Construction and Application of Learner Corpus for Chinese Language E-Learning Systems

中国語 e ラーニングシステムのための 学習者コーパスの構築と応用

February 2011

早稲田大学国際情報通信研究科
国際情報通信学専攻
情報ネットワークシステム研究 II

Song LIU / 劉　松

PREFACE

The recent advancements in information and communication technology (ICT) have changed our learning and teaching methods. With the availability of technologies such as the Internet, multimedia, database, and artificial intelligence, educators and learners need not remain in one place and can use their time effectively at their own discretion. ICT-based language learning systems can also be realized with the development of these technologies. At Waseda University, a new language education model named Tutorial Chinese (TC) has been introduced since 2002. TC classes are small, usually consisting of one native speaking tutor and four students. One of the educational goals of this course is to provide more opportunities to Japanese students to learn Chinese with native speakers. The tutors who are located in Tokyo, Beijing, and Taipei employ a conversational style in conducting lessons. In order to support this education model, it was necessary to create an e-Learning system named Tutorial Chinese Platform (TCP), which is the focus of this research.

In the TC model, many components of language such as new vocabulary and sentences taught in the classroom go far beyond the contents of a static textbook and are valuable to be utilized in pedagogy. This kind of information is called a learner corpus, which is the collection of spoken or written texts produced by learners in the process of learning a language. Learner corpus is important in the field of second language education because the specific learning situations of learners can be closely analyzed and appropriate advice can be offered for future pedagogic design, based on this kind of information. In this research, correspond to three important language skills—vocabulary ability, writing ability, and tone discrimination ability, that need be practiced outside of TC classrooms, three kinds of corpora—vocabulary corpus, composition/correction corpus, and error corpus of tone discrimination are dynamically generated every week from pedagogical activities of TC and need to be utilized by the students as soon as possible. Therefore, the objective of this research is to integrate both construction and application of learner corpus to daily pedagogical activities effectively.

In order to achieve this objective, a mechanism that combines dynamic construction of learner corpus and its applications has been implemented in the TCP. Appropriate corpora have been utilized to create learning exercises for students. With the support of this system, teachers, tutors, and students are involved in the process of generating, gathering, analyzing, and utilizing learner corpora. This process works as a dynamic cycle with valuable learner corpora being stored and applied. This dissertation consists of six chapters and is organized as follows.

Chapter 1 introduces related research fields, including e-Learning, learner corpus, and the current situation of Chinese language education at Waseda University. The importance and necessity of the TCP and the objectives of this research are presented.

Chapter 2 describes the research proposal — the working mechanism is needed to integrate the construction of learner corpora and their applications to daily pedagogical activities, and the design for Waseda Tutorial Chinese Corpus (WTCC) are described. Corresponding to three important language skills that need be practiced through e-Learning system, the following three chapters described three applications and corresponding corpus respectively.

Chapter 3 introduces the Computer Assisted Language Learning (CALL) drill sub-system for improving of vocabulary ability, which combines the lesson report module of the TCP with WTCC to collect new vocabulary used in the classroom into vocabulary corpus. Meanwhile, the mechanism of analyzing the corpus and semi-automatically creating learning exercises has also been implemented. This system has features such as (a) web-based user interface for students and teachers to facilitate system access, (b) integration with WTCC corpus to make drills more effective, (c) layer architecture to make authoring flexible and reusable, and (d) a mobile phone-based ubiquitous environment to make the drills accessible to users anytime and anywhere. The mechanism implemented in the CALL drill sub-system has been employed in TC since 2004 and achieved satisfactory results.

Chapter 4 describes the composition/correction sub-system with a corpus retrieval function. The concept of this sub-system is to take advantage of the Internet, implement online composition/correction functions similar to that in a paper-based work environment, and simultaneously store the compositions and correction information into the composition/correction corpus through the correction process. This sub-system is effectively used not only by students in writing composition but also by teachers while

correcting and providing comments on the compositions. Moreover, the compositions together with the correction information have been tagged and constructed automatically into the learner corpus, and the retrieval function has also been developed, which makes it convenient for researchers to analyze this kind of language resource. The result of the usability evaluation shows that the composition/correction sub-system has dramatically increased teachers' work efficiency.

Chapter 5 cites a typical example named the Computer Assisted Instruction (CAI) sub-system for self-teaching of discriminating Chinese four tones. Since Mandarin Chinese is a tonal language, it is a big barrier for Japanese learners in the understanding of the features of the tones when studying Chinese. In order to give them a more direct understanding of the Chinese tones, this sub-system presents a visualized pitch pattern of the tones and gives effective guidance derived from the analysis of tone discrimination error corpus. The system is mainly designed for elementary level students. Precise acoustic data of the Chinese tones have been utilized for designing the system. Characteristics of the errors found in the examination have been analyzed, and these characteristics have been carefully considered while constructing the system. Examination of the improvement in the students' scores after practicing with the CAI system has clearly confirmed the effectiveness of the system. Moreover, the result of the analysis of the students' answers has provided valuable insights for teachers to make effective education plans.

Finally, chapter 6 concludes the dissertation and proposes the possibility of future development.

In this study, the application of the mechanism of e-Learning systems that helps construct and utilize learner corpora for language education has been proposed. It is a new attempt to integrate both the construction of learner corpora and their applications to daily pedagogical activities through e-Learning systems. This mechanism has provided a necessary foundation for corpus-based adaptive learning environment. The TCP has been operating since 2003 and the vitality of its role in Chinese language education at Waseda University has been confirmed by the improvement in students' language skills and the positive results of its usability evaluation. Moreover, teachers have analyzed the constructed learner corpora to understand students' learning aptitude and give appropriate advice to individual students. On the basis of the results mentioned above, it can be concluded that this research has achieved success in the fields of ICT-based Chinese language learning.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Yoshiyori URANO, for giving me the opportunity to work on this study and encouraging me to submit this dissertation. I am very grateful to him for giving me appropriate advice at various times.

I would also like to express my heartfelt gratitude to Professor Hidenori NAKAZATO, Professor Yoshinori SAGISAKA and Professor Wataru KAMEYAMA for their valuable suggestions during this research.

I am indeed grateful to Professor Shizuo HIKI for giving me valuable inputs and critiquing on my research. From him, I learnt a lot about the qualities of a good researcher.

I am truly grateful to Professor Kazuko SUNAOKA for providing me the opportunity to commence this challenging study in the Tutorial Chinese project at Waseda University. My sincere thanks to Professor Liming YANG for directing me into this research field and giving me sound advice when I became confused.

I would like to thank my teammates at the laboratory's e-Learning group for their time, support, and encouragement. When I encountered problems, they always helped me selflessly.

I would also like to thank the entire staff of GITS for their assistance and enthusiasm, which enabled me to achieve my goal.

Finally, I would like to thank my family for their support and understanding. They have taught me to appreciate and cherish the value of every day's endeavor. Their love enabled me to complete this long, yet rewarding journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

The recent advancements in information and communication technology (ICT) have changed our learning and teaching methods. With the availability of technologies such as the Internet, multimedia, database, and artificial intelligence, educators and learners need not remain in one place and can use their time effectively at their own discretion. ICT-based language learning systems can also be realized with the development of these technologies. Traditional forms of e-Learning technology, which used print, audio and video materials, have already existed for decades. The recent development and wide availability of Internet technology for connecting learners and educators supplies opportunities for interaction and collaboration online.

