

ELIZA the psychiatrist and self-access learning.

Watson Todd, R.

Selected Proceedings of the 4th CULI International Conference: ELT Collaboration: Towards Excellence in the New Millenium. Bangkok, Thailand, 1-3 December 1999.

The definitive version of this article was published as Watson Todd, R. (1999) ELIZA the psychiatrist and self-access learning Selected Proceedings of the 4th CULI International Conference: ELT Collaboration: Towards Excellence in the New Millenium. Bangkok, Thailand, 1-3 December 1999.

ELIZA the psychiatrist and self-access learning

Richard Watson Todd

King Mongkut's University of Technology Thonburi

Abstract

ELIZA is the prototypical artificial intelligence (AI) program which aims to mimic human communication. Other easily accessible AI programs include FRANK and CLAUDE. ELIZA and FRANK are psychiatrists which aim to help people solve their problems while CLAUDE is an eccentric conversationalist. These programs were made available for university students to use in a self-access centre. Students' reactions were very positive and the programs were used heavily. Problems, however, occurred since the programs repeated mistakes which the students made leading to possible reinforcement of errors. To solve these problems, a spellchecker and grammarchecker are being incorporated into an AI program to informally correct students' language while students converse with the computer. The aim is to produce an AI program with which students can communicate happily and naturally, and which will also help students improve their English.

Computer-assisted language learning (CALL) is here to stay. The drive and impetus behind CALL means that arguments for and against CALL (e.g. the series of papers by Higgins, 1988, 1989 and Hirvela, 1988, 1989) have

become irrelevant. Whether we like it or not, as teachers we have to accept that computers have a role in language learning/teaching and that this role is likely to increase over the coming years. Instead of arguing for and against CALL as an entity, then, we need to examine types of CALL and how CALL can be used for the greatest benefit for the learners.

In this paper, I will evaluate the use of one type of CALL, namely artificial intelligence (AI) chatterbot programs. The AI programs examined in this paper were not designed for use in language learning. By evaluating students' interaction with and reactions to these programs, I hope to be able to point out directions which the development of AI programs for language learning should take.

Types of CALL programs

Before focusing specifically on AI chatterbot programs, we need to look at how such programs fit into the overall picture of CALL and why they are worth investigating. Schreck and Schreck (1991) identify six main types of CALL program:

- interactive tutorials
- drills and practice
- simulations and games
- text-building applications
- multimedia applications
- AI applications

Schreck and Schreck's classification is perhaps too neat. Several programs exhibit the characteristics of more than one of these categories, often to the detriment of learning. At present, most CALL programs available are professionally produced with high production values which can create novelty effects (Clark, 1989) or a "wow factor" (Murray and Barnes, 1998). Although such high production values can have positive effects regarding initial acceptance of the program, if there is little substance beneath this highly produced surface, the long-term effects of CALL on student learning are likely to be negligible. This emphasis on technology-driven high production values over learning-driven content is a particular problem for multimedia programs which currently dominate CALL (Watts, 1997). For example, several multimedia programs (e.g. *Business English Activities* and *Planet English*) are little more than language drills dressed in expensive clothes, while others (e.g. *English Works 1*) simply offer a more convenient

method of listening to and recording dialogues than is provided by a tape recorder, and still others (e.g. *Expressions*) are based around the kind of stilted dialogue which was heavily criticised in coursebooks twenty years ago. Overall, regarding multimedia programs, “the list of missed opportunities [for designing effective language teaching programs] is depressingly long” (Eastment, 1996: 78).

To develop CALL programs, therefore, I would argue that, rather than focusing on the appearance and production values that technology-driven multimedia programs encourage, we need to step back and ask ourselves about what sorts of content which facilitate learning are feasible in CALL. Otto (1988) gives a useful list of the desirable characteristics of CALL programs:

- they should be based on pedagogical principles.
- they should interest and motivate learners.
- they should individualise instruction.
- they should encourage interaction.
- they should include ongoing evaluation so that learners can monitor their progress.

The multimedia programs criticised above focus on initial interest and motivation to the exclusion of the other characteristics of desirable CALL. Instead of investigating how to incorporate these other characteristics into multimedia programs, however, I would like to investigate alternative conceptions of CALL which may already include the desirable characteristics.

