FACE TO FACE: AN INTERACTIVE CHINESE LANGUAGE LEARNING MODEL ON THE INTERNET

JIANNA J. ZHANG
Department of Computer Science, Western Washington University, Bellingham, WA, USA

DONGYAN BLACHFORD
Department of International Languages, University of Regina, Regina, Sask., Canada

JAMES HEARNE
Department of Computer Science, Western Washington University, Bellingham, WA, USA

JANET XING
Department of Modern and Classical Languages, Western Washington University, Bellingham, WA, USA

We describe a theoretical and technological framework for designing and developing an interactive Chinese language learning model based on current language learning theories [Dobrovolsky 1996], existing technology and established techniques in Artificial Intelligence and Machine Learning. This model is intended to provide students with a near reality learning environment which can help them effectively practice the four major language skills: listening, speaking, reading, and writing. The design of our system has seven components: (1) the natural language user interface; (2) Chinese character and sentence recognition; (3) Chinese speech recognition; (4) student face recognition; (5) the dynamic knowledge base; (6) the virtual teacher; (7) the self-improvement element. We believe that through the collaboration of expertise in computer technology and in language instruction, this intelligent and flexible software can be developed to respond the real challenges in computer aided natural language learning.

Keywords: Machine Learning, Chinese Language Learning Model, AI, Internet

1. INTRODUCTION AND BACKGROUND

The ideal way to learn a language is to be immersed into the target language environment and without an environment to practice and use what they have acquired in the classroom, students competence soon erodes [Kaplan et al.1998]. Although more and more students have managed to study in China and maintained and/or improved their Chinese proficiency, such study is not feasible to support the anticipated demand for people who can communicate with the Chinese world. What we have identified is a need to fill in the gap with a tool which can provide students with a simulated environment where they can effectively develop major communicative skills: speaking, listening, reading, writing, and ways to keep conversation going with appropriate cultural manners. An intelligent computer program simulating this learning environment will serve as a bridge or short cut from classroom to reality [Moore1996, Lehnen2000].

2. CURRENT COMPUTER TECHNOLOGY IN LANGUAGE TEACHING AND LEARNING

We surveyed the literature on CALL (computer assisted language learning) to understand what tools are available for language related activities, what they have to offer to improve language learning, their major advantages and disadvantages, and most importantly what techniques we can borrow from such to solve the problem as we have characterized. In the

© 2003 Pacific Association for Computational Linguistics
past decade or so, CALL programs and on-line materials have flooded the field of language teaching and learning and bombarded the learners as well as educators with a variety of activities.

2.1. Pronunciation

There are all sorts of online programs and software on the market specially designed to help with pronunciation such as English Pro Web Edition for English learners, HyperChinese [Fu1996], and Pronunciation Modules [Smitheram and Ma1996], that is designed to help students with the learning and practicing of various Mandarin sounds (Pinyin) and tones. There are even commercial programs, which demonstrate the way of pronouncing by showing mouth, teeth, and tongue positions through animated visual displays. Many programs also offer immediate feedback, mini quizzes and game like drills to boost interest in pronunciation practice.

2.2. Grammar Exercises and Vocabulary Building Games

A large number of on-line and CALL programs are designed to teach and practice grammar and vocabulary. Electronic dictionaries are also available in various languages. For example Hall [Hall1998] talked about how a varied application of CALL and IT (Information Technology) can be used to teach grammar more effectively by a combination of explicit, implicit, and exploratory approaches.

2.3. Character Writing

This is unique to Japanese and Mandarin learning. The development of computer animation has been of great assistance to students to practice and learn to write Chinese characters on line. The animated program, for example WenLin [WenLin Institute2002], is capable of demonstrating, stroke by stroke, each character chosen by students.

2.4. Authentic Materials

One of the greatest advantages that the computer has brought to foreign language learners is to provide exposure to the target language and culture in various authentic forms: on-line newspapers, magazines, songs, novels, videos, movies, photographs etc. The information is at the fingertips for research and increasing knowledge.

2.5. E-mail and Chat Rooms

The e-mail and chat rooms are so far the most interactive tools provided by computer technology. They offer a real opportunity for communication in written form between a language learner and a native language speaker. They partially fill in the gaps we have identified by providing an authentic language environment in which several aspects of language learning are addressed but there are several drawbacks: lack of face-to-face communication and the lack of practice in the use of oral skills.

2.6. Microworlds

This tradition in CALL is the closest to the approach we follow. In this class of programs there is an attempt to integrate grammar, vocabulary and sometimes speech production by enabling students to interact with a simulated world [Kaplan et al.1998]. But because
they deploy few techniques from machine learning, artificial intelligence and computational linguistics, they are quite limited in their interactive potential, often being restricted to a limited repertoire of sentences [Kaplan et al.1998, Graesser et al. 2001] We have found that many of the existing programs are, at best, capable of demonstrating and reinforcing the student’s knowledge of the target language and culture. Most of them still play a passive role, as they are mainly oriented and limited to providing information, amplifying explanations, and mechanical drills. Technically, what is missing is the ability to react to unpredictable outcomes as these would occur in real life communication.

What we envision is a model that provides an environment strongly simulating reality that can arouse learners’ creative ability as well as engage their attention. That is, the model should provide advice and guidance and, most of all, lead the students to the realization that there is a way to advance communicative competence without the investment of foreign travel or other means of more plastic linguistic interaction. Hence, we are designing and developing an interactive language model, making use of established techniques in Artificial Intelligence, Machine Learning and Natural Language Processing. The approach we follow is built upon the insights and techniques of existing programs but is significantly more ambitious in pressing into service recently codified techniques in supporting sub-disciplines of computer science.

3. THEORETICAL AND TECHNOLOGICAL BASIS

Through decades of research and experimentation, the current field of second and foreign language acquisition focuses the research on how a second language learner acquires the target language and what knowledge and skills contribute to the ability to use a language effectively. So far, most language educators believe that the focus should be on the development of communicative competence on the part of the learner. There are many books and journals

![Diagram](image_url)
devoted to this topic. Other than merely teaching the forms of a language as has been done for centuries, communicative competence includes a wide array of related elements contributing to successful language learning. [Dobrovolsky 1996] model of communicative competence, as depicted in Figure 1, provides a good idea of the major elements included in this concept. The belief is that while knowledge of a language’s grammar such as phonetics, phonology, morphology, syntax, and semantics allows learners to distinguish between grammatical and ungrammatical sentences, successful communication requires many more skills and knowledge as indicated in Figure 1. All of these aspects of language competence need to be considered in learning a language for the purpose of effective communication, including the interaction between culture and usage. However, as discussed earlier, many of the existing learning programs are inadequate in addressing the areas of strategic competence and pragmatic competence. These are no doubt very difficult areas to tackle.

Language is actual use is far richer than textbooks would suggest and the cultural elements attached to it make it unique to each situation. For example, there are number of ways of expressing a greeting in Mandarin: "ni hao!", "ni hao ma?", "zemu yang?", "zuixiu hao ma?", "jinlai hao ma?", "chi fan le ma?", "hao jiu mei jian le!", "ni qu na le?", "zao wa!", "shen ti hao ma!", or simply "hei!". The native speakers would never give a thought before answering any form of greeting. However, it would be extremely difficult for students to arrive at this level for there have been no effective tools available to assist them in developing such competence.

Classroom instruction and texts are usually confined to a few forms because spelling-out the cultural circumstances in which each form is appropriate would require digressive elaboration. Even a short exchange of greetings involves an array of knowledge and skills, such as cultural tradition and communication strategies in addition to grammar, vocabulary, pronunciation. But the forms can be taught in a simulated setting quite easily, as they would in the first few weeks of an immersion experience. It is the nature of the language and language use. Therefore, we intend to venture into the task of providing the learners with such a model which is designed intelligent enough so that it can help the learners to explore, be advised, prompted, given alternatives, feedback, and ultimately lead them through various social situations. What is needed is a computer program with the ability to act intelligently to guess the users intentions, to assess their abilities, to analyze and pinpoint their weakness and to provide help accordingly. In short, what we need is a tool or a boat to allow students to try but not sink and eventually build up enough courage, knowledge, and skill to successfully reach their destination.

4. DESIGN OF THE INTERACTIVE LANGUAGE LEARNING MODEL

The interactive language learning model we are developing contains a selection of virtual environments (VI), such as “airport”, “bookstore”, or “hospital”, in which students can practice a particular language (current application is Mandarin Chinese) that they have learned in class rooms. We want students to feel as if face to face with native language speakers and at the same time not to feel being watched or examined by a third party. In such worry-free environments, students would feel more relaxed and therefore be able to concentrate on the learning tasks.
To achieve such effects, we design a natural language (Chinese) oriented user interface shown in Figure 2. The communications between students and the virtual teachers are realized in Chinese writing using the mouse or a pen pad, and/or through conversations by computer speakers and microphones. A set of intelligent machine learning programs, as the center brain of F2F, lay behind this interactive user interface. These machine learning elements are: Chinese sentence recognition (CSR), Chinese speech recognition and generation (CRG), student face recognition (SFR), virtual teachers (VT), dynamic knowledge base (DKB), and performance evaluation (PE). The detailed design for them is stated as the
4.1. The Natural Language User Interface

The natural language user interface of F2F are dynamic Web pages containing virtual environments classified by learning tasks (Figure 3 is a screen shot which shows a selection of virtual environments), virtual teachers chosen by the students, and a dialog box to display the conversation in Chinese sentences. An initial input, either entered by a student or selected by a virtual teacher, will start the learning process. The virtual teachers, supervised machine learning programs, will play different roles in different environments.

For example, in the airport environment, students can choose to talk to a front desk clerk, another passenger, or airport attendants. If the student makes a mistake in their conversation, the virtual teacher, who plays as one of the above selected characters, will try to correct the student after they guess what the student meant. The commonly made mistakes and the learning progress of individual students will be recorded and analyzed. Suggestions, and advises will be given according to the profile of individual students. A student face recognition component will enable the virtual teachers to “remember” their students so that they can provide better supervision on individual students.

4.2. Input/Output Enhancements

F2F will take advantage of several input/output technologies that have matured in recent years. But unlike previous computer aided language learning tools F2F will be trained to recognize Chinese sentences instead of single characters. Although much research effort has been done towards Chinese character recognition [CHO02, CHA01, TON00, FU99, CHE97, HAN97, ROM96], very few has been attempted recognize sentences. We want to investigate current research and adapt the available and suitable technologies to our implementation. For example, Chong et. al. have developed a simple but effective Two-stage coding scheme which has achieved close to 99%[TON00]. This scheme combines the projection and regional decomposition coding methods to achieve extracting global and local features of the input Chinese characters. We want to research on this method further and use is as the starting point of F2F’s Chinese sentence recognition scheme.

There is also available software such as SharpEye Chinese Optical Character Recognition System (SharpEye Cor) and Chinese Pen by TwinBridge. SharpEye can recognize traditional Chinese characters up to three fonts of 5,543 characters. It can recognize up to 1000 characters per minute, and the accuracy rate can be as high as 98%combined with the character recognition software SharpEye may contribute greatly to F2F’s Chinese sentences recognition, the techniques used by SharpEye and Chinese Pen are not discussed anywhere. We must test and analyze it carefully before a further conclusion is drawn.

Speech processing is a well developed research area. There exist many world class commercial products, free ware, and research projects. For example, Sphinx, a real-time, large vocabulary, speaker independent speech recognition system, is developed at Carnegie Melon University [LEN02]; Chinese ViaVoice, an extension and enhancement of IBM’s speech recognition technology, offers a speech interface for Mandarin Chinese [IBM02]; ScanSoft’s RealSpeak, the most widely used text-to-speech solution in the world, enables interactive voice response applications [SCA02]; VBVoice 5.0 Text-to-Speech synthesis, developed by Pronexus, enables a computer to translate text into audio output [PRO02]; TwinBridge Smart Ear Chinese Speech Recognition System allows the users to interact with machines in the same way of interacting with people [AS100]; and TTS developed by AT & T is a creation of audible speech from computer readable text [ATT02]. We want to test some of
the available software and select one or combine some of them into new software for F2F.

To enhance habit-ability, we wish also to implement a student face recognition component such that the virtual teacher will be able to recognize students and provide specific supervision for individual students. Many successful commercial products on the market will help us to realize this user friendly feature. However, modifications and improvements must be made within reasonable scope. Our investigation will start with three commercial face recognition software:

1. FaceIt, an award-winning facial recognition software engine that allows computers to rapidly and accurately detect and recognize faces [Incorporated2002].
2. Mandrake, a neural network that can recognize faces by comparing templates [TSSI2002].
3. FaceTOOLS the complete software developer’s kit to meet all Windows based face recognition development needs [Inc.2002].

All three leading research projects on face recognition have generated successful results and have practical applications. Indeed, FaceIt and FaceTOOLS have discovered that faces also have prints similar to fingers. The faceprint is not sensitive to lighting, skin tone, eyeglasses, and facial expressions. In spite of this significant discovery, both FaceIt and FaceTOOLS require static face databases, and none uses machine learning techniques. Although FaceIt uses a neural network, it requires retraining the entire network if a new face needs to be recognized. Based on our study and discussion on previous face recognition research, we propose a new method that automatically generates a hierarchical neural network called NN-Tree (Neural Network Decision Tree) to learn recognizing faces utilizing the faceprint technology discovered in [Incorporated2002] and [Inc.2002].

4.3. Natural Language Processing: Background Assumptions.

In spite of our goal of building a teaching environment maximizing flexibility and habit-ability, we are acutely aware of the limitations on current technology. The strategy we will follow will be an extension of one that has strong precedent in the literature [Fu1996, Wang et al.1999, Wenlin Institute2002, Xing2000, Xing1998]. Syntax will be centered around

![Figure 4](image-url)  
**Figure 4.** The Overall Design of the Dynamic Knowledge Base
lessons in a specific text. Secondly, we presuppose that a precise formal representation of stereotyped domains will permit a precise mapping from the semantics of natural language and the knowledge representation of those domains. Such a formal representation, including a representation of syntax that includes in its grammar a representation of student errors—misplaces qualifiers, violation of transitivity conditions and many more—allows for a rule-governed response that (1) properly corrects and (2) tests the same condition at a later stage of the dialog.

4.4. The Dynamic Knowledge Base

The F2F knowledge database will be organized so that the primary knowledge stored in the system can be updated and retrieved dynamically. Figure 5 shows the general architecture of the dynamic knowledge base in the F2F language learning model. Each “KB” in Figure 5 is a set of sub-knowledge bases corresponding to different virtual environments. Each virtual environment, such as Airport, Bookstore, and Supermarket, is associated with the index file that contains keys to different KBs.

4.5. The Virtual Teacher

Since the purpose of F2F to support highly active interactions with the students, the virtual teacher’s help should be specified for each student according to the his progress. It can give advice and correct errors according to the history of a student’s learning behavior. The model maintains a history database and a neural network to recognize and remember the students whom the teacher had met previously. This maintenance is done by a knowledge accumulator shown in Figure 6.

Obviously the virtual teacher is an intelligent machine learning program which learns in order to help students learn. We will explore two approaches, of which space limitations precluded extended discussion, will be explored: The IVSA approach, [Zhang2002,

5. CONCLUSION

We find that although current computer technology cannot replace a classroom teacher, it can certainly enhance language learning. We agree that the ultimate goal in using technology for teaching and learning should be to enhance what we already do well, and allow us to explore approaches we have never been able to consider. [JOH89]. We believe that implementing the F2F language learning model is a useful effort to apply state-of-the-art computer technology to difficult linguistic problems. We also believe that such a model is not limited to Mandarin learning. It has greater capacities and implications for the teaching of other foreign languages. It can also be adapted to be used by students of all ages and for all instructional levels.

Currently, we are still at a very initial stage. There are many tasks, such as how to represent the internal knowledge base intelligently, or how to make the model "aware" of different types of questions which actually mean the same thing, that have not been solved completely yet. As we have discussed earlier in the previous sections, we will start from a simple model that will fulfill the basic requirements of our design first, and then continue to reach our future goals one after another. We believe our model holds the promise of providing near reality learning environments for language students. We believe that the modular approach to be followed in this project, which separates formal representation of knowledge, semantics, pragmatics and syntax, entail that many components can be deployed the languages other than Mandarin without modification. Since we intend to publish our results in a vigorous way to the communities outside Computer Science, we hope that have a broad general influence on second language learning.

REFERENCES