At Waseda University, a new language education model named Tutorial Chinese (TC) has been introduced since 2002. TC classes are small, usually consisting of one native speaking tutor and four students. One of the educational goals of this course is to provide more opportunities to Japanese students to learn Chinese with native speakers. The tutors are located in Tokyo, Beijing, and Taipei and employ a conversational style in conducting lessons directly in Chinese language. In order to support this education model, an e-Learning system named TCP (Tutorial Chinese Platform) has been constructed and many modules and functions have been designed. During the system design process, concept of integrating learner corpora construction and their applications to daily pedagogical activities has been taken. Reasonable learner corpora have been tried to utilize to make system more effective for language education.

This chapter describes the research background, previous studies, research objective and outline of whole dissertation.

1.1 RESEARCH BACKGROUND

1.1.1 E-LEARNING

The meaning of e in “e-Learning” is electronic. In its broadest definition, e-Learning includes instruction conducted through all electronic media, such as Internet, satellite broadcasts, audio/video tape, TV, CD-ROM, etc. Although education supported by information and communication technologies (ICT) has existed for long time, the term of e-Learning just appeared in the 1990s, coined by Jay Cross [Cross and Hamilton, 2002]. Even now there is no unified viewpoint in the definition of e-Learning. In this research, the Internet based definition is preferred, as the follows:

> *The use of Internet technologies to deliver a broad array of solutions that enhance knowledge and performance. It is based upon three fundamental criteria:*
> *Networked;*
> *Delivered to the end-user via a computer using standard Internet technology;*
> *Focuses on the broadest view of learning* [Rosenberg, 2001]

In this research, an e-Learning system has been constructed based on Internet, to gather and deliver contents and information to end-user—teachers, tutors and students—for the purpose of Chinese language pedagogy.

With the support of e-Learning, education has become more mobile, that is, educators and learners need not remain in the traditional classroom and may be in different countries in different time zones. E-Learning can be divided into synchronous and asynchronous based on the classification of time. Synchronous e-Learning uses technologies such as video conference or chat rooms to realize communication in ‘real-time’ which can be highly motivating in that distant learners feel less isolated and gain energy and advices from the learning group. Asynchronous e-Learning makes use of technologies such as print, video, and computer assisted discussions as e-mail, BBS, etc. The advantage of asynchronous interaction is that learners can participate and respond at their convenience, there is time for thought and reflection between responses, and it is possible to revisit discussions at a later date [C.White, 2003:9]. Nowadays, the trend of e-Learning is to combine synchronous and asynchronous forms in order to bring together the benefits of both. In this research, the pedagogy model named Tutorial

Chinese (TC), and e-Learning system named Tutorial Chinese Platform (TCP), also combined both synchronous and asynchronous elements:

- Interactive video conferencing system which links remote tutors and students (synchronous);
- Face–to–Face Tutorial Chinese class (synchronous);
- Internet-based discussion place for holding chats in classrooms (synchronous), and for posting and reviewing homework and class announcements (asynchronous);
- E-mail for sending messages, homework and feedback (asynchronous).

1.1.2 TUTORIAL CHINESE PROGRAM IN WASEDA UNIVERSTIY

An increasing number of students have applied for admission to the Chinese language program at Waseda University. Japanese students have an advantage in studying Chinese language due to similarities in the cultures of the two countries and the shared use of KANJI (Chinese characters). However, Chinese is a tonal language, and most Chinese syllables are pronounced with one of four tones. The same syllable will often have completely different meanings if pronounced with different tones. Chinese characters communicate ideas and meanings but provide only limited information about pronunciation; therefore, Japanese students learning to speak Chinese need to keep putting in more effort.

One of the main objectives of language education is to promote students' communication skills. In traditional classrooms, language is usually taught in classes of 20–30 students and students, with few opportunities for each student to speak aloud. In response to this situation, Waseda University began a new Chinese language education model named Tutorial Chinese (TC) in April 2002. The TC classes are small, usually consisting of one native speaking tutor and four students, and are supported by information and communication technology (ICT) tools such as video conference systems, computer-assisted instruction (CAI) applications, and an e-Learning platform [Sunaoka, 2002]. TC is part of a Waseda University project called Cross-Cultural Distance Learning (CCDL) [CCDL]. In each TC class, tutors who are located in Tokyo,

Beijing, and Taipei employ a conversational style in conducting lessons. TC consists of two modes: Distance Mode and Face–to–Face Mode [Sunaoka and Murakami, 2003] [Liu et al., 2004: a]. In Distance TC Mode, the tutor is in Beijing or Taiwan, and students are on Waseda University campus. Video conference systems are utilized to connect the tutor and students. In Face–to–Face TC Mode, the tutor and students are in the same classroom.

Distance Tutorial Chinese　　Face–to–Face Tutorial Chinese

Figure 1-1　Two modes of Tutorial Chinese

Because the pronunciation of Chinese syllables and tones is complex for Japanese students, coaching by native speaking tutors is essential from the elementary stage of learning [Sunaoka, 2002]. TC tutors must meet the following requirements.

- As a conversation supporter, the tutors should have the ability to advise students freely in Chinese, the ability to create conversation topics, and the leadership to let students express ideas spontaneously.
- As a judge of students' Chinese representations, the tutors should have knowledge and skills to determine the correctness of students' expressions.
- As a creator and provider of model sentences, the tutors should have the ability of designing lessons combined with teaching of new words and syntax.

Figure 1-2 below shows the contrast between traditional education and TC. TC has the following features:

- The small class style provides sufficient time for conversation and learning.

- The native speaking tutors teach students the most practical and living form of language thorough direct use of Chinese.
- The students, rather than the teacher, are main speakers, which is a better practice for improving their speaking and listening abilities.
- Grammar, vocabulary and conversation are integrated around conversation topics, and practice topics that reflect real life situations are used.

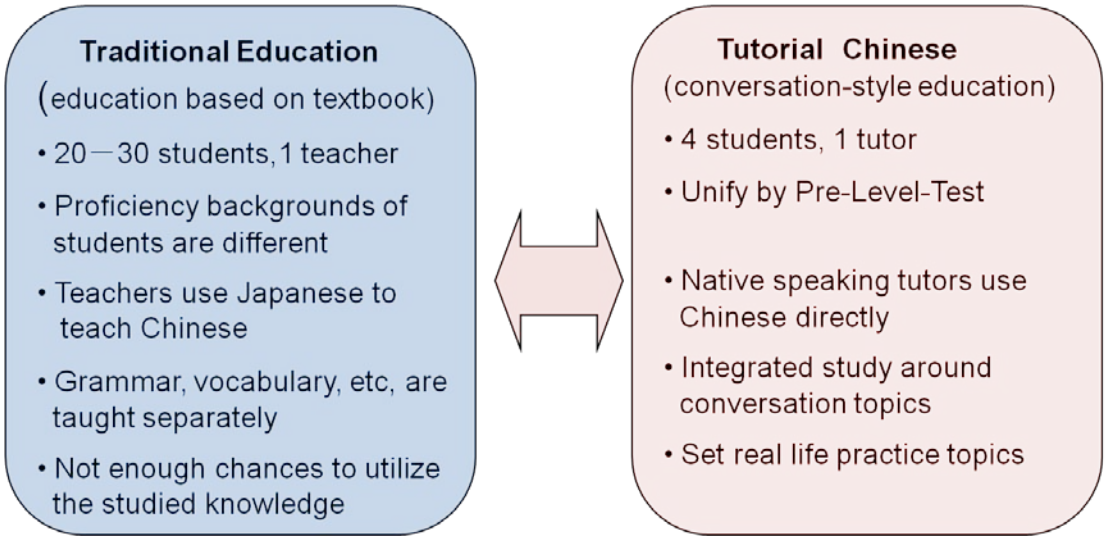


Figure 1-2 Contrast between traditional education and Tutorial Chinese

There are four roles in the management structure of TC [Waseda Univ. DCC] [Sunaoka, 2008]:

- Teachers take overall control of the course design, including the contents, schedule, and evaluation method of lessons;
- Tutors teach practical subjects; and report the status of students and lesson situation to teachers;
- Students take the course with the support of communication tools such as chat rooms, a video conference system and CAI tools;
- System managers provide technical support for the whole system.

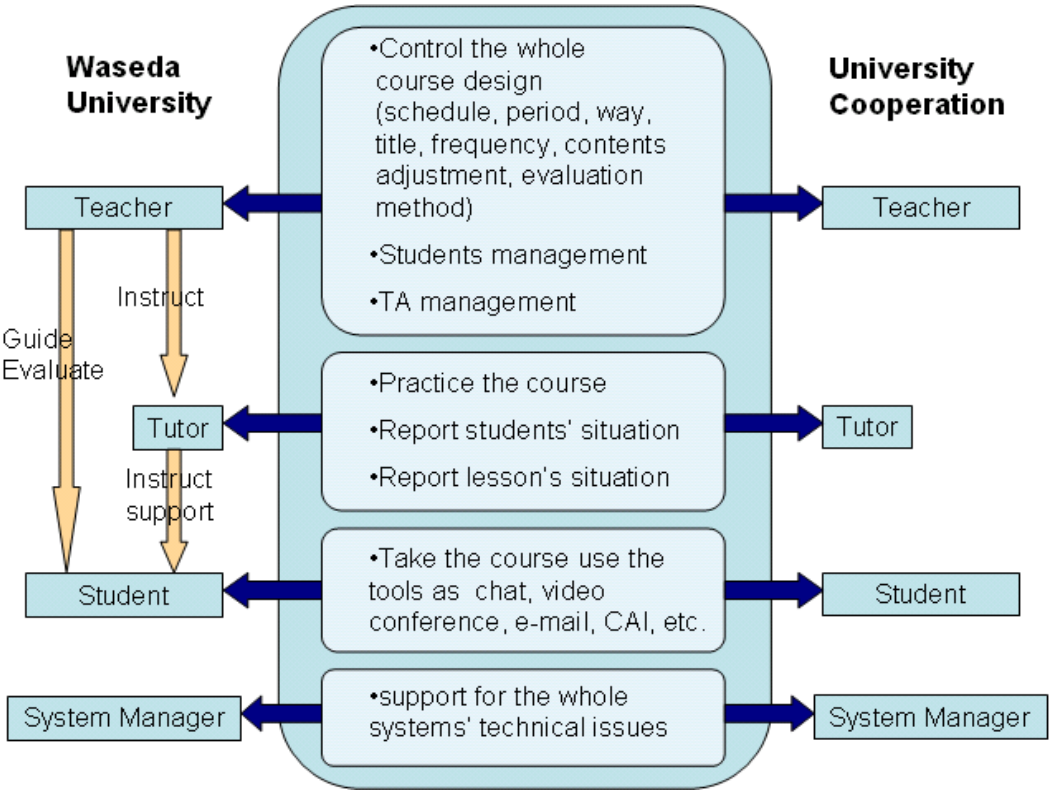


Figure 1-3 Tutorial Chinese management structure

In the TC model, in addition to face–to–face education, many synchronous and asynchronous e-Learning tools have also been utilized to resolve the problem of the scarcity of opportunities for communication. These include Bulletin Board System (BBS), an online composition/correction system, online chat rooms, and video conference system. With the assistance of these technologies, students receive many more opportunities to communicate with overseas native speaking tutors. Today, connections have already been established among Waseda University, Keio University, Peking University, Capital Normal University (Beijing), National Taiwan Normal University, and Korea University.

1.1.3 LEARNER CORPUS

The main concept of system design in this research is to integrate both the dynamic construction of learner corpora and their applications to daily pedagogical activities through e-Learning systems.

In principle, any collection of more than one text can be called a corpus: the term ‘corpus’ is simply the Latin for ‘body’ (the plural is corpora). Hence a corpus may be defined as body of text. However, in modern linguistics this term may be considered not fewer than four main headings:

- Sampling and representativeness
- Finite size
- Machine-readable form
- A standard reference

So a corpus in modern linguistics might more accurately be described as finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration. [McEnery and Wilson, 2001:32]. Nowadays, the common research areas where corpora can be used include: computational linguistics, cultural studies, discourse analysis, language education, machine translation, NLP (Natural Language Processing), etc. Specific to the field of language education, corpora have been considered by many teachers as useful tools currently. Corpora examples are important in language education as they expose students to the kinds of sentences that they will encounter in real life situations. The use of real examples of texts in language learning is not a new issue in the history of linguistics. However, corpus linguistics has developed considerably in the last decades due to the great possibilities offered by the processing of natural language with computers. The availability of computers and machine-readable text has made it possible to get data quickly and easily and also to have this data presented in a format suitable for analysis.

Until recently, most of corpora used in language education are collected from native speakers. The most common application of corpora in language education is to construct pedagogical materials, especially reference materials, to make them closer to native speakers’ standard usage. However, for language education, not only corpora

from native speakers are important, information from nonnative speakers—learners, is very important as well, because not only the standard model of language, but also the difficulties that learners encountered are essential to be known by language educators.

The first attempt in application of learners' information can be traced back to Error Analysis, but in the early era without support of advanced ICT, the scale of those data is always small. Nowadays, the practical use of technologies in language education classroom provides a platform where much direct interaction between learners and corpora can be expected. From the late 1980's, along with the combination of corpus linguistics and foreign/second language education, a new research area has appeared—research about learner corpus.

Learner corpus is a relative recent research topic. The simple meaning of learner corpus is just collections of spoken or written texts produced by second language learners. The definition of learner corpus discussed here is mainly connected with computer. According to Granger's definition, "the computer based learner corpus is electronic collection of authentic foreign language textual data assembled according to explicit design criteria for a particular foreign language teaching purpose. It is encoded in a standardized and homogeneous way and documented as to its origin and provenance" [Granger, 2002: 7]. In the definition, the "authentic" means "All the materials is gathered from the genuine communications of people going about their normal business" unlike data gathered "in experimental conditions or in artificial conditions". Because the learner corpus is mainly from classroom for language education, some degree of 'artificiality' will be involved. "Design criteria" are also very important because there is so much variation in language education. Some meta information about learners such as the learning context, mother tongue, level of proficiency, etc, should be considered when construct learner corpus.

Learner corpus makes more comprehensive studies possible. More general questions such as the frequency of different types of mistakes can be analyzed with learner corpus. Aspects of pragmatics and discourse, including communication strategies, can also be studied. Through comparison of languages of different type learners, the individual identities and common features of learners at a certain studying stage can be found, thus contributing to syllabus design of foreign language education, education contents development and classroom pedagogy to be more adaptive to learners' needs.

According to the collected method and pedagogical usage, learner corpus can be divided into two categories: corpus for delayed pedagogical use and corpus for immediate pedagogical use. Granger shows the contrast between the two categories as Figure 1-4 and Figure 1-5 below [Granger, 2009: 21].

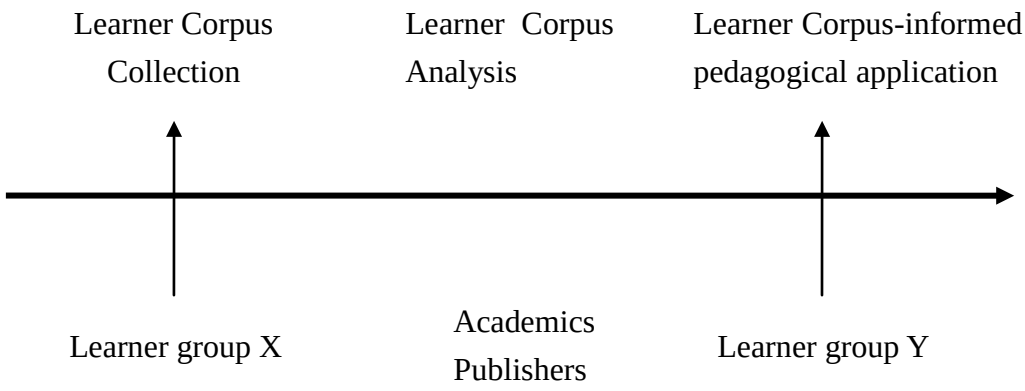


Figure 1-4 Learner corpus for delayed pedagogical use

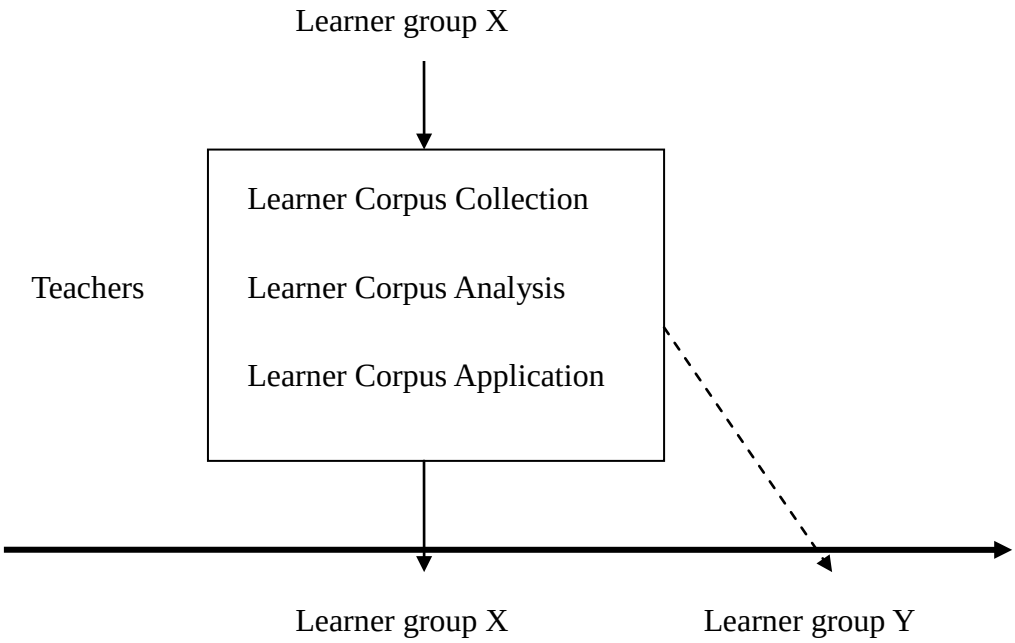


Figure 1-5 Learner corpus for immediate pedagogical use

Learners who have produced corpus for delayed pedagogical use (learner group X) cannot use the corpora directly as teaching/learning materials. The learner corpus will be analyzed and compiled by academics or publishers, and then improved pedagogical tools based on the learner corpus will be created and provided to the later learners (learner group Y). Here, learner group Y will be usually be similar type of learner group X, such as same mother tongue, background, same level of proficiency, etc. With the development of web-based learning platforms, the type of corpus for immediate pedagogical use has begun to appear. According to Franca, corpus for immediate pedagogical use is collected by teachers as part of their normal classroom activities and learners are at the same time producers and users of the corpora data (learner group X) [Franca, 1999]. Of course, other similar-type learner groups (learner group Y) can also utilize the data, as shown by the dotted arrow in the figure 1-5.

In corpus for immediate pedagogical use category, because the learner corpora will be used by the creators themselves, education will be more targets oriented. As Mukherjee and Rohrbach mention, the exploration of learner data by learners themselves will motivate more learners to reflect on their language use and thus raise their foreign language awareness [Mukherjee and Rohrbach, 2006: 228]. It also provides necessary foundation for the construction of adaptive learning environment. The focal point of this research is to achieve this kind of learner corpus.

1.2 RESEARCH OBJECTIVES

TC tutors are located in Tokyo, Beijing, and Taipei; therefore, the program educators and learners are not in the same place. In order to support this education model, an e-Learning system named the Tutorial Chinese Platform (TCP) has been constructed. In the system design process, demands are considered from the viewpoints of teachers, students, and researchers.

TC gives students more chances to study and practice language with native speaking tutors. However, only tutors and students are in the classroom in this education model; the curriculum designer—teachers—cannot gather real-time educational information directly from the classroom. Using the normal classroom education model, teachers can obtain this kind of information from students through face–to–face communication and adjust the lessons or class schedule in a timely manner. For example, when most students in the classroom cannot understand a topic, teachers will immediately adjust their focal point to this topic. If it is a very representative problem, they can even add additional lessons on the topic to future teaching programs. However, teachers are not in the classroom in the TC education model and thus cannot obtain real-time educational information directly from students. This results in the first requirement of the TCP—to support teachers to gather timely educational information from the classroom.

The TC curriculum is carried out by integrated study around conversation topics, and tutors should only speak Chinese. Through this type of education, students' listening and speaking abilities greatly improve, but other language knowledge and skills such as vocabulary ability, writing ability, tone discrimination ability, and so on, cannot be practiced enough within of the context of classroom instruction. Support CAI tools should be developed and utilized to supplement education in the classroom. Thus, the second requirement of the TCP is following:—to support students to improve their language skills that cannot be practiced in the classroom.

Learner corpora are important in language education as they expose students to the kinds of sentences that they will encounter in real life situations. The data created by learners in daily language education can be of use to language educators and linguistics researchers. The third requirement of the TCP—to support corpus linguistics researchers to collect corpora generated in real language educational environments.

Based on these three demands, the TCP has been constructed. The main concept of the system design is to integrate the construction of learner corpora and their applications to daily pedagogical activities through e-Learning systems. The learner corpora are dynamically generated from TC classrooms every week, and the system's goal is to utilize these learner corpora as the contents of feedback to students as soon as possible. Therefore, the objective of this study is to find solutions to the following problems:

- How can the construction of learner corpora be efficiently integrated to daily pedagogical activities of TC?
- How can the constructed learner corpora be effectively applied to TC pedagogy?

The challenge of this research is how to integrate the solutions of above two problems to pedagogical activities through e-Learning system without teachers extra workload.

1.3 PREVIOUS STUDIES

A growing number of studies based on learner corpora have developed in the last twenty years. Tono has summarized the application fields into four main categories: studies related to error analysis, investigating quantitative differences between native and non-native language, describing features of interlanguage, and applying learner corpora-based research to language pedagogy [Tono, 2003]. Subject of this research is belonging to the last category.

According to Tono, there are still very few studies which relate the findings from learner corpora to actual classroom practice currently [Tono, 2003]. Even if there are some examples of pedagogical innovations, most of them are applications of learner corpora for delayed pedagogical use. For example, Milton investigates Chinese learners' problems in their English writing and implemented a CALL system named AutoWord to assist learners in improving the lexical, grammatical, and discoursal aspects of their writing in English. The components such as error recognition exercises and grammar tutorials have been constructed in the system. The research is based on the Hong Kong University of Science and Technology (HKUST) learner corpus, which is comprised of English texts of Chinese students at the senior high school level [Milton, 1998]. Smrz's project at Masaryk University in Czech Republic makes use of static corpora such as the British National Corpus (BNC) [BNC] and the Times Corpus [Time Corpus] to create vocabulary test queries. The standard form of the vocabulary test is generated by teacher based on concordances of words and students are required to fill the gap [Smrz, 2004]. These two applications are both using constructed learner corpora to create exercise to improve some kind of language ability. The learners who produced those corpus cannot use corpus-based applications for themselves.

Another example involves construction of learner corpus. Wible has constructed a language learning environment named Intelligent Web-based Interactive Language Learning (IWiLL), which allows both teachers and students to create and search an online database of student essays and teacher annotations [Wible et al., 2001]. In this project, learner corpora are constructed timely through e-Learning platform, however, there is no specific proposal on the pedagogical application of these constructed learner corpus.

In contrast to previous studies, the environment of learner corpus for immediate pedagogical use has been implemented in this research. An e-Learning system has been designed to integrate both construction and application of learner corpora to daily pedagogical activities. Corpora are constructed dynamically with the process of pedagogical activities through e-Learning system. Because learner corpora will be used by the creators themselves, corpus-based pedagogy will be more targets oriented. The following table shows the contrast among corpus-based pedagogical applications of previous studies and this research.

Table 1-1 Contrast among corpus-based pedagogical applications

Research Project	Corpus	Pedagogical Applications	Features
John Milton, HKUST	HKUST Learner Corpus: essays of senior high school level	AutoWord: error recognition exercises, grammar tutorials	Apply the **constructed corpus** to design pedagogical tools for improving writing ability. LC for delayed pedagogical use
Pavel Smrz, Masaryk Univ.	The British National Corpus, Times Corpus	Corpus-based vocabulary test. Test queries are generated based on concordances of words.	Use **static corpus** to generate vocabulary test for improving vocabulary ability LC for delayed pedagogical use
David Wible, Tamkang Univ.	Corpus consists of students' essays and teachers' annotations	Online writing platform	**Construct** and search an online database of student essays and teacher annotations for improving writing ability LC for delayed pedagogical use

Tutorial Chinese Platform, Waseda Univ.	WTCC: vocabulary corpus, composition/correction corpus, tone discrimination corpus	Corpus-based vocabulary CALL , online writing/retrieval platform, tone discrimination CAI	Integrate both **construction** and **application** of LC to pedagogy. Corpus is **dynamic**, reflects students' real time situation. LC for immediate pedagogical use

This research has provided a necessary foundation for the construction of corpus-based adaptive learning. On the basis of environment of learner corpora for immediate pedagogical use, learners' real time learning aptitude, language ability and learning requirements can be analyzed from the combination of learners' profile and learner corpora. After that, learning materials adapt to individual learner can be selected based on the result of analysis. The methods of learner's aptitude analysis and mapping of appropriate learning materials are remained two topics in the next step research.

1.4 DISSERTATION ORGANIZATION

This dissertation consists of six chapters, Chapter 1 introduces the research background; Chapter 2 describes proposal of this research; Chapter 3, Chapter 4, and Chapter 5 introduces the three proposed applications corresponding to three focused language skills in this research; Chapter 6 describes the conclusion and future prospect. Detail of dissertation is organized as the follows:

Chapter 1 introduces related research fields, including e-Learning, learner corpus, and the current situation of Chinese language education at Waseda University. The importance and necessity of the TCP and the objectives of this research are presented.

Chapter 2 describes the research proposal — the working mechanism is needed to integrate the construction of learner corpora and their applications to daily pedagogical activities, and the design for Waseda Tutorial Chinese Corpus (WTCC) are described. Corresponding to three important language skills that need be practiced through e-Learning system, the following three chapters described three applications and corresponding corpus respectively.

Chapter 3 introduces the Computer Assisted Language Learning (CALL) drill sub-system for improving of vocabulary ability, which combines the lesson report module of the TCP with WTCC to collect new vocabulary used in the classroom into vocabulary corpus. Meanwhile, the mechanism of analyzing the corpus and semi-automatically creating learning exercises has also been implemented. This system has features such as (a) web-based user interface for students and teachers to facilitate system access, (b) integration with WTCC corpus to make drills more effective, (c) layer architecture to make authoring flexible and reusable, and (d) a mobile phone-based ubiquitous environment to make the drills accessible to users anytime and anywhere. The mechanism implemented in the CALL drill sub-system has been employed in TC since 2004 and achieved satisfactory results.

Chapter 4 describes the composition/correction sub-system with a corpus retrieval function. The concept of this sub-system is to take advantage of the Internet, implement online composition/correction functions similar to that in a paper-based work environment, and simultaneously store the compositions and correction information into

the composition/correction corpus through the correction process. This sub-system is effectively used not only by students in writing composition but also by teachers while correcting and providing comments on the compositions. Moreover, the compositions together with the correction information have been tagged and constructed automatically into the learner corpus, and the retrieval function has also been developed, which makes it convenient for researchers to analyze this kind of language resource. The result of the usability evaluation shows that the composition/correction sub-system has dramatically increased teachers' work efficiency.

Chapter 5 cites a typical example named the Computer Assisted Instruction (CAI) sub-system for self-teaching of discriminating Chinese four tones. Since Mandarin Chinese is a tonal language, it is a big barrier for Japanese learners in the understanding of the features of the tones when studying Chinese. In order to give them a more direct understanding of the Chinese tones, this sub-system presents a visualized pitch pattern of the tones and gives effective guidance derived from the analysis of tone discrimination error corpus. The system is mainly designed for elementary level students. Precise acoustic data of the Chinese tones have been utilized for designing the system. Characteristics of the errors found in the examination have been analyzed, and these characteristics have been carefully considered while constructing the system. Examination of the improvement in the students' scores after practicing with the CAI system has clearly confirmed the effectiveness of the system. Moreover, the result of the analysis of the students' answers has provided valuable insights for teachers to make effective education plans.

Finally, chapter 6 concludes the dissertation and proposes the possibility of future development.

CHAPTER 2.
PROPOSAL OF THIS RESEARCH

This research proposed a mechanism that integrates both construction and application of learner corpus to daily pedagogical activities of Tutorial Chinese through e-Learning systems. Corpus-based exercise generation function has been designed as a core component in this mechanism to convert learner corpus to pedagogical application. Corresponding to three important language skills that need be practiced outside of TC classroom—vocabulary ability, writing ability, and tone discrimination ability, three applications and three corresponding corpora have been constructed.

2.1 WORKING MECHANISM OF TUTORIAL CHINESE

In order to efficiently integrate the construction of learner corpora to daily pedagogical activities of TC and effectively apply the corpora to TC pedagogy, the TCP has been designed and constructed. Figure 2-1 shows the working mechanism implemented in this research. The TCP works as a bridge to connect the pedagogical activities in TC classrooms and the learning process through e-Learning systems outside of classrooms.

The TCP provides fundamental functions to support teachers, tutors, and students in TC pedagogy, such as schedule control, user management, and lesson reports. Tutors will send reports to the teacher after every week’s class through the TCP, enabling the teacher to receive timely educational information in their classrooms. Meanwhile,

through functions such as the tutor lesson report and composition/correction system, different kinds of corpora will be generated and gathered into the learner corpora of this study, called the Waseda Tutorial Chinese Corpus (WTCC) [Sunaoka, 2005]. In order to utilize these learner corpora as feedback to students as soon as possible, the CALL drill system has been designed and functions such as automatic or semi-automatic queries generating from a vocabulary corpus have been developed. Students will be required to do a CALL drill after class, and the results will be sent to their tutors and teachers through the TCP. This information will be provided to tutors as reference material on student progress for the next week's classroom instruction. In this proposed mechanism, teachers, tutors, and students are involved in the process of generating, gathering, analyzing, and utilizing learner corpora. This process works as a dynamic cycle with valuable learner corpora being stored and applied.

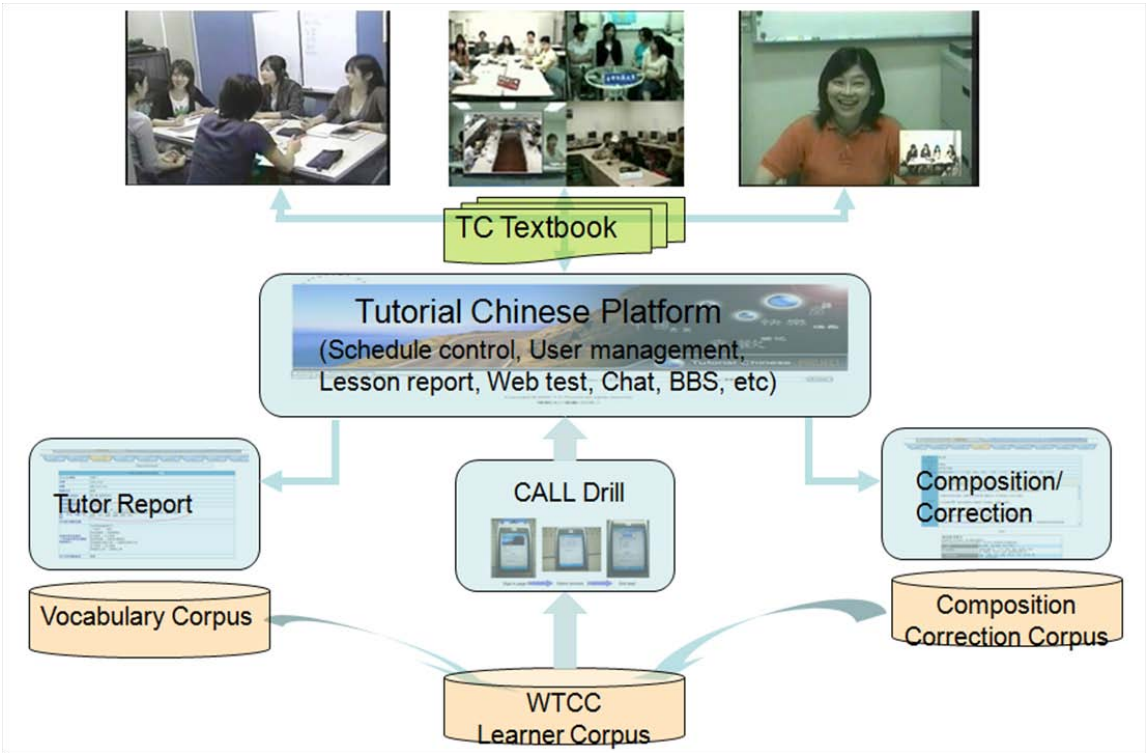
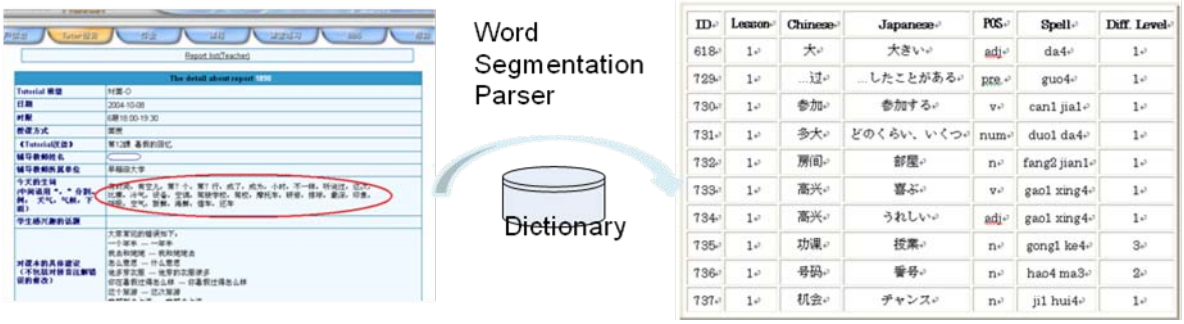


Figure 2-1 Working mechanism of Tutorial Chinese

2.2 OVERVIEW OF THE CORPORA IN THIS RESEARCH

In this study, educational information has been collected into the WTCC, which is mainly composed of a vocabulary corpus and composition/correction corpus. The former is dynamically incremented by tutor reports, and the latter is automatically tagged and gathered through teachers' daily correction activities through composition/correction systems. Retrieval tools have also been developed for the composition/correction corpus, and analysis can be implemented with the combination of the corpus and learners' profile information stored in the system, such as gender, age, mother tongue, ability level, studying context, and experience.

The vocabulary corpus is the initial and most basic part of the WTCC. It has already been utilized in CALL exercise construction. The base of the WTCC vocabulary corpus comes from three sources: The grammatical knowledge base of contemporary Chinese corpus by Beijing University (approx. 10,000 words) [Yu et al., 1998], HSK (Hanyu Shuiping Kaoshi, approx. 8000 words) [HSK, 2001], and the Japanese basic university educational vocabulary corpus (approx. 3000 words) [Shi, 2003]. More information has been added to these basic corpus sources, including Japanese translations, Chinese pronunciation in pinyin, and difficulty level tags. The vocabulary corpus is linked with the lesson report module in the TCP in order to be incremented with new vocabulary. The work process of linkage between the tutor lesson report module and the WTCC vocabulary corpus is shown in Figure 2-2 (Details will be described in Chapter 3):



ID	Lesson	Chinese	Japanese	POS	Spell	Diff. Level
618	1	大	大きい	adj	da4	1
729	1	...过	...したことがある	pre	guo4	1
730	1	参加	参加する	v	can1 jia1	1
731	1	多大	どのくらい、いくつ	num	duo1 da4	1
732	1	房间	部屋	n	fang2 jian1	1
733	1	高兴	喜ぶ	v	gao1 xing4	1
734	1	高兴	うれしい	adj	gao1 xing4	1
735	1	功课	授業	n	gong1 ke4	3
736	1	号码	番号	n	hao4 ma3	2
737	1	机会	チャンス	n	ji1 hui4	1

Figure 2-2 Construction of WTCC vocabulary corpus

The composition/correction corpus is automatically tagged and gathered through teachers' daily correction activities through the composition/correction system. In order to reduce the discomfort of online operation, the system has provided template-based coding rich-text tools, such as underline, strikethrough, and bold mark. Paper-based operations such as insert, replace, delete, and comment can also be performed online. Meanwhile, all the comments and correction information will be automatically transferred into XML tags and stored in the correction data corpus. Details of the construction issues of the WTCC composition/correction corpus will be described in Chapter 4. The process is shown in Figure2-3:

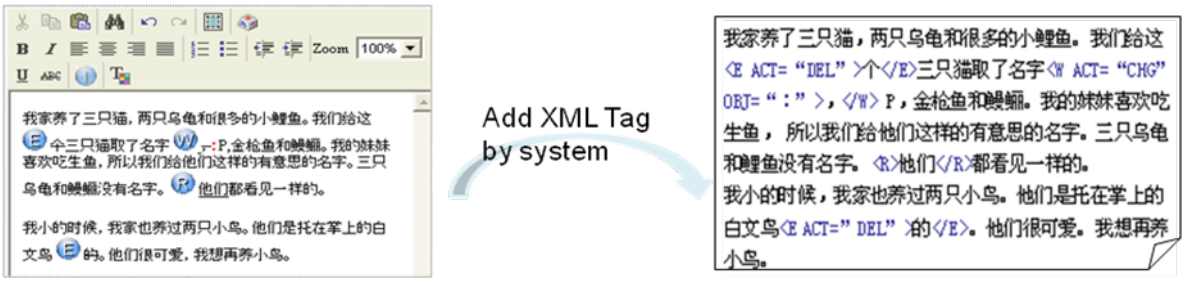


Figure 2-3 Construction of WTCC composition/correction corpus

In addition to these two corpora in the WTCC, there is another kind of corpus collected in the CAI sub-system to record students' exercise results in the self-teaching of discriminating the four tones of Chinese. Based on this corpus, characteristics of mistakes made by Japanese students during their discrimination of Chinese tones can be analyzed and carefully considered during construction of the system. Details of this system will be described in Chapter 5.

2.3 SYSTEM DESIGN OF TCP

The first version of the TCP was constructed in 2003. In order to support teachers and tutors to carry out language education in TC with less effort than and even in a better way in the normal classroom, many function modules have been designed and constructed [Liu et al., 2004: b, c]. For example, the web test module and schedule and class management module have been designed to help teachers create classes with similar level students; this is important for making reasonable arrangements of instructional design [Murakami et al., 2003, 2004, 2005]. The lesson report module has been designed to help teachers master the latest information about the classroom and students, which is important for adjusting the curriculum to adapt to students' current situations [Sunaoka et al., 2004b, 2005]. In the later version, some sub-systems have also been designed and constructed to help students improve their listening, speaking, reading and writing abilities. For example, the composition/correction sub-system with a corpus retrieval function has been designed to support students to improve their writing ability [Liu et al., 2006a, b] [Sunaoka and Liu, 2006]; a mobile phone–based CALL drill sub-system has been designed to support students to improve their vocabulary ability [Liu et al., 2004a, 2005a, 2005b].

Figure 2-4 below shows the function structure of the whole system. The system consists of many modules supporting basic pedagogical activities of TC. There are also some systems designed to support students to improve their language skills.

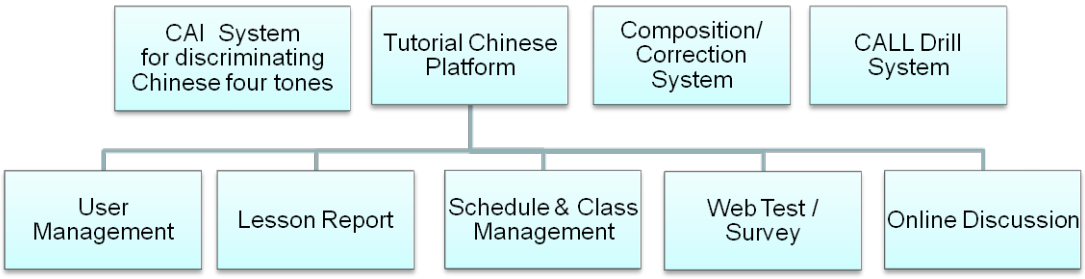


Figure 2-4 Function structure of the whole system

1. User Management Module

The user management module will authorize different interfaces based on different

user roles. Students can read information on their profile, lesson reports, compositions, and results of web tests and drills; they can also submit homework and take tests or surveys. However, they cannot delete the information or read other students' information. Tutors can read the information of students in their class, correct homework submitted by these students and submit and modify lesson reports. However, they cannot read or modify information from outside of their class. Teachers have the highest authority and can read or modify all the information. The table of user in database is shown in Table 2-1.

Table 2-1 Attribute of table TC_USER

Field Name	Attribute	Interpretation
User_ID	int(11)	user id
User_Name	varchar(30)	user name
User_Name_Katakana	varchar(100)	pronunciation of name
User_Password	varchar(12)	user password
User_Group	tinyint(4)	type of user(teacher/tutor/student)
User_Sex	tinyint(4)	user gender
User_Age	year(4)	the birth year
User_Email	varchar(100)	user email
User_Tel	varchar(14)	user telephone number
User_Student_No	varchar(9)	user student id (for login)
User_Department	tinyint(4)	user department in university
User_Grade	tinyint(2)	user grade in university
User_Class	varchar(10)	user's class in Tutorial Chinese
User_Level	varchar(20)	initial language level of user
User_Extra	text	other extra information
User_Delete	tinyint(1)	deleted or not

2. Lesson Report Module

In order to support teachers and tutors to carry out education effectively in TC, the educational information of the classroom should be gathered and reported to teachers as soon as possible for instructional design. The online lesson report module has been designed for this purpose. Three types of information are included in lesson reports.

(1) Information about the lesson with a fixed format, such as class name, date, period, type of class (distance or face-to-face), which lesson in the textbook, and tutor's information.
(2) Educational information generated in the classroom, such as new words that appeared outside of the textbook, hot topics in the classroom, common errors and advice to the class, and comments on the hardware environment.
(3) Evaluation information given to an individual student, such as attendance record, status of preview, status of speech, grammar, vocabulary, and advice to the student.

Tutors should input this report through the TCP as soon as possible after finishing a lesson. All the reports will be sent in a list to the teacher. Some special reports, for example, the reports of students who are absent many times or do not actively participate in discussion frequently will be listed in a significant place so that even though the teacher is not in the classroom, the real-time information of the class can be received. Moreover, information gathered in the Lesson_New_Words field, after analyzed by the teacher, will enhance the vocabulary corpus in the WTCC for future pedagogical activities.

The database tables related to lesson report module are shown in Table 2-2 and Table 2-3. Table TC_LESSON is for gathering information appeared about one lesson. Table TC_REPORT is for gathering information of individual student.

Table 2-2 Attribute of table TC_LESSON

Field Name	Attribute	Interpretation
Lesson_ID	int(11)	lesson id
Lesson_Class_ID	tinyint(4)	the class id
Lesson_Tutor_ID	int(11)	id of tutor who teach this lesson
Lesson_Date	date	lesson date
Lesson_Period	tinyint(4)	university period of lesson
Lesson_Type	tinyint(4)	type of class(Distance or F-to-F)
Lesson_Room	tinyint(4)	classroom
Lesson_Text_ID	tinyint(4)	id in textbook
Lesson_New_Words	text	new words appeared
Lesson_New_Topics	text	hot topics appeared
Lesson_Textbook_Advice	text	advice to textbook

Lesson_Requirement	text	comments to environment
Lesson_Extra	text	other extra information
Lesson_Edited	tinyint(1)	lesson report be submitted or not

Table 2-3 Attribute of table TC_REPORT

Field Name	Attribute	Interpretation
Report_ID	int(11)	report id
Report_Lesson_ID	int(11)	lesson id in this report
Report_Student_ID	int(11)	student id in this report
Report_Speaking_Ability	tinyint(1)	speaking ability of this student
Report_Presence	tinyint(1)	attendance of this student
Report_Delay	tinyint(2)	how many delay minutes if the student be late
Report_Prepare	tinyint(1)	prepared situation this student
Report_Communication_Skill	tinyint(1)	communication skill of this student
Report_Speaking_Effect	tinyint(1)	Speaking effect of this student
Report_Speaking_Time	tinyint(2)	how many minutes the student spoke
Report_Listening_Ability	tinyint(1)	listening ability of this student
Report_Pronunciation	tinyint(1)	pronunciation skill of this student
Report_Pronunciation_Problems	Text	what is the problems appeared in the student's pronunciation
Report_Grammar	tinyint(1)	grammar ability of this student
Report_Grammar_Problems	Text	what is the problems appeared in the student's grammar
Report_Advice	Text	advice to this student(if any)

3. Schedule and Class Management Module

Each year, over 200 students take the TC course. If the language abilities of students vary too greatly, it is difficult for tutors to implement lessons. The schedule and class management module, together with the web test module described in (4) below

have been designed to settle this problem. Students will be asked to take a pre-test through the web test module, and the results will appear in the schedule and class management module. Then, the teacher will group students with the similar basic abilities into the same class. The information of class is saved in the table TC_USER.

4. Web Test / Survey Module

As mentioned in (3) above, a pre-test will be given to students before the semester starts. All questions in the test are single choice. Based on the results of the pre-test, students with similar levels of ability will be grouped into the same class. The table of the web test in the database is shown in Table 2-4.

Table 2-4 Attribute of table TC_WEBTEST

Field Name	Attribute	Interpretation
Webtest_ID	int(11)	web test result id
Webtest_User_ID	int(11)	id of user who took webtest
Webtest_Time	datetime	time of web test be taken
Webtest_Done	tinyint(1)	web test submitted or not
Webtest_Result	text	result of webtest
Webtest_Correctness	text	the correctness of each question
Webtest_Detailtime	text	time of each question

In the Webtest_Result, Webtest_Correctness and Webtest_Detailtime field, the format is decided as Figure 2-5. This design makes it possible for an unlimited number of question results to be saved to the database.

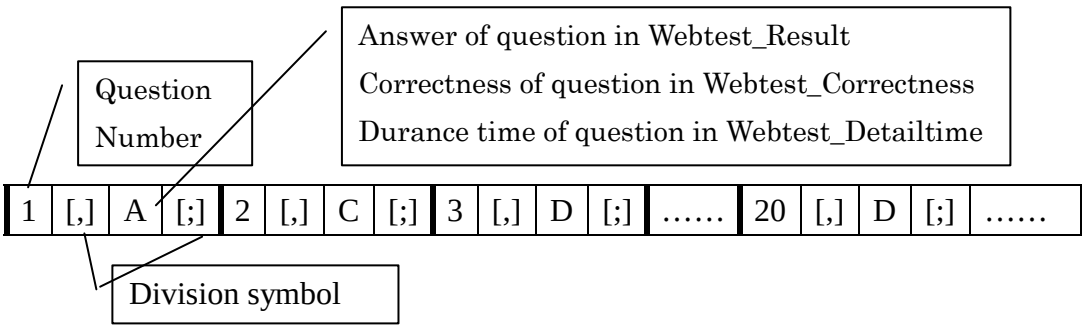


Figure 2-5 Format of data field in web test module

The format of survey's table is similar to web test. The table in database is shown in Table 2-5.

Table 2-5 Attribute of table TC_SURVEY

Field Name	Attribute	Interpretation
Survey_ID	int(11)	survey result id
Survey_User_ID	int(11)	id of user who took survey
Survey_Time	datetime	time of survey be taken
Survey_Done	tinyint(1)	survey submitted or not
Survey_Result	text	result of survey

In recent version of survey, I have extended the survey module to a data collection system supporting both PC and mobile phones. The key features of new survey module are [Liu et al., 2009: a,b] :

(1) Ease of use. The creator needs no special programming skill. The operation of creating, deleting, modifying, analyzing of survey is very easy, just like operation in normal Word or Excel.
(2) No special request for users. Access is through a standard Web page so there is no need to download or install special tools.
(3) Templates are used to standardize question formats. These templates are designed to be flexible so questions can be added or modified.
(4) The results can be collected and viewed in real time. The results can be saved into a standard Excel file and could be downloaded.

5. Online Discussion Module

The role of the online discussion module is to supply communication functions for interactive Chinese language education. In the TCP, not only a public discussion space for all students and teachers, but also discussion spaces for each unit of a class have been designed. The table of discussion in database is shown in Table 2-6.

Table 2-6 Attribute of table TC_DISCUSSION

Field Name	Attribute	Interpretation
Dis_ID	int(11)	discussion thread id
Dis_User_ID	int(11)	id of user who submit a thread
Dis_Time	datetime	time of thread be submitted
Dis_Parent	tin(11)	the id of thread in the upper level of current thread
Dis_Title	text	title of thread
Dis_cotent	text	body content of thread

2.4 SYSTEM IMPLEMENTATION

2.4.1 IMPLEMENTATION ENVIRONMENT

1. Browser/Server Structure

B/S (Browser/Server) mode has been taken in Tutorial Chinese Platform. In the server side, web server has been used as the bridge between user and database. Through *Forms*, users' requests will be sent to database and then feedback the results to users. In the client side, no special software installs are required and it ensures that fairly standard computers will be ready to use the system. Figure 2-6 shows the working structure of system. [Liu and Urano, 2003a]

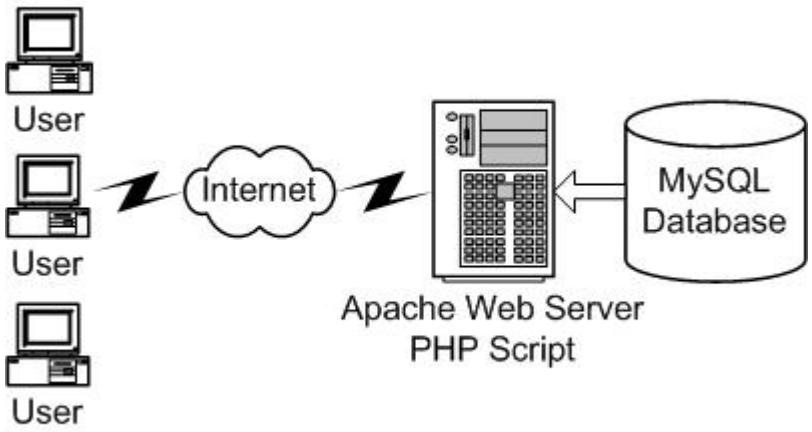


Figure 2-6 TCP system working structure

2. LAMP Environment

The set of LAMP (Linux, Apache, MYSQL, and PHP) has been used as the development environment in this system.

(1) Linux [Redhat Linux]

The operating system of server is the standard RedHat Linux. I choose it because it provides the best support of execution for open source software specified later. Linux also provides a multitasking environment that is required if the platform is to be used by more than one user. It also has the advantages of stability, support for industry standard communication protocols, user community support and large software application base. Therefore Linux may be considered a low risk platform for this kind of development.

(2) Apache Web Server [Apache]

Apache web server is used to provide the hosting of both static and dynamic webpage from platform. It has a reputation for robustness, stability and good support for integration with other open source software – compatibility with PHP are strong points of this software.

(3) PHP Scripting Language [PHP]

PHP named Hyper Text Pre-Processor, which is a server-side, cross-platform, HTML embedded scripting language. Much of its syntax is borrowed from C, Java, and Perl, so it is very easy to learn and use. It has powerful database access functions. It has nearly all the features needed for building a web site: designing, objects, database access, network protocols, and security. Its quick response time, multi-threaded nature, and transparency to the end user make it ideal for developing dynamic web sites and applications.

(4) MYSQL Database Server [MySQL]

MYSQL offers a rich and very useful set of functions. The performance, connectivity, speed and security make MYSQL highly suited for accessing databases on the Internet. MYSQL is a client/server system that consists of a multi-threaded SQL server that supports different back-ends, several different client programs and libraries, administrative tools, and a programming interface.

LAMP mode has many advantages: (a) the entire technology stack is available through open-source; (b) it works fine for most applications; (c) it is easy to learn; (d) it allows one to build a web application quickly; and (e) there are many open source code examples available that make creating an entire web application even easier. The combination of Linux, Apache, PHP and MYSQL has formed an ideal network database environment. So this kind of technology stack has been widely adopted.

2.4.2 INTERFACE OF TCP MODULