,000	
Clerks and Operator Trainees	8,000	
Computer Operator	2,000	
<u>Equipment</u>		12,800
ATS Terminals (8)	12,800	
<u>Materials and Supplies</u>		4,740
Magnetic Tape	960	
Braille Stock	3,080	
Paper (ATS)	500	
Miscellaneous	200	
<u>Computer Time</u>		10,500
		<u>98,040</u>

SELECTED REFERENCES

The following list of references is included as a cross section of the materials reviewed. Several of these references are sources of comprehensive bibliographies. In particular, reference 7 gives comprehensive coverage through 1967. The Research Index published by the American Foundation for the Blind provides a continuing coverage of work in progress.

1. Anderson, Gary B., and David W. Rogers, "An Inexpensive Braille Terminal Device," Communications of the ACM 11:6 (June 1968).
2. Boyer, John J., "Brailletran: A Comprehensive Braille Transcription Program." Proceedings, Braille Research and Development Conference. Cambridge, Mass., Massachusetts Institute of Technology. 1966.
3. Beddoes, M. P., "Toward an Improved Optophone - Experiments with a Musical Code." Research Bulletin of the American Foundation for the Blind, No. 16 (May 1968).
4. Black, William Lawrence, "An Acoustic Pattern Presentation." Research Bulletin of the American Foundation for the Blind, No. 16 (May 1968).
5. Bliss, James C., "An Information Processing Model for Tactile Perception." American Automatic Control Council, Joint Automatic Control Conference, University of Colorado, Boulder Colo., August 5-7, 1969.
6. Dupress, John K., Dwight M. B. Baumann, Robert W. Mann, Towards Making Braille as Accessible as Print. Cambridge Mass., Massachusetts Institute of Technology. (Department of Mechanical Engineering), 1966. (Report No. DSR 70249-1)
7. Goldish, Louis H. Braille in the United States: Its Production, Distribution and Use. American Foundation for the Blind, 1967.
8. Goldish, Louis H. "A Study of Braille Production, Distribution and Use." Proceedings, Braille Research and Development Conference. Cambridge, Mass., Massachusetts Institute of Technology, 1966.
9. Grunwald, Arnold P. "A Braille-Reading Machine." Research Bulletin of the American Foundation for the Blind. No. 16 (May 1968).

REFERENCES (Continued)

10. Mason, Samuel J., and Jan K. Clemens, "Character Recognition in An Experimental Reading Machine for the Blind." Recognizing Patters: Studies in Living & Automatic Systems. Cambridge, Mass., Massachusetts Institute of Technology Press, 1968.
11. Staack, Gerald Francis, A Study of Braille Code Revisions., Cambridge, Mass., Massachusetts Institute of Technology, 1962. (Thesis, M.S.)
12. Thornhill, Daniel Earl, "Translation from Monotype Tape to Grade 2 Braille," Research Bulletin of the American Foundation for the Blind. No. 5 (July 1969).