

LEARNING ENGLISH WITH *IT'S ENGLISH*

J.G.M. Jaspers, G. Kanselaar, W.A.M. Kok
Department of Educational Sciences, Utrecht University
Mailing address: Prof. dr. Gellof Kanselaar
Department of Educational Sciences
Heidelberglaan 2, 3584 CS Utrecht, The Netherlands
Tel: 31.30.534893, Fax: 31.30.532352
Email: kanselaar@fsw.ruu.nl

ABSTRACT:

This paper describes a new Computer Assisted Instruction program for foreign language (english) teaching based on a communicative approach.

Due to the greater mobility and the closer ties in the European community, foreign languages have become an even more important subject in the school curriculum. Based on the results of a study in 1989 [3], a new Computer Assisted Instruction program for foreign language teaching (English) has been developed. Main features of this program are: the communicative approach, a 70.000 word dictionary, multimedia and a syntactic parser.

BACKGROUNDS IN THE DEVELOPMENT OF *IT'S ENGLISH*

To a large extent foreign language teaching is directed towards communicative goals. Socio-cultural developments (mass media, holidays abroad etc.), scientific developments in teaching methodology, the use of new media, linguistics and psycholinguistics have led to a more functional approach to foreign language teaching.

The traditional grammar-translation method, in which vocabulary was learned mostly by memorizing L2/L1 language pairs and in which grammar was taught as a system of rules, did not directly lead to fluent oral proficiency.

In developing the CAI programme we chose the communicative approach to foreign language teaching. In this approach language is viewed as an instrument of human behaviour. The behavioral "situation" is carefully selected in teaching. The linguistic context, the situation, speakers' roles and

types of texts are all taken into careful consideration in the communicative approach. Linguistic function is seen as more important than linguistic form. The structuring of the learning materials is first and foremost determined by communicative needs.

When designing the learning materials for the CAI programme we used Neuner's typology as a starting point for the construction of exercises. The three types of exercises in the programme, including receptive, reproductive (cloze texts) and productive (free sentence input) exercises can be placed on this continuum.

BASIC AIMS AND ASSUMPTIONS

The aim of the project is to create a computer-assisted learning environment for communicative foreign language teaching [1]. This form of foreign language teaching is oriented towards an active and passive command of a foreign language in everyday situations. It avoids translation as the only or the most important means of conveying meaning. Vocabulary is acquired on the basis of context-specific sentences and meanings in the foreign language, to ensure the development of a language system that is separate from the learner's native language. In addition to the syntactic and morphological characteristics of words, their pronunciation is of great importance, as is sentence intonation. The sound of a new word is a basic element in assigning meaning to that word. In order to effect this a great number of words and sentences were stored on CD-ROM. The aim of communicative language teaching is mainly achieved by having students use the non-native language productively. In order to achieve random language-input a sentence parser was developed for the program. This makes use of the grammatical information contained in the CD-ROM dictionary.

The combination of dictionary, phonetic word-representations and parser makes it possible to create different receptive and productive exercises in this interactive learning environment.

The software program IT'S ENGLISH is primarily aimed at teaching vocabulary. However, as an intelligent tutorial system are working on making it usable in the teaching of writing. In the following sections the program itself will be described.

A short history of It's English

During a previous study [3] a program called "PRESTIGE" written in SD-PROLOG and running on MSDOS was created. This program allowed, among other things, students to enter their own text. This limited dictionary (± 2000 words) proved too small. Student often used words not contained in this small dictionary.

To overcome the problem of the limited dictionary we used a 70.000 word dictionary that was made available to us. The size of this dictionary however and the addition of sound capabilities to the program made it inevitable to use another storage medium. At the time the CDROM-XA standard was the most adequate choice. This resulted in a CDROM-XA disk which held both the dictionary databases and a great number of soundfiles. This version of the program was tested in four groups of students in secondary education [7].

The MSDOS version of the program had two major drawbacks: the limited amount of memory available to the program and the limited capabilities of the 80*25 character screen. In order to remove memory limitations and improve on the user-interface a new version of the program is being developed. The new version runs under Microsoft-Windows version 3.1 in enhanced mode. This version was implemented in Prolog-2 for Windows, with some extensions in TurboPascal for windows. This transition removed the memory problem and improved the user-interface.