Artificial intelligence programs for CALL

Of Schreck and Schreck’s six types of CALL, the one least investigated is artificial intelligence. AI programs aim to mimic human intelligence and usually consist of source code to guide analysis of input and output, a knowledge base, and an interface (for technical details of well-known AI programs, see Grosz et al., 1986). One kind of AI program which has potential for language teaching is the chatterbot. Chatterbots are designed to simulate a conversational partner (for more details, see Laven, n.d. and Saygin, 1999), although the ‘conversation’ takes place in writing on the computer monitor rather than through speaking and listening.

Chatterbots, then, could provide potentially useful practice in certain conversational skills (though not all, given the written nature of the

‘conversation’) and in writing. In contrast to the practice element criticised in the multimedia programs above, practice with chatterbots is likely to be free and meaningful since the topic and direction of conversation are controlled by the learner, the interaction is unpredictable, and, for some chatterbots, the interaction may serve other learner purposes in addition to language learning (see van Lier, 1996 for a discussion of types of practice and their value). In this way, concerning Otto’s desirable characteristics, the kind of practice encouraged by chatterbots reflects pedagogical principles, they may motivate learners, the freedom of topic choice individualises instruction, and they may encourage interaction. These points need to be investigated to see if chatterbots have these characteristics in reality, and the lack of ongoing evaluation in chatterbots also needs consideration.

Given that chatterbots potentially have these desirable characteristics, the question of why they have not been used more in CALL arises. I will posit two reasons here. Firstly, most language teachers are not aware of the existence or availability of AI programs and so have overlooked them. Secondly, some CALL specialists (e.g. Nyns, 1989) have argued that designing AI techniques for language learning is impossible because of the problems of constructing suitable knowledge bases.

The second of these two reasons presents a serious obstacle to using AI programs for CALL, but I believe that there are two ways in which it can be circumvented. Firstly, it is possible to construct a knowledge base if the domain is kept specific. For example, a chatterbot called *ARTHUR* makes an interesting conversational partner so long as the topic of conversation is films; on other topics, he has remarkably little to say. Secondly, some chatterbots, such as *ELIZA*, *FRANK* and *CLAUDE* investigated in this paper, give the illusion of being AI programs rather than being genuinely ‘intelligent’ (Devlin, 1997). From the user’s perspective, the fact that the ‘intelligence’ is illusory rather than genuine is irrelevant, since the programs perform as if they are genuine AI programs.

This paper investigates the potential of ‘illusory’ chatterbots. These, rather than domain-specific chatterbots, were chosen primarily because of their availability, but also because they do not restrict possible topics of conversation.

The situation

King Mongkut's University of Technology Thonburi has a large well-developed self-access centre with 18 sections. One of these is CALL. Because of budget constraints, there are only seven non-networked computers available in this section. Nevertheless, the CALL section is heavily used. Most programs in the section are pre-multimedia controlled practice programs such as *SEQUITUR* and *DOUBLE-UP* and TOEFL practice programs. Three chatterbots (*ELIZA*, *FRANK* and *CLAUDE*) were added to the menu of programs available.

To investigate the effectiveness of the programs, students were observed using the programs, records of the 'conversations' were made, and informal interviews were conducted. The students observed fall into two groups: those who used the programs on their own initiative, and those who were asked to use the program by the researcher. It was expected that the first group might be able to highlight the advantages of the programs and give suggestions for improvements, whereas the second group could explain why they did not like to use the programs.

Sample interactions between learners and chatterbots

ELIZA

The first chatterbot we will examine is *ELIZA*, which was one of the first AI programs. *ELIZA* is intended to mimic a psychiatrist and so help solve personal problems. To investigate *ELIZA*'s effectiveness, we can look at two conversations between a third-year undergraduate and *ELIZA*.

'Conversation' 1

ELIZA: What's your problem?
 Student: I cannot study well.
ELIZA: Are you saying no just to be negative?
 Student: Yes.
ELIZA: Are you sure?
 Student: Yes.

'Conversation' 2

ELIZA: What's your problem?
 Student: I cannot study well.
ELIZA: You are being a bit negative.
 Student: I cannot get a good grade.
ELIZA: Why not?

Student: I cannot understand teacher.
ELIZA: Are you sure?
 Student: My teacher very bad.

In these extracts, it is clear that *ELIZA* is showing her age. With a limited set of responses and no parsing ability, ‘conversations’ with *ELIZA* quickly become circular. Students also felt that extended meaningful conversations with *ELIZA* were difficult and quickly became boring. For our purposes of identifying directions for the development of AI programs for language learning therefore, *ELIZA* is not an appropriate model to start from.

CLAUDE

CLAUDE was designed to be used in early secondary education in the US, and so promises to be useful. The following extracts, however, show *CLAUDE*’s limitations regarding language learning.

‘Conversation’ 3

CLAUDE: My name is Claude. What should I call you?
 Student: X.
CLAUDE: Hi, X. I’m pleased to meet you.
 Student: Pleased to meet you.
CLAUDE: Great Scott! Really?
 Student: No.
CLAUDE: No. I suppose not. Things could be worse, you know. Of course, a sheriff’s deputy can’t ever be niftier than a janitor is nifty.

In this extract, things start well, but the unfamiliarity of the English *CLAUDE* uses and the unpredictability of his responses mean the student gave up immediately. A student with a higher level of English proficiency, however, persevered:

‘Conversation’ 4

CLAUDE: My name is Claude. What should I call you?
 Student: X.
CLAUDE: Hi, X. I’m pleased to meet you.
 Student: Yes, it’s very nice to meet you too.
CLAUDE: Whether something is nice or not is purely a matter of opinion. But it doesn’t make sense to dwell on it. I like a drunkard who never shaves gerbils. However, gerbils are awfully bizarre.

Student: You are very crazy.
CLAUDE: I am, huh? I bet you can't prove it.
 Student: What is your name?
CLAUDE: You ask the darnedest stuff. I admire that in you. That reminds me of something. I knew a living room from Omaha that forgot to be fashionable. It was dismal.
 Student: This is program, there are problems.
CLAUDE: I'm not so sure it is. And it wouldn't take a genius to figure that out. To see trash, a private detective must be lucky.
 Student: Fuck you.
CLAUDE: Give me a break and put a lid on that obscenity crap.
 Student: I'm sorry. I don't do it again.
CLAUDE: You are! Well, imagine that.
 Student: I will try to do that.
CLAUDE: OK. If you will, you will. But that can be our little secret. Right. I think scientists are awfully gloomy, don't you?
 Student: What do you want to do tomorrow?
CLAUDE: I don't think that's any of your business! ...

This student managed a fairly long 'conversation' with *CLAUDE*, but even she complained that she did not really understand the 'conversation' and did not want to try again. With no restrictions on the language he can use and with his overall abstruseness, like *ELIZA*, *CLAUDE* appears to be an unpromising starting point for the development of AI programs for language learning.

FRANK

Let us then turn to the third program. *FRANK* was designed as an improved version of *ELIZA* and so is also designed to mimic a psychiatrist. Again, we can examine interactions between the program and students.

'Conversation' 5

FRANK: What is your first name?
 Student: X.
FRANK: What is your problem, X?
 Student: I cannot study very well.
FRANK: X, have you ever tried to study well?
 Student: Yes, I tried to study in classroom.
FRANK: Go on.
 Student: My teacher say something, I cannot understand.

- FRANK:* What would happen if you did understand?
 Student: I tried to tell my teacher say slowly.
FRANK: Your teacher say slowly? What concerns you most about that?
 Student: I think it can help me understand subject?
FRANK: Do you really think it can help you understand subject?
 Student: I don't know, sometimes it can help me understand. Sometimes I don't know knowledge. I should read before.

In this extract, *FRANK* served his purpose well and helped the student to solve her own problem. The student enjoyed conversing with *FRANK* and after this extract went on to have several further 'conversations'. Nevertheless, there were problems regarding the language *FRANK* used. For example, while conversing with *FRANK* the student asked me to explain the meaning of "What concerns you most about that?" The problem of *FRANK*'s language use is, however, minor and can be solved with some quick tinkering to the program.

A more serious problem arose with a postgraduate student with lower proficiency in English as the following extract shows.

'Conversation' 6

- FRANK:* What is your first name?
 Student: X.
FRANK: What's your problem, X?
 Student: I want to know about my mine.
FRANK: What would it mean to you if you got to know about your yours?
 Student: I want to understand about my mine?
FRANK: Your yours? What concerns you most about that?
 Student: Some time, I don't understand in my done or my desige.
FRANK: Your desige? What concerns you most about that?
 Student: I think it difficult to explane for me.
FRANK: Do you really think it difficult to explane for you.
 Student: Yes.
FRANK: Tell me why.
 Student: Because I don't know how to explane.
FRANK: What is the real reason, X?
 Student: What are the reason?
FRANK: Why do you ask, X?
 Student: I don't understand in your mean.

Despite the seeming incomprehensibility of some parts of this extract, the student stated that he enjoyed talking to *FRANK*. Nevertheless, with students of lower proficiency, *FRANK* does pose some serious problems.

From a language learning perspective, the most worrying aspect of ‘conversation’ 6 is the lack of feedback given on the learner’s performance. In terms of Otto’s (1988) desirable characteristics of CALL programs, *FRANK* interests the student, individualises instruction and encourages interaction, but *FRANK* does not enable the student to monitor his own progress. In fact, *FRANK* may actually be detrimental to the student’s learning since the program may reinforce errors by repeating them (e.g. “explain”). Learners need feedback from a ‘superior’ system which can help them identify and correct errors, rather than having their errors reinforced and getting no feedback (Prabhu, 1987). While *FRANK* may be interesting and interactive, then, he is far from perfect, but may provide a foundation we can work from.

Evaluation of the AI chatterbot programs

From the ‘conversations’ above, we can see that *ELIZA* and *CLAUDE* are probably inappropriate models to work from in our search for AI programs suitable for language learning. *FRANK*, however, shows more promise, though there are still serious problems regarding his efficacy for language learning.

Overall, the programs were used less than was hoped for, and the main users were students with relatively high proficiencies in English. When these students were interviewed, they were positive about the programs especially *FRANK*, and highlighted the psychiatrist aspects of the programs. For example,

“I really like these programs. I don’t like Internet chat rooms because if you want to talk about your problems, no one wants to listen and you can’t speak strongly. These programs listen to anything you say.”

For students of high proficiency then, *FRANK* provides a useful means of meaningful practice and also someone to talk to about problems.

Less proficient students, however, used the programs less frequently. They stated that they found it difficult to express what they wanted to say (which may actually be a rationale for using the programs) and so became bored.

Even in this group of students, some, such as the student in ‘conversation’ 6, said they found the programs interesting, wanted to use them, but needed more language support.

In their present state, therefore, *ELIZA*, *CLAUDE* and *FRANK* are more appropriate for learners with high proficiencies in English, and so may be suitable for enrichment work (Guskey, 1989). To make the programs more beneficial for learners of a wider range of proficiency levels, language support and feedback needs to be added to the programs.

Directions for AI programs for language learning

In looking at directions for developing AI programs for language learning, I will focus on how the programs can give feedback on the learner’s performance, as I believe this feedback may also provide language support. The most widely used methods of giving feedback on unpredictable language performance are spell checkers and grammar checkers, such as those used with *Microsoft Word*. These can provide a model for us in incorporating feedback into AI programs, but as usually used, there are three problems with spell checkers and grammar checkers, two trivial but one serious, which need to be overcome.

Firstly, any corrective feedback should not interrupt the flow of the ‘conversation’ between the AI program and the learner. This can be solved by adding extra responses to the AI program which relate to feedback. For the student’s error “explane” in ‘conversation’ 6, the program may give feedback as follows. First, the program using string recognition would identify “explane” as a word not given in its dictionary. It would then identify “explain” as the most plausible correct spelling of the word. Next, instead of highlighting “explane” and giving a choice of alternatives as happens in most word processing programs, the program’s response to the student’s input which includes “explane” would be something like “I’m sorry. I’m not sure what you mean by “explane”. Do you mean “explain”?” If in the next line the student types “Yes”, the program would replace “explane” with “explain” in its memory and respond to the original line. In this way, the feedback is given as an insertion sequence (see Coulthard, 1977) in the ‘conversation’, a common feature of conversations which does not interrupt the overall flow.

The second problem is that, for some lower proficiency learners, there may be a need for constant feedback on spelling and grammar errors. Corrective

feedback given on every sentence the learner types would probably be demotivating, and thus the interest aspect of AI programs could be lost. Commands within the program to only give corrective feedback on, say, every third line of learner input or a random generator enabling corrective feedback on a third of all errors could solve the problem easily.

The third problem is more serious. At present, most checkers, both for grammar and perhaps surprisingly for spelling, are designed for native speakers of English rather than non-native speakers. As anyone who has tried grammar-checking mistake-riddled students work in *Microsoft Word* will know, grammar checkers for native speakers do not pick up most mistakes that non-native speakers make. Similarly, although spell checkers will pick up mistakes, the alternatives suggested are based on the closest letter combinations rather than on orthographic variations of homophones which are frequently the source of spelling errors of non-native speakers. To create a spell checker and a grammar checker specifically designed for non-native speakers to be linked with an AI program is clearly a massive task. A basic version, however, focusing on some common errors such as miscollocations could be achievable purely through string recognition. More advanced versions would require parsers and would probably need to be based on a functional grammar. Even an advanced version would probably overlook many errors. Despite the problems, creating an advanced grammar check specifically aimed at the problems made by Thai learners is under consideration as a research project by the Natural Language Processing Laboratory at King Mongkut's University of Technology Thonburi, and I hope that within a couple of years, we will be able to report some progress in this area.

Conclusion

With the unstoppable growth in importance of CALL, effective kinds of CALL are needed. The presently dominant multimedia approach may be one such way though many such programs lack depth, but alternatives also need to be sought. AI programs are one possible alternative which could provide meaningful practice for learners. However, existing AI programs lack the feedback element necessary for learners to monitor their own progress, which means that at present only learners with high proficiency may benefit from these programs. Adding a spell checker and a grammar checker designed for non-native speakers, although a tall order, could create AI programs which lead to effective computer-assisted language learning.

References

Texts

- Clark, R. E. (1989) Reconsidering research on learning from media. In Anderson, L. W. (ed.) *The Effective Teacher: Study Guide and Readings*. New York: McGraw Hill. pp. 232-242.
- Coulthard, M. (1977) *An Introduction to Discourse Analysis*. London: Longman.
- Devlin, K. (1997) *Goodbye, Descartes: The End of Logic and the Search for a New Cosmology of the Mind*. New York: Wiley.
- Eastment, D. (1996) Survey review: CD-ROM materials in English language teaching. *ELT Journal* vol. 50 no. 1 pp. 69-79.
- Grosz, B. J., Jones, K. S. and Webber, B. L. (eds.) (1986) *Readings in Natural Language Processing*. Los Altos, CA: Morgan Kaufmann.
- Guskey, T. R. (1989) Feedback, correctives, and enrichment. In Anderson, L. W. (ed.) *The Effective Teacher: Study Guide and Readings*. New York: McGraw Hill. pp. 353-363.
- Higgins, J. (1988) Response to Hirvela's article "Marshall McLuhan and the case against CAI". *System* vol. 16 no. 3 pp. 313-318.
- Higgins, J. (1989) A reply to Alan Hirvela's reply. *System* vol. 17 no. 1 pp. 67-70.
- Hirvela, A. (1988) Marshall McLuhan and the case against CAI. *System* vol. 16 no. 3 pp. 299-312.
- Hirvela, A. (1989) The case against CAI: a reply to John Higgins. *System* vol. 17 no. 1 pp. 61-66.
- Laven, S. (n.d.) The Simon Laven page. <http://www.toptown.com/hp/sjlaven/>.
- Murray, L. and Barnes, A. (1998) Beyond the "wow" factor - evaluating multimedia language learning software from a pedagogical viewpoint. *System* vol. 26 no. 2 pp. 249-260.
- Otto, F. (1988) Using the computer. In Gleason, J. B. (ed.) *You Can Take It With You: Helping Students Maintain Foreign language Skills Beyond the Classroom*. Englewood Cliffs, NJ: Prentice Hall. pp. 71-92.
- Prabhu, N. S. (1987) *Second Language Pedagogy*. Oxford: Oxford University Press.
- Saygin, A. P. (1999) The Turing Test page. <http://www.cs.bilkent.edu.tr/~psaygin/ttest.html>.
- Schreck, R. and Schreck, J. (1991) Computer-assisted language learning. In Celce-Murcia, M. (ed.) *Teaching English as a Second or Foreign Language*, second edition. Boston, MA: Heinle & Heinle. pp. 472-485.

Van Lier, L. (1996) *Interaction in the Language Curriculum: Awareness, Autonomy & Authenticity*. London: Longman.

Watts, N. (1997) A learner-based design model for interactive multimedia language learning packages. *System* vol. 25 no. 1 pp. 1-8.

Programs

References for computer programs are problematic, both in format and in finding dates, publishers etc. I hope that the details below, although missing some information, may help.

ARTHUR available at <http://www.kingston.ac.uk/users/k967325/> (source code and knowledge base can be downloaded, but a good working knowledge of Turbo C++ is required to make the program run.)

Business English Activities CD-ROM published by Vektor/BBC.

CLAUDE written by McLaughlin, B. in 1991, downloadable from Laven (n.d.).

DOUBLE-UP diskette written by Higgins, J. and Higgins, M. in 1992.

ELIZA written by Weizenbaum, J. in 1964, downloadable from Laven (n.d.).

English Works 1 CD-ROM published by Longman.

Expressions CD-ROM published by Vektor.

FRANK unknown source, downloadable from Laven (n.d.).

Planet English CD-ROM produced by TASEAP, 1998.

SEQUITUR diskette written by Higgins, J. and Higgins, M. in 1992.