The paper describes a project designed to develop and implement a computer-based system to be utilized in the design and analysis of vocabulary-controlled dictation material. The system is based on the use of the 1,500 most frequently used words in business communications (as established by Mellinger), and words are coordinated with the theory lessons in beginning Gregg shorthand textbooks. As such, the dictation material generated under these conditions provides a fundamentally structured approach to speed building. The use of vocabulary-controlled dictation material tends to alleviate many of the dysfunctions associated with the use of nonvocabulary-controlled dictation materials. (Author)
A COMPUTER BASED SYSTEM
FOR THE
DEVELOPMENT AND ANALYSIS
OF
VOCABULARY CONTROLLED DICTATION MATERIALS

by
Leona M. Gallion
Professor of Business
Indiana State University
Terre Haute, IN 47809

and
C. Bruce Kavan
Senior Systems Analyst
Indiana State University
Terre Haute, IN 47809

"A Computer Based System for the Development and Analysis of Vocabulary Controlled Dictation Materials" by Leona M. Gallion, Indiana State University and C. Bruce Kavan, Indiana State University.

SUMMARY

Objectives of the Project

The purpose of this project was to design and implement a computer based system to be utilized in the design and analysis of vocabulary-controlled dictation material. As such, the following objectives were established for the successful completion of this project:

1. To select a word base dictionary upon which the dictation material could be predicated. This base was limited to the 1500 most frequently used words as established by Mellinger. The dictionary for these selected words is comprised of:
   a. the word
   b. the syllable count
   c. the lesson the word could first be used in beginning Diamond Jubilee shorthand
   d. the word type(s)—brief form, brief form derivatives, word endings and word beginnings

2. To design the necessary computer software that would comprise the components of the system. Each of these components or subsystems consists of one or more phases with their respective objectives as follows:
   a. to statistically analyze dictation materials
b. to access the dictionary
c. to provide working tools for writing dictation materials

3. To design a computer system with maximum simplicity of operation for use by the novice student user while simultaneously achieving maximum efficiency of computer resources.

Research Basis

Prior research by Gallion and Anderson has established that shorthand learning achievement in beginning shorthand may increase with the use of vocabulary-controlled dictation material. Vocabulary-controlled dictation material is material based upon the most frequently used words in business correspondence. Dictation material presently available in shorthand textbooks is not controlled according to vocabulary level. Thus, a teacher had to individually prepare dictation material in order to utilize vocabulary-controlled material. Therefore, a need existed to design and implement an efficient method to assist in the development of high-quality vocabulary-controlled dictation material.

Methods Used

The words for the word base dictionary were compiled from the 1500 most frequently used words in the Mellinger word list. The other information in the word base dictionary was researched and compiled by the designers from the various Gregg shorthand publications including the beginning shorthand textbook and the shorthand dictionary. All information contained in the dictionary was verified to insure accuracy (see Figure 1).
Table 1

<table>
<thead>
<tr>
<th>Word</th>
<th>Lesson Word</th>
<th>Number of Syllables</th>
<th>Word Frequency</th>
<th>Word Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Able</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>About</td>
<td>15</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Above</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>.5</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Acceptance</td>
<td>21</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Accepted</td>
<td>14</td>
<td>3</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Accident</td>
<td>21</td>
<td>3</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Accidents</td>
<td>21</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Accommodate</td>
<td>20</td>
<td>4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Accommodations</td>
<td>27</td>
<td>5</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Accomplished</td>
<td>20</td>
<td>3</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Accordance</td>
<td>11</td>
<td>3</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>According</td>
<td>11</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Accordingly</td>
<td>37</td>
<td>4</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Account</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Accounting</td>
<td>20</td>
<td>3</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Accounts</td>
<td>20</td>
<td>2</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Acknowledge</td>
<td>23</td>
<td>3</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Across</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Act</td>
<td>5</td>
<td>2</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Active</td>
<td>5</td>
<td>2</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Activities</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Activity</td>
<td>5</td>
<td>4</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Actual</td>
<td>31</td>
<td>3</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>Actually</td>
<td>31</td>
<td>4</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Add</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Added</td>
<td>14</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Addition</td>
<td>27</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Additional</td>
<td>27</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Address</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Addressed</td>
<td>5</td>
<td>2</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Adequate</td>
<td>14</td>
<td>3</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Adjustment</td>
<td>19</td>
<td>3</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Administration</td>
<td>16</td>
<td>5</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Administrative</td>
<td>16</td>
<td>5</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Advance</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Advantage</td>
<td>21</td>
<td>3</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Advantages</td>
<td>21</td>
<td>4</td>
<td>15</td>
<td>2</td>
</tr>
</tbody>
</table>
The computer system components were specified in non-syntactical form. The various phases were coded in either PL/1 or FORTRAN IV to take advantage of the particular structure of the language to the phase application. All phases were catalogued into the core-image library for simplicity of execution.

User oriented documentation was developed for complete use of the computer system.

Results of the Project

The computer based system as developed has been used successfully at Indiana State University. The evaluation component verifies and statistically analyzes the user's dictation material. The system is used first to verify a passage of dictation material which has been coded for use in a specific lesson in beginning shorthand. If a word in the dictation passage is not in the dictionary or is a word that cannot be written either in or prior to the lesson introduction code, that word will be underscored by astericks in the output (see Figure 2).

After all words in a dictation piece are only those among the 1500 most frequently used words and are those which can be written in either the introductory lesson or the previous lessons, the statistical phase will execute. The edited text image is then outputted into standard word groups of 10 for constant-level dictation (indicated by /) and 20 for traditional dictation (indicated by /#/ see Figure 3). The
DEAR SIR:

YOUR LETTER OF MAY FIRST CAME TO ME THIS DAY. IT WILL BE ONLY THREE

DAYS BEFORE I CAN MAIL YOU THE TWO HUNDRED DAILY /01/ PAPERS FOR WHICH YOU

ASKED. IS THIS ALL RIGHT WITH YOU? I WOULD PUT A FAST SERVICE ON

THIS PACKAGE. THE COST WILL BE THREE DOLLARS /02/ FOR THE FIRST ONE

HUNDRED, BUT ONLY TWO DOLLARS FOR THE REMAINING / ONE HUNDRED.

CORDIALLY,

NUMBER OF TEXT CARDS READ -------------------------- 7
NUMBER OF GOOD WORDS ON TEXT CARDS -------------- 60
NUMBER OF WORDS REJECTED ON TEXT CARDS ----------- 8
NUMBER OF GOOD SYLLABLES ON TEXT CARDS ----------- 74
SYLLABIC INTENSITY (GOOD SYLLABLES / GOOD WORDS) - 1.233
DEAR SIR:

I read your letter of May first this day. It will be only three days before I can mail you the two daily papers for which you asked. Is this all right with you? I would put good service on this package. The cost will be three dollars for the first, but only two dollars for the remaining.

CORDIALLY,

NUMBER OF TEXT CARDS READ ----------------------------- 7
NUMBER OF GOOD WORDS ON TEXT CARDS --------------- 61
NUMBER OF WORDS REJECTED ON TEXT CARDS ----------- 0
NUMBER OF GOOD SYLLABLES ON TEXT CARDS ----------- 75
SYLLABIC INTENSITY (GOOD SYLLABLES / GOOD WORDS) -- 1.229
syllabic intensity of the passage is computed as well as the following statistics: percent of words from each lesson, percent of words in each hundred of the 1500 most frequently used words as well as the percent of brief forms, brief form derivatives, word endings, and word beginnings. The number of different words and the percent of different words is also given. Further, all words that can first be written in each of the lessons are listed as well as brief forms, brief form derivatives, word beginnings, and word endings (see Figures 4 to 8).

Importance of the Study

Writing usable shorthand dictation material involves a considerable expenditure of time especially when the passage is to be vocabulary controlled. This is particularly evident when writing for beginning level classes as only a portion of a highly limited vocabulary can be utilized in the construction of the dictation material.

Research has indicated the importance of vocabulary-controlled dictation materials in beginning shorthand. However, very little of this type of material is currently available. In addition, very little material contains useful statistical analysis which can be used in the selection of materials of comparable levels of difficulty.

Thus, this system provides the user an efficient method to verify and statistically analyze dictation material for beginning shorthand classes. This system has been used in shorthand methods classes and workshops as a basis for writing new dictation material.
**Figure 4**

**COMPUTER OUTPUT OF WORD FREQUENCY COUNT BY HUNDRED GROUPINGS**

**SCHOOL OF BUSINESS**  
**SHORTHAND EVALUATION**  
**INDIANA STATE UNIVERSITY**

<table>
<thead>
<tr>
<th>GROUP</th>
<th>FREQUENCY</th>
<th>PERCENT FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41</td>
<td>67.2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6.5</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>8.1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>3.2</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>4.9</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**TOTAL**  
**61**
### Figure 5

**COMPUTER OUTPUT OF LESSON FREQUENCY COUNT**

SCHOOL OF BUSINESS  
SHORTHAND EVALUATION  
INDIANA STATE UNIVERSITY

<table>
<thead>
<tr>
<th>LESSON</th>
<th>FREQUENCY</th>
<th>PERCENT FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>17</td>
<td>27.8</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>8.1</td>
</tr>
<tr>
<td>5</td>
<td>22</td>
<td>36.0</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3.2</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>13.1</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>8.1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>61</strong></td>
<td></td>
</tr>
</tbody>
</table>
SCHOOL OF BUSINESS
SHORTHAND EVALUATION
INDIANA STATE UNIVERSITY

LESSON 08
BE
BEFORE
CORDIALY
DAILY
FOR
GOOD
ONLY
PUT
THIS
WHICH
WILL
WOULD

LESSON 07
ALL
COST
DOLLARS
ON

LESSON 05
ASKED
BUT
CAN
FIRST
IS
LETTER
OF
PACKAGE
SERVICE
SIR
THE
THREE
WITH
YOU
YOUR

LESSON 04
TWO

LESSON 03
DAYS
I
IT
PAPERS
WILL

LESSON 02
DEAR
MAIL
READ
REMAINING
RIGHT

LESSON 01
DAY
MAY
Figure 7

COMPUTER OUTPUT OF WORD FORMS BY TYPE

SCHOOL OF BUSINESS
SHORTHAND EVALUATION
INDIANA STATE UNIVERSITY

LESSON 8

WORD FORMS

TYPE 1
BE
BUT
CAN
FOR
GOOD
I
IS
IT
OF
PUT
THE
THIS
WHICH
WILL
WITH
WOULD
YOU
YOUR

TYPE 2
BEFORE

TYPE 3
CORDIALLY
DAILY
ONLY

TYPE 4
REMAINING
Figure 8

COMPUTER OUTPUT OF WORD FORMS BY COUNT AND PERCENT

SCHOOL OF BUSINESS
SHORTHAND EVALUATION
INDIANA STATE UNIVERSITY

LESSON 8

<table>
<thead>
<tr>
<th>TYPE</th>
<th>FREQUENCY</th>
<th>PERCENT OF WORD FORM BY TYPE</th>
<th>PERCENT USAGE OF WORD FORM BY TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>83.7</td>
<td>50.8</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2.7</td>
<td>1.6</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>10.8</td>
<td>6.5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2.7</td>
<td>1.6</td>
</tr>
</tbody>
</table>

NUMBER OF DIFFERENT WORDS------------------------ 43

PERCENT OF DIFFERENT WORDS---------------------- 70.49