The CDROM-XA standard for sound however is not supported by the windows multimedia environment. Conversion of the soundfiles to the standard multimedia format of Windows (.WAV-files) is currently in progress.

DESCRIPTION OF THE 'IT'S ENGLISH' PROGRAM.

A computer program that connects with the above basic aims and assumptions should enable the user to retrieve semantic definitions and context sentences in English, as well as synonyms, antonyms, pronunciation and the syntactic and morphological properties of words. In addition such a program should be capable of evaluating the user's use of language to enable him or her to apply the vocabulary in (re)productive exercises. To achieve all this the program consists of several components:

1. An automatic sentence parser which checks random language input for syntactic, orthographic or inflectional etc. correctness.
2. A CD-ROM database, which contains definitions and context sentences in everyday-English, as well as the grammatical properties and synonyms and antonyms of over 70,000 words.
3. A CD-ROM database with the digitized pronunciation of over 5000 words, 1200 context sentences for the presentation of phonetic word representations and English intonation.
4. An exercise generator, which generates exercises for passive, active and reproductive language use.
5. A feedback-generator, which uses the domain-input of the parser to reveal and remedy learner errors in (re)productive exercises.

The Collins Cobuild English Language Dictionary [2] was chosen as a basis for the lexical CD-ROM database. This dictionary meets the following requirements. It was compiled for 'non-native learners of English'. Each word is accompanied by a semantic definition in simple English, and a number of authentic, explanatory citations. In addition, each word meaning is accompanied by systematic and moderately detailed grammatical information, for use by the parser. This direct connection between grammatical information and the semantic section makes the meaning of every word that is recognized by the parser instantly available for retrieval, as well as a number of context sentences and other relevant information such as synonyms, antonyms and superordinate

terms. One inherent restriction in the application is that the parser's "power", that is, the number of English structures that can be automatically parsed, is limited by the grammatical information available in COBUILD. An important advantage of using such an extensive dictionary is the availability of information on nearly all English words. In previous studies [3] we found that in productive exercises with limited dictionaries (2000 words) pupils often use words that are not in the selected subset of English.

CD-ROM AS A STORAGE MEDIUM.

Until recently it had not been possible in computer-aided education to devote much attention to foreign-language pronunciation, in view of the large amounts of external memory involved. To a large extent the arrival of CD-ROM has removed this restriction. In this project CD-ROM is used as a storage medium both for digitalized pronunciation and the lexical database based on Collins Cobuild Dictionary. The choice of the Extended Architecture (XA) standard for the CD-ROM in our project is exclusively related to the audio-component.

The most important considerations in the choice of the XA-standard for the audio-component were the following:

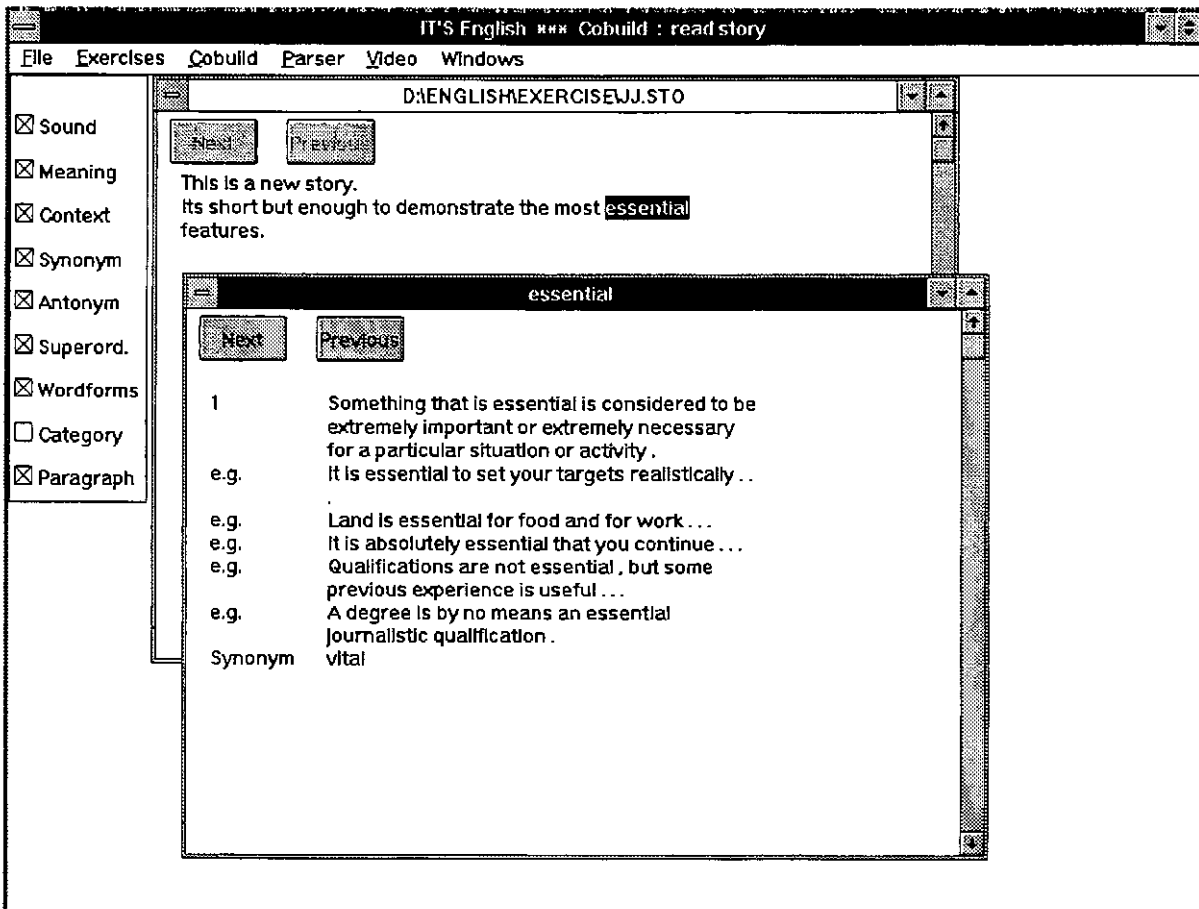
1. The sound-quality had to be optimal. In particular the absence in speech of the higher sound-frequencies (3500 - 11000 Hz) makes it very difficult to distinguish the pronunciations of certain words in English, whereas more or less subtle sound differences in fact signal differences in meaning. This is especially important in speech from female speakers, because their speech is naturally richer in higher frequencies.
2. In the acquisition of a good foreign-language pronunciation it is of importance for learners that the phonetic representations of words should be presented not only in isolation but also in sentential contexts. The learner should experience the pronunciation of words as being connected with and influenced by their contextual position as well as by the intonation of the sentence.

PROGRAM EXERCISES.

The exercise generator.

The generator is capable of generating different types of exercises on the basis of texts stored in the database. Exercises are been developed for the program which range from reading exercises to free writing. These exercises follow a route from receptive via reproductive to productive language use, and may be completed in a teacher- or learner-determined sequence. Three types of exercises have been developed so far in the making of the program. These are prototypical of the type of application that is possible with such programs. In developing these exercises we have followed Neuner [4], who distinguishes three main phases in an exercise sequence that leads to communicative language use. As a result of the newly-developed parser, adding exercises in more elaborate versions of the program is a relatively simple matter.

Figure 1: A reading exercise.



A. Reading texts (receptive use).

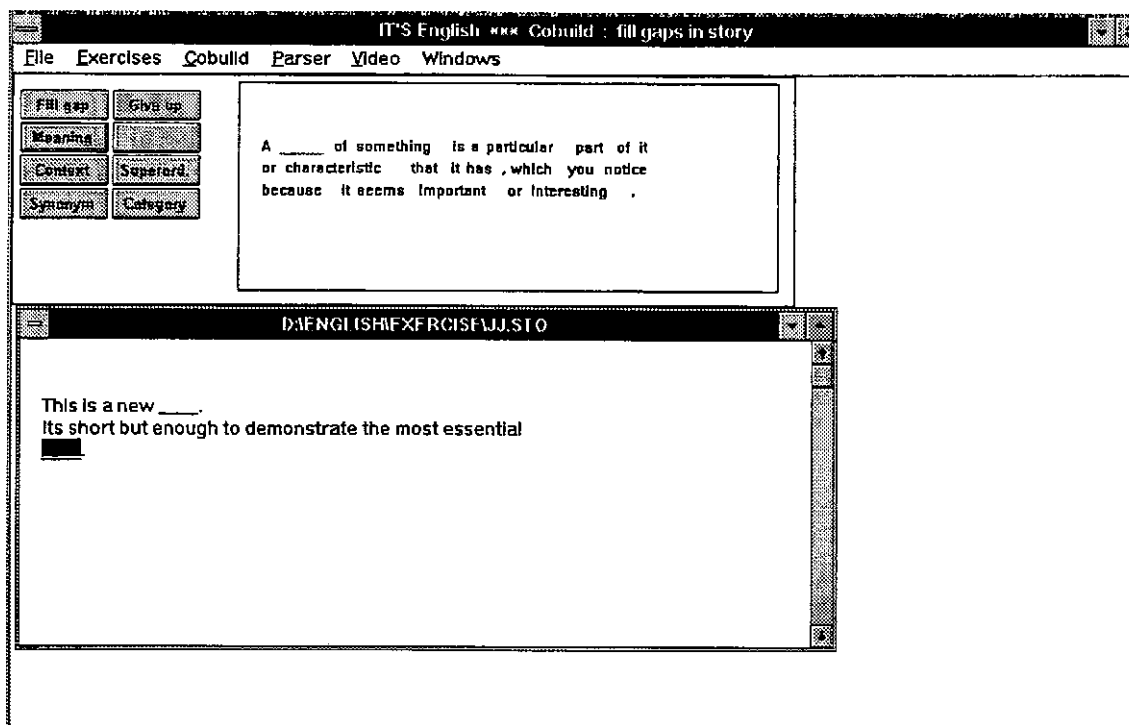
A fragment of a story is selected by the student or teacher from the text database. Learners are free to retrieve a definition of each text-word in English, or the context sentence that accompanies each word. As soon as a text has appeared on the screen, the user can click on any word on the screen to search for the word in the lexical database. Definitions and contexts are retrieved and presented. If the selected word occurs in the sound-database that soundfragment will be presented to the student. The first meaning paragraph from the dictionary will be shown. Appropriate synonyms, antonyms and superordinate terms will be displayed. The user can select the type(s) of information by enabling options in the box in the upperleft corner of the screen. The NEXT and PREVIOUS buttons enable the user to browse through the available information (see figure 1). It will also be possible for the teacher to mark a word in advance to indicate that the learner should call up information about that word later. In this

process word-specific feedback can be provided by the teacher.

B. Texts with blanks (reproductive use)

In the creation of texts with blanks one or more properties of a word, all of which are recorded in the parser's property-list, may serve as a criterion for the omission of that word from the exercise, or for its inclusion in it. For example, texts could be created without any prepositions or adjectives. The decision to leave out all nouns, verbs, prepositions or adjectives can be taken by the teacher or the learner. Sentence constituents such as verb phrases can also be left out. This type of filling-in exercise is clearly superior to the cloze-texts of current prevalence, in which only every n-th word may be omitted. Since the database can be put into operation in the analysis of the learners' answers, feedback can be provided which is aimed at various kinds of errors. This makes IT'S ENGLISH different from traditional programs, in which every erroneous answer, plus the accompanying feedback,

Figure 2: A gapfilling exercise



has to be anticipated and dealt with. The type of feedback that is provided by IT'S ENGLISH for filling-in exercises can be of the following kind: the typed-in answer is synonymous with a superordinate concept; the word class is correct but the actual word is not; information about possible 'typing errors' which is based on an algorithm; information about conjugation and other grammatical characteristics. It is also possible to provide help by giving a context sentence provided by the dictionary from which the target word has also been omitted (see figure 2)

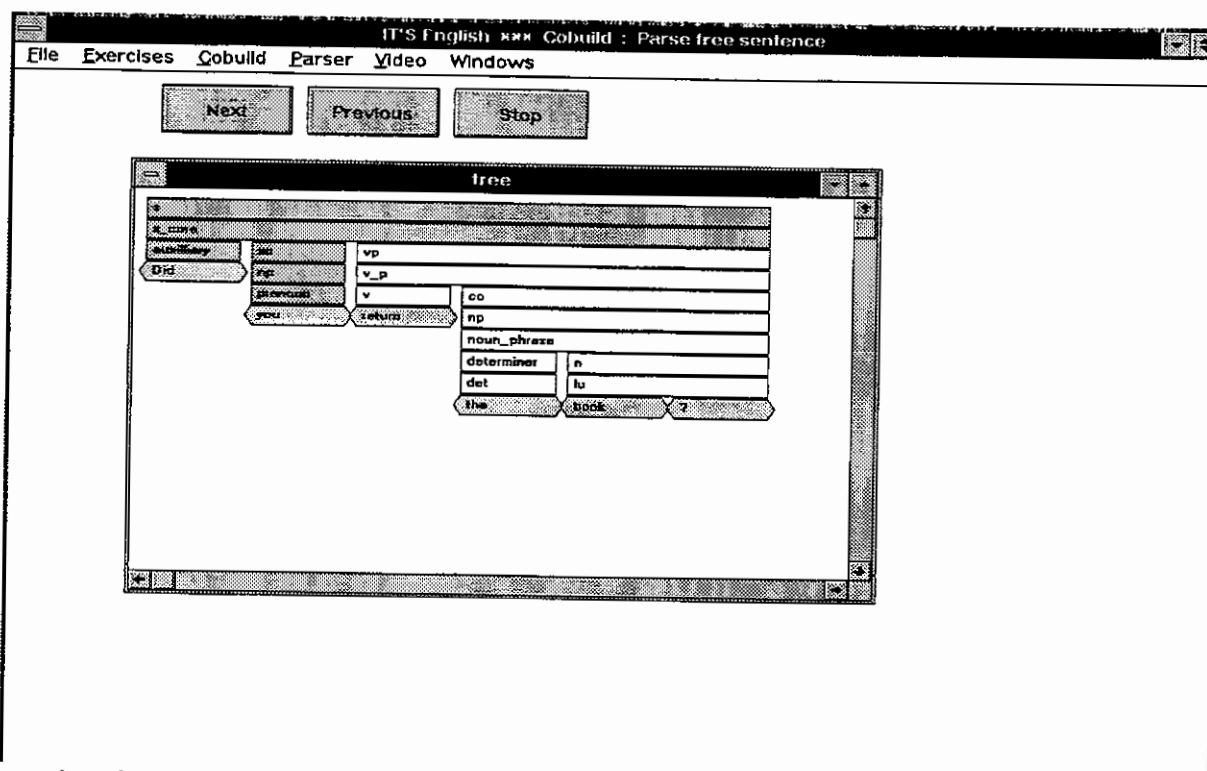
C. Random text-input (productive use).

This type of exercise allows the learner to type in a story. The parser checks each sentence for grammatical correctness. Parsing of natural language is quite a complicated process. The current version of the parser knows over 400 rules to connect words to form grammatical sentence-constituents. The current version of the parser is able to analyse some quite complicated sentences. Most parsers are built around a relatively small dictionary where the information is tailored to the needs of the parser. Our parser is completely dependent on the information from the Cobuild dictionary. One consequence is the programs

inability to obtain gender information from the dictionary.

For a parser to be able to pinpoint an error in a sentence we will have to adapt the parsing mechanism to generate a hypothesis for possible errors. This will require extensive modifications to the parsing mechanism and research into the most common errors of students. When an error occurs a feedback message will be generated and the learner is given an opportunity to correct the error. The final story is stored in the text database, and can thus serve as a basis for new exercises. This enables learners, among other things, to create fill-in exercises for each other. All the help options based on the dictionary that are available for the receptive exercises are similarly available for this type of productive exercise. The structure of a sentence can be made visible in tree-form on the screen. This possibility of random language input has one drawback: it makes great demands on the speed and capacity of the computer. The dictionary generates all possible grammatical uses of each word in a sentence. This gives rise to a combinatorial explosion of possible ambiguities. This slows down the parsing process. We are looking into ways of optimizing the parsing process to cope with this problem. However, in our opinion the program is uniquely valuable at present precisely because of its

Figure 3: A parsed sentence



capacity for processing random input, with graphical feedback about the grammaticality of input-sentences (see figure 3)

In the teaching of writing various aspects of writing practice can be dealt with, such as planning and organization, formulation and revision. The program offers help in formulation at sentence level. A particularly powerful aid is the ability to call up context sentences for a particular word in order to examine its use in frequently occurring contexts.

Learner registration.

The program stores information concerning the words that each learner has studied up to a given moment, the help that has been requested and the answers to the exercises that have been given. These data can be scrutinized by the teacher.

Exercise development.

The system applications allow learners as well as teachers to enter random texts for the development of exercises. For example, a teacher could in

principle enter a text for later use by the learner to develop exercises. There are two entry-checks. The first one tests whether the word with the desired grammatical properties is present in the computer lexicon. The second check tests whether the parser is capable of parsing the input-sentence and whether the sentence is grammatical. Both checks are required for the creation of exercises.

Stories can also be entered after the parser has been cancelled. In such cases the only type of operation to be carried out is the connection of each word to its first dictionary meaning. The advantages of this option are that the input-process is faster and that texts of any degree of difficulty can be entered. This kind of text can only be used for receptive exercises.

The generative potential of the parser and the great amount of information in the lexical database make it possible for the program to adjust quickly to different teaching methods by allowing for the addition of method-specific texts. The menu-driven environment and the extensive context-sensitive help-options make the program very user-friendly for both teachers and learners. The unambiguous screen layout makes it relatively simple to enter

commands. It should be borne in mind, however, that the program's full potential can only be realized with suitable machines (i.e. computers with 80486 processors, 50 Mhz, 8 Mb RAM, hard disk, Windows 3.1 and a soundcard).

Multimedia

The Windows environment supports a wide variety of ways to present information. Text and graphics are the most common types used in applications.

With the addition of a soundcard sound can be easily recorded and played back. This way listening comprehension tests could be created.

A relatively new addition is the ability to display digital video in a window on the screen. The performance in frame-rate and the size of the playback-window is currently quite limited. Advances in image compression will soon overcome these problems. Use of digital video in language teaching opens other powerful ways of creating interesting exercises.

References

- 1 Kanselaar, G. & T. van Maanen, *Vreemdetalenonderwijs met computers in het voortgezet onderwijs: verkenning en inventarisatie van een onderzoeksgebied*. In: *Cluster V onderwijs-onderzoek: twee studies van het Instituut voor het onderzoek van het onderwijs*. PSOI-reeks nr.31, 1987, p. 3-154.
- 2 Sinclair, J. (Ed.), *Collins COBUILD English Language Dictionary*. Collins, London and Glasgow, 1987.
- 3 Kanselaar, G., M. Giezeman, J. Jaspers, L. Koster, J. van der Veen, H. Wichmann, J. Zuidema. *Computerondersteund vreemde talenonderwijs* (eindverslag van SVO-project 7108). Utrecht: ISOR, Universiteit te Utrecht, 1989.
- 4 Neuner, G., M. Krüger & U. Grever, *Übungstypologie zum kommunikativen Deutschunterricht*, Berlin: Langenscheidt KG, 1981.
- 5 Markham, P. Effects of contextual versus definitional computer-assisted vocabulary instruction on immediate and long-term vocabulary retention of advanced ESL students. *Educational Psychology*, 9(2), 121-126, 1989.
- 6 Gein, J. van de, *The sense of sentences. A study into the effects of grammar instruction upon junior writing*. Proefschrift, 17 mei 1991 te Utrecht.
- 7 Kanselaar, G., Kok, W.A.M, Jaspers, J.G.M. *It's English: Engels leren met computer en CDROM*. ISOR-rapportnummer 91.07, ISOR / Universiteit Utrecht, 1991.