DOCUMENT RESUME

ED 043 025 AL 002 588

AUTHOR Kuno, Susumu
TITLE Computer Aids to Language Instruction.
INSTITUTION Language Research Foundation, Cambridge, Mass.
SPONS AGENCY Peace Corps (Dept. of State), Washington, D.C.
PUB DATE [30 Jun 70]
NOTE 31p.; From the "Final Report on Peace Corps Language Coordinators Workshop," Rockport, Massachusetts, April 26 to May 16, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.65
DESCRIPTORS *Computational Linguistics, *Computer Assisted Instruction, Computer Graphics, Computer Science, Evaluation, Instructional Materials, *Language Instruction, Language Laboratories, Language Tests, Orthographic Symbols, Pattern Drills (Language), Teacher Role, *Textbook Preparation, Transformation Generative Grammar

ABSTRACT

This paper consists of three parts. Section 1 considers how difficult it is to analyze natural languages by computer, and, therefore how difficult it is to evaluate automatically the student's responses to some types of questions in computer-aided instruction systems for language teaching. Section 2 discusses drawbacks of conventional computer aided instruction systems for language teaching and of conventional language laboratories, and presents a picture of an idealized computer-aided instruction system as an extension of classroom instruction--a long range goal which cannot be technologically and financially achieved in the immediate future. Section 3 outlines a more modest goal that is within the bounds of the present hardware and software developments in computer sciences--that is, a system for aiding textbook authors in compiling and revising language textbooks, and for aiding classroom instructors in using textbooks in a way more suitable for their students' needs. Added to such a system is the capability for printing non-standard characters on computers, which will make it possible to produce language textbooks printed in the orthographics of the languages. (Author/FWB)

COMPUTER AIDS TO LANGUAGE INSTRUCTION

Susumu Kuno

0. Introduction

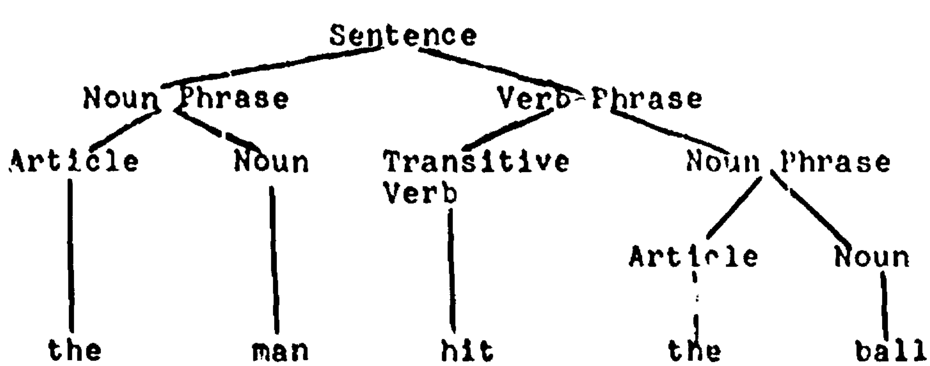
This paper consists of three parts. In section 1, I will discuss how difficult it is to analyze natural languages by computer, and, therefore, how difficult it is to evaluate automatically the student's responses to some types of questions in computer-aided instruction systems for language teaching. In section 2, I will discuss drawbacks of conventional computer aided instruction systems for language teaching, and of conventional language laboratories, and present a picture of an idealized computer-aided instruction system as an extension of classroom instruction --- a long-range goal which cannot be technologically and financially achieved in the immediate future. In section 3, I will outline a more modest goal that is within the bound of the present hardware and software developments in computer sciences --- that is, a system for aiding textbook authors in compiling and revising language textbooks, and for aiding classroom instructors in using textbooks in a way more suitable for their students' needs. Added to such a system is the capability for printing non-standard characters on computers, which will make it possible to produce language textbooks printed in the orthographics of the languages.

1. Syntactic Analysis of Natural Language by Computer

Syntactic analysis is a process which determines the part of speech of each word in a given sentence, and what function each word or sequence of words plays in the sentence. For example, given "The man hit the ball.", it must first be determined that this sentence consists of the following parts of speech:

"Article + Noun + Transitive Verb + Article + Noun"

Furthermore, "the man" and "the ball" should be identified as Noun Phrases in the sentence, that "hit the ball" as a whole plays the role of a Verb Phrase, and, finally, that "The man hit the ball" in its entirety, functions as a Sentence. These syntactic functions can be represented in an upside-down tree form called a phrase-marker:



It is comparatively an easy matter to assign such a structure to a given sentence by hand: we know intuitively that "man" in this sentence is a noun, and not a transitive verb, as in "We man the boat." Similarly, we know that "hit" here is a transitive verb, and not an intransitive verb, as in "He hit well," or a noun as in "The song was a great hit." However, the computer has no basis to decide a priori to what parts of speech these words belong. The computer must try all the other possibilities to figure out to which parts of speech these words belong in the given sentence. The computer must try out all the possibilities to see if there is more than one way of analyzing given sentences.

Syntactic analysis of natural language is carried out by computer using formally stated rules such as the following:

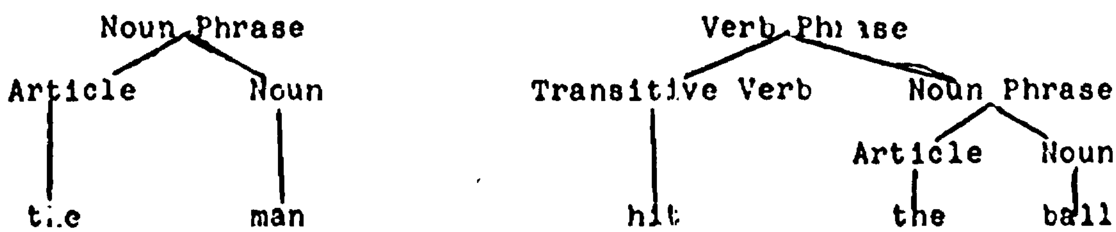
1. Sentence --- Noun Phrase + Verb Phrase
2. Noun Phrase --- Article + Noun
3. Verb Phrase --- Intransitive Verb
4. Verb Phrase --- Transitive Verb + Noun Phrase
5. Article --- the, a, etc.
6. Noun --- man, hit, etc.
7. Intransitive Verb --- hit, walk, cry, etc.
8. Transitive Verb --- hit, accuse, man, etc.

(1), for example, is to be interpreted as "A Sentence has two components: a Noun Phrase and a Verb Phrase." (3) and (4) should be interpreted as "A Verb Phrase consists of either an Intransitive Verb itself, or of a Transitive Verb and its object Noun Phrase."

A computer program for syntactic analysis compares a given sequence of words with the pre-stored set of grammar rules of this or some other form. For example, as one possibility, the computer will definitely select "Article + Noun + Transitive Verb + Article + Noun" for "The man hit the ball." It will recognize "Article + Noun" at the beginning and at the tail end as Noun

Phrases due to Rule (2) yielding

It will further recognize that "Transitive Verb" and "Noun Phrase" can form a Verb Phrase due to Rule (4), yielding



Finally, it will recognize that "Noun Phrase" and"Verb Phrase" can form a Sentence, due to Rule (1). Thus, the computer will successfully discover a phrase-marker, as shown at the beginning of this paper, for the entire sentence. On the other hand, the computer will undoubtedly assume, as one possibility, the part-of-speech sequence

"Article + Verb + Transitive Verb + Article + Noun"

for the same sentence. As before, the computer will identify "hit the ball" as a Verb Phrase. Now, there are no rules in the grammar for combining "Article" and "Verb," or "Verb" and "Verb Phrase," or "Article," "Verb" and "Verb Phrase." Therefore, it cannot arrive at any higher-order phrases which eventually get grouped into a Sentence. This particular analysis path thus comes to an impasse, because of a wrong selection of a part of speech for "man." The computer tries some other sequences of parts of speech to see if there are any other ways of arriving at the topmost Sentence. All computer programs for syntactic analysis have a systematic way of exhaustively checking all the possible analysis paths. They should be able to find two analyses for "They are flying planes": one with "flying" as modifier of "planes" as in "They are planes which are flying," the other with "flying" as a transitive verb with the preceding "are," as in "What are they doing? --- They are flying planes."

The production of all possible syntactic analyses of a given sentence is not a problem with automatic syntactic analysis; the difficulty is producing only the correct ones. In this connection, I should report on our own experiments in computer analysis of English sentences at Harvard University during 1961-64. We developed a grammar containing some 2,000 rules of a form somewhat similar to that discussed above, and a dictionary of about 30,000 inflected word-forms. We analyzed sentences such as "They are flying planes," and our system successfully produced all the ambiguous interpretations. However, it also produced ambiguous interpretations for sentences which to native speakers of English are unambiguous, sentences such as "Time flies like an arrow." Possible parts of speech for the words in the sentence are:

time: Noun, as in "Time is an important factor."
Transitive Verb, as in "We should time the music."
Adjective, as in "Time factor should not be ignored."

flies: Noun, as in "Flies are ugly."
Transitive Verb, as in "He flies airplanes."
Intransitive Verb, as in "This airplane flies well."

like: Transitive Verb, as in "I like apples."
Conjunction, as in "Do it like I do."
Adjective, as in "John's exactly like his father."
Noun, as in "Men of his like are scarce these days."

an: Article

Arrow: Noun

In spite of the fact that the sentence is unambiguous to native speakers of English, the computer system produced the following five analyses:

(i) Time passes as quickly as an arrow. (Correct Analysis)

(ii) You should time (the) flies as you time an arrow.

(iii) You should time (the) flies as an arrow times (the) flies.

(iv) You should time (the) flies which are like an arrow.

(v) There is a species of flies called "time flies" which are fond of an arrow.

It should be easy to tell why these analyses have been produced by computer if one compares them with the parts of speech that each word can potentially play in the language. For example, in the last analysis, "time" is taken as an adjectival modifier of noun "flies," and "like" is taken as a transitive verb, meaning "are fond of," with "an arrow" as its object.

Some of the semantically anomalous syntactic analyses given above can be eliminated by refining the grammar rules. For example, analysis (v) has as the object of "like" the singular form of a countable noun "arrow." Now, there is a general rule in English which says that if a transitive verb (except for the verb "to have" and a few others) is used in the present tense, and if its object is a countable noun, one must use the plural form of the noun unless it is followed by a time adverb. For example, "John reads a book." is ungrammatical, although all of the following sentences are acceptable:

John reads books

John reads a book every morning.

John is reading a book.

John has read a book.

John will read a book.

Since analysis (v) violates this principle, it can be eliminated if rules are added to the grammar to that effect. There are some messy problems in formulating such rules for all constructions related to the one under consideration. Assume, however, that this has been done. Then, this principle is no exception to the principle "All rules have exceptions." Observe the following sentences:

John smokes a pipe.

John drives a sportscar.

If only the above stated principle were to be incorporated into the grammar, it would reject the above two sentences, which are perfectly grammatical. Therefore, it is necessary to incorporate rules for taking care of exceptions. Now, it should be clear to the reader how difficult it is to take care of this single grammatical feature.

Turning now to analysis (iv), note that there is an agreement of number between the subject of "like" and its object. For example,

These children are like monkeys.

is grammatical, but not "These children are like a monkey." Assume that we have incorporated such a rule into the grammar. This rule can be used only when the computer program identifies "flies" of our example as an understood subject of "like," as in "flies are like an arrow." First of all, it is not easy to make such an identification. Second, again, there are some exceptions to this rule. For example,

All my children are exactly like their father.

is grammatical, and the form with the number agreement, "All my children are exactly like their fathers." would be ungrammatical unless that is what the situation calls for. Therefore, more rules to take care of such exceptions.

Analysis (iii) can be eliminated if we pay attention to "time" as a transitive verb. It requires as its subject either a higher animal or an instrument. We can time something, a frog can time something, but a desk cannot. Assume that all the nouns are classified with respect to whether they are higher animals, instruments, or not. An "arrow" does not belong to either of these two classes. Now, in analysis (iii), an arrow is the semantic subject of the verb "fly," which requires as was mentioned above either a higher animal or an instrument, neither of which an "arrow" is. On this ground, this analysis can be rejected. However, this is based on the assumption that the computer program can identify the fact that, in this particular analysis, "an arrow" of "Time flies like an arrow" is the subject of the understood "times (the) flies." This identification is not an easy matter to achieve. Also observe what a task it would be to assign these features to all the nouns in English, and to classify verbs as to whether they can take noun phrases of these features as their subjects. It is easy, comparatively speaking, to do it just for these two features. However, there is no knowing how many such features we would need: several hundreds? several thousands? or even more?

Now it should be clear what a difficult task it is to construct a grammar to obtain the correct and only the correct analyses for given sentences in English, or for that matter, in any language of the world. Assume, however, that this has been accomplished. We would still have an insurmountable problem of semantic analysis. Observe the following sentence:

The toy is in the pen.

The native speakers of English would regard this sentence as unambiguous: however, the computer will produce two semantic analyses to it, although syntactically it is unambiguous: namely, "The toy is in the (play)pen" and "The toy is in the (writing) pen." The reason that the sentence is taken only in the first sense by native speakers is that they know that the writing pen usually does not contain anything except an ink-cartridge, and that they know that a playpen is a fairly large three-dimensional object which can contain objects such as children's toys. This judgment is based on the native speakers' knowledge of this universe. No one has ever succeeded in systematically describing such knowledge in the form that can be manipulated by computer, and there is little hope that anyone will succeed in doing so in any foreseeable future.

I believe that the ability to use foreign languages is an indispensable part of foreign-language acquisition, and foreign language teaching should have, as its integral part, exercises for producing sentences or discourses in the language creatively. The sentences that the students have produced should be evaluated in terms of grammar, meaning, and discourse continuity. Since automatic analysis of natural language is in the state that I have described above, it is next to impossible to evaluate automatically sentences creatively produced by the students. Such evaluation must be left in the hands of classroom instructors. Therefore, although a part of foreign language teaching can be conducted by computer in a computer-aided instruction system, a completely automatic foreign language teaching program, with no classroom instructors personally guiding the students, will not be achieved in our lifetime.

2. Drawbacks of Conventional Computer-Aided Instruction Systems for Foreign-Language-Teaching

The objective of computer-aided instruction in the past has been to use a computer as an instructor and to establish direct communication between the computer and the student. The computer asks a question of the student and the student gives an answer via a scope or a console connected to the computer. The student's answer is evaluated by the computer, which determines which branch of the tree of programmed instructions it should take for the next question. Accurate evaluation of the student's response thus is essential to a successful computer-aided instruction system which incorporates branched instruction. Therefore, computer-aided instruction systems (the CAIS) have been limited to those questions which can be answered "yes" or "no," by multiple choice, by numerical values, or by a small number of fixed paradigmatic answers.

The above requirements of the CAIS impose a very serious limitation on the design of CAIS for foreign-language teaching. First of all, verbal responses by students are out of the range of consideration for such a system. Questions will be limited to one of the following types:

1) Substitution drills: The student is presented with a pattern sentence. Below this sentence, a single word or a phrase appears. The student's task is to generate a new sentence that would result from the substitution of the word or phrase into the pattern

sentence.

2) Cue transformation drills: The student is presented with a sentence. Below this sentence appears a cue, for example, the word PASSIVE. The student's task here is to transform the original sentence into the PASSIVE voice.

3) Fill-in-a-slot: The student is given a sentence with a slot to be filled in: for example,

   "I am interested (   ) music."

   The student is expected to type in "in" as the correct preposition for the slot.

4) Question-answering: The student is presented with a question and a cue for the answer. For example,

   "What is this? (a book)"

   and the student is expected to type in "This is a book."

All these exercises are rote, and constitute an indispensable part of the language-learning process. However, they do not constitute the major portion of this process. For example, the ability to express one's ideas freely can never be tested by the CAIS. Translation from the mother tongue into the foreign language and vice versa constitutes an important part of foreign language learning. Again, since the CAIS cannot evaluate free translation, nothing but elementary translation drills with constraints on the vocabulary and the syntax of the target sentences can be given in the CAIS. The problem here is the following:

there is more than one way of translating a given sentence into the target language. Since it is impossible to prestore all the possible translations in the system, the student's translation will be compared with only a limited number (usually one) of prestored possible answers. If the student's answer is correct, but is not one of these prestored answers, he will receive "incorrect" from the system, and will not know whether his sentence is really ungrammatical or not. This problem is a source of great frustration on the part of the student. One system I know of includes some kind of a syntactic and morphological evaluation routine for checking to see if the student's answer is correct or not, and if incorrect, at what point in the sentence the error occurs. However, on the basis of what I have described in the previous section, the present state of the art of linguistic research and computer analysis of natural languages, one cannot expect that such an evaluation routine will be able to handle anything but the most rudimentary translations.

In summary, the existing CAIS for foreign-language teaching suffers from the following drawbacks (which are not independent but mutually dependent):

1) Only a small portion of what constitutes language-learning is tested and evaluated.

2) Questions that can be asked are of very limited types.

3) The student tends to get bored with questions of the same type.

4) Verbal aspects (especially verbal responses) of language learning are completely ignored.

5) In consequence,the CAIS for foreign-language teaching would be useful only for an elementary level (at most, through the end of the first year of non-intensive study).

Now, let us turn to conventional Language Laboratories (LL). The objective of language laboratories is to supplement classroom instruction. However, in most institutions which use the LL facilities, the LL has become not an extension of the classroom, but an isolated, nonintegrated portion of language learning. Most students have bitter and frustrating experiences with their use of LL facilities. These, I assume, stem from the following drawbacks of the LL systems:

1) Nonintegration of the LL with classroom instruction. Classroom instructors usually do not have time to monitor the students' use of the LL. Monitoring is a heavy requirement on the instructor and also on the students. This would nullify one of the few advantages of the LL -- that students may elect to use the LL facilities whenever they choose to do so. Since the LL sessions are conducted without the presence of any monitors, or with the presence of monitors who in many cases are not classroom instructors, little feedback is given to the classroom instructors regarding the students' performance in the LL.

2) Individual needs are ignored.

The same sequence of drills is given to students of all levels of achievement. For advanced students, repetition of too many drills of the same type are boring; for less advanced students, the drills may go too fast. Some students may require more drills on a certain grammatical pattern, and others may require fewer because they have already mastered the pattern. The LL system does not discriminate between these two groups, and thus tends to push the first group and frustrate the second.

3) No evaluation of students' performance.

Since the students' performance at the LL is not evaluated at all, they do not know whether their performance is up to the expected level or not.

4) Limitations on drills.

Drills given in the LL are all of the rote type (since the correct answer is always given by tape after the students have given their own answers, there is little that a monitor can do except to correct pronunciation. Therefore, the need for monitors for supervising students' LL sessions decreases drastically after the first several weeks of the first-year study of the language). The LL does not test students' ability to express what they want to express in the foreign language.

No drills for free translation or free conversation are given because the correct answers cannot be prestored on tape. No drills for spelling are given, either.

5) In consequence, the classroom instruction will proceed as if there were no LL sessions unless the students' difficulties or proficiencies at the LL were also manifested clearly in the classroom.

I visualize that future computer-aided language laboratories (CALL) will have the following organization: each console for student use will consist of a tape-recorder unit, a teletypewriter, and a display scope. Modes of communication between the computer and the students will thus be multi-medial, and the most efficient mode(s) will be selected for each objective. For example, the tape will give the verbal explanation of a grammatical pattern that the student is expected to master in the current lesson. To reinforce the student's memory, visual explanations such as sentence structure diagrams, some pattern sentences, and so on will be displayed on the scope. Drills may be given both in the aural mode and the visual mode. Students may be asked to type on the teletypewriter what has been said on tape, thus making it possible to practice dictation and spelling.

Student answers that can be evaluated automatically by the computer will be evaluated immediately. Depending upon the student's performance for such drills, different branches of instruction will be followed by the system, thus enabling the

LL system to meet each individual student's need. Answers that cannot be automatically evaluated on the spot will be accumulated, and will be given to classroom instructors for later evaluation. Statistics of the student's performance in the LL will be compiled by the system, and at the end of the session, each student will be told by the system how well or poorly he has done. In addition, such statistics will be given to the classroom instructors so that they know what difficulties the students have had in their "home-work." These statistics may contain diagnoses showing where in drills the students have made common errors, thus suggesting future modifications of the LL materials and/or classroom instruction materials. Communication between the computer and classroom instructor will not be unidirectionally limited to that from the former to the latter. The instructor, having observed the students' performance in class, might instruct the system to emphasize or deemphasize drills of certain grammatical patterns. Even further, he may instruct the system to do so for each student. Thus, the CALL will become an integral part of the whole language-teaching program.

The CALL system will have the following advantages over the CAI system and/or the LL system:

1) The students will have a motivation for using such a system because they would realize (a) it is an integral part of their language program, and (b) that they are being looked after individually.

2) Various types of drills and questions can be handled by the system, and therefore, the system would be less monotonous and boring.

3) Various types of communication media are utilized, and therefore, the system can reinforce students' learning of aural, oral, visual, and orthographical aspects of the language.

4) The system can automatically evaluate some portions of students' responses, and the students will be graded on these responses. Each will know his level of achievement. Reward and punishment is an indispensable factor for learning.

5) Statistics of students' performance will be given to classroom instructors, enabling them to modify course materials for the next meeting accordingly.

6) Since classroom instructors are available for off-line evaluation of students' performance, sophisticated exercises such as free translation and free conversation can be handled.

7) Classroom instructors can control the CALL sessions for each student according to his performance in class.

8) Since the CALL system will have branch-instruction capabilities, it will meet the individual need of individual students.

9) Each student may want to learn the additional vocabulary or expressions needed for communication in his special field. For example, an economics student would want to learn technical terms in economics as well as the basic words which all students are required to learn. The CALL system can meet such a demand.

10) The system would relieve classroom instructors from drills that can be conducted outside the classroom, and enable them to concentrate on the subjects that can be discussed and drilled only in the classroom. Thus, they can reduce the number of classroom meetins per week, while conducting each meeting to its best advantage.

3. Intermediate steps toward the CALL system

Modern computer technology has not yet advanced to the stage in which the CALL system described in the previous section would be practicable. First of all, a display scope is still an expensive device, both in terms of its own hardware cost and in terms of the costs for driving it by the central computer. High-speed access micro-film readers, on the other hand, would be inadequate for our purpose because of branched instruction and the constant need for revising instructions. Secondly, coordination between branched instruction and audio tape is not an easy matter. Tape-recorders are designed for sequential playing, and are extremely inadequate for looping or branching on any sophisticated level. On the other hand, storing the information on audio tape in the digitized form would require an enormous amount of storage space, thus making it economically unfeasible. Research in vocoders has not advanced yet, and will not in any foreseeable future, to the stage in which it is possible to produce a natural speech directly from any formulaic representation of sentences. Further, the hardware and software costs for the computer system at present would be beyond the reach of most institutions engaged in foreign language teaching.

For the reasons mentioned above, I feel that the time is not yet ripe for conducting a large-scale experiment for research and development of the CALL system and the evaluation thereof. Instead, we have undertaken, sponsored by Peace Corps, a project with a much more modest goal -- a development of a computer system as an aid to authors in their compilation and revision of language textbooks, and to classroom instructors in preparing course materials for the next class meetings.

(1) Step 1: CAI system as an aid to language textbook authors

This is essentially an editing program which can amass statistics on word occurrences in the text. Elementary language textbooks have requirements such as (a) a given set of words or expressions must be introduced in the first year, second year, and so on, and (b) a given set of grammatical patterns must be introduced in the first year, second year, and so on. Before the final version of a textbook is completed, the author goes through the repeated processes of revising various portions of the textbook. Each time a major revision is made, new statistics of word and grammatical pattern occurrences of the textbook must be compiled and compared with the required occurrences. Also, each time a revision is made, portions of the text that are indirectly affected by the revision because of line, paragraph, and page changing must be retyped, thus introducing a new source of typing errors. The CAI system as an aid to language textbook authors will enable authors to revise their preliminary drafts at will, and will produce statistics of word and pattern occurrences, and will take care of typesetting problems as well. The system also can give suggestions to authors in the following manner: when a word is introduced in a given lesson, it should not be the last occurrence of that word. The same word should be reintroduced in later lessons for reinforcement. The system can remind authors that such and such a word should be used in the lesson that the authors are now writing for reinforcement. Or the system can compare the words and patterns that have been introduced in the lessons that the authors have written so far, with the list of words and patterns that must be introduced by the end of the textbook, and remind them that such and such words or patterns must be introduced in the current or subsequent lessons. By the time that a textbook is completed, the authors will have a complete index of the vocabulary and the grammatical patterns. The system can also have the following capability: by a given set of control words, the authors can print out only the selected portions of the textbook compiled. This capability will enable authors to prepare a master textbook, and then produce a textbook for student use, a teacher's manual, and an exercise book, or various types of student textbooks, depending upon how many hours, for example, can be used for each lesson.

(2) Step 2: The CAI system as an aid to language teachers in preparing for the next class

The system lets the instructor compile a scenario, so to speak, for the next class meeting, which he thinks fits in best with his students' needs. If, based on the previous class meeting, the instructor thinks that his students need more drills on a certain grammatical pattern, he can get them from the system's resources by requesting them on the console of a time-sharing system. He can also compile homework assignments meeting students' needs. Thus, from the same resources of the system, flexible, dynamic, non-monolithic textbooks which reflect students' requirements, their levels and their achievements at present can be compiled.

To be combined with these programs is a special program, which has already been developed at Harvard, for computer printing of non-standard orthographies. This is not a place to go into any detail on how non-standard characters are printed by computer. Suffice it to say that a text in non-standard orthographies is displayed on a cathode-ray tube of a special computer as an image is displayed on a TV screen, and the microfilm and hardcopy of the image on the screen is produced by a built-in camera automatically. The special computer which is currently used for this purpose can produce a printout of a 3,600 Chinese-character text in three seconds. Our current program has the repertoire of 10,000 distinct Chinese characters, all Korean characters, Japanese characters in two styles (Katakana and Hiragana) Hebrew, Persian, Tamil, Greek and some other orthographies. The orthographies of languages needed for Peace Corps language training will be added to the repertoire.

The text-editing program previously described and this program for printing non-standard orthographies combined would produce a very powerful system for editing and printing language textbooks in non-standard orthographies. For example, a textbook can be printed by computer in the orthography of the language, with translations and grammatical explanations all in alphabet. From the same text, the computer system can produce a transliterated text if desired, and the transliteration can appear on the right-hand page corresponding to the orthographic representation on the left-hand page, or inter linearly, or in a separated volume. The microfilms produced by computer of any such texts can be used as they are on microfilm printer to produce as many copies as are desired. Or, the hard copies can be used for photo-offset printing.

For the benefit of those who are not familiar with computerized printing in general, I will conclude with an outline of what it is and how efficient it is in terms of time and money.

4. Computerized Printing

In the past few years, advances in the technology for computerized typesetting have been such that more than 200 computers are now in use or on order by the printing business throughout the world. Nearly all the major U.S. computer manufacturers have entered this field, and competition for the market is keen.

Although newspapers have been the primary practitioners of computerized printing, book publishers and government agencies have also begun computerized operations. In its book publishing application, a typical system would consist of the following operations:

1) The author types his manuscripts in the customary way and hands them to the publisher.

2) The manuscripts are punched on paper tape (or on punched cards and then converted to magnetic tape), together with control words for capitalization, italicization, bold lettering, indentation, line-changing, etc.

3) The paper tape thus produced is run on a photocomposition machine. The printer takes care of the justification of printed lines by simply adding the width of the characters and spaces in each line, and comparing the sum with the column width, the computer is able to apply the proper spacing techniques (e.g., insertion of thin spaces, ens, ems, or hyphenation) for justification. A microfilm of the text is produced.

4) A hardcopy of the microfilm is printed.

5) The author indicates on the hard copy what changes he desires to be made.

6) Only the necessary corrections are punched up, together with the control words indicating where these corrections are to be made.

7) The paper tape for the corrections is merged to the original paper tape for the whole book, and a new revised paper tape is produced.

8) Steps 3-7 are repeated until no more corrections are necessary.

There are several advantages for use of such a system for printing.

(a) The fastest line-casting machines are capable of an output of only 15 newspaper lines a minute, whereas the newest photocomposing machines are capable of printing 1,000 - 2,000 lines a minute.

(b) Since no line-casting is involved, it is easy to make corrections. It does not matter whether one word in a specified line is to be replaced by another word or by ten words, or by several sentences. The replacement takes place not in terms of physical typecast, but in terms of punched paper tape. When the revised punched paper tape is produced , the photocomposition machine will take care of line-changing, page-changing, etc. automatically.

(c) When the text is such that it undergoes constant changes but the bulk of the text remains the same, computerized printing is a must. Otherwise, the text would have to be typed manually from the beginning to the end, necessitating the proofreading of the entire newly typed text and the correction of newly introduced errors.

We do not have any extensive statistics of costs for computerized printing. The fact that many publishers of professional books and journals are using the computerized printing system is witness to its efficiency in speed and cost.

The need for computerized printing becomes more acute when non-standard orthographies are involved. For example, a textbook in Tamil will have to be hand-written if the Tamil orthography is to be used. Updating such a textbook would be a highly expensive operation because the whole textbook would have to be manually rewritten. In such a case, our computer graphics system plays an irreplaceable role by producing a print-out both in the Tamil orthography and its alphabetic transliteration.

The following is a rough figure for the operations involved at Harvard for computerized printing in two modes: one printing of standard characters using ordinary chain printers, the other printing of non-standard characters using a special computer (Stromberg-Carlson's 4020) for computer graphics.

A. Initial Keypunching of a 300-page Textbook

(i)	Assignment of computer control words Encoding of non-standard characters if there are any, etc. at 20 pages/hour for $3/hour		$	45
(ii)	Keypunching at 3 pages/hour for $6/hour		$	600
(iii)	Printing for Proofreading 1 hour to print on IBM 360/30		$	40
		Total	$	685

B. Printing of a 300-page Textbook by Chain-printer

(i)	Printing on Multilith paper plates 1 hour to print on IBM 360/30		$	40
(ii)	$ .10 per multility plate x 300		$	30
(iii)	100 copies each of 300 multilith plates at $ .015/copy		$	450
		Total	$	520

C. Updating a 300-page Textbook to the extent of inserting a paragraph every 3 pages

(i)	Assignment of control words 3 hours at $3/hour		$	9
(ii)	Keypunching of updating data cards 10 hours at $6/hour		$	60
(iii)	Updating 20 minutes of IBM 360/65		$	83
(iv)	Printing (by chain-printer) for proofreading		$	40
		Total	$	192

D. Printing of a 300-page Textbook by Stromberg-Carlson's 4020 or Comparable Computer*

(i) Processing for Formatting 30 minutes of IBM 360/65**		$ 125
(ii) Test Printing for Checking 50 pages at $ .75/page		$ 38
(iii) Final Printing on S-C 4020 at $ .50/page hardcopy and microfilm		$ 150
(iv) Microfilm Printing of 100 copies at $ .02/copy		$ 630
	Total	$ 943

E. Comparison of computer and human updating of a textbook (Assume that a 300-page textbook undergoes 5 revisions, 100 copies printed for each of the total six versions)

I. Computer Updating

(i) Initial Cost (see A)		$ 685
(ii) 5 Revisions (See C) $ 192 x 5		$ 960
(iii) 6 printings (See B) $ 520 x 6		$3,120
	Total	$4,765

---

* In case textbook contains non-standard characters.

** See note at conclusion of summary.

II. Manual Typing

For each version,

(i)	Typing at 3 pages/hour at $4/hour	$ 400
(ii)	Proofreading	$ 100
(iii)	Printing (Photo-offset) 300 plates at $0.10/plate	$ 30
	100 copies of 300 pages at $0.015/copy	$ 450
	Total	$ 980

6 versions altogether: $ 980 x 6 = $5,880.

From the above calculationsit would appear that the advantages of the Computer-Printing system lie in the ease with which editing may be done. Also, this system has the following features which make it preferable to hand-typing.

(1) Outputs only the specified sentences or paragraphs.

(2) Locates words or phrases and outputs the list in alphabetical order as aid to index creation.

(3) Creates an alphabetical dictionary of words used, with a frequency count of each as another aid to index creation and also is good as a spelling check.

(4) Right-hand justify text for a cleaner-looking print. Justification may be stopped for output which one wishes to leave "as is." (As in a vocabulary list, for example.)

Note

This figure is based on the assumption that the average number of complex non-standard characters (such as Chinese) per page of the Textbook is 40.

Processing Textbook tape for formatting for printing use on an SC4020 takes a comparatively large amount of time if the textbook contains many characters of complex configurations, e.g., Chinese characters. For example, our present program takes 1 minute of IBM 360/50 time to format a page of 1,200 Chinese characters. We are in the process of making the program work more efficiently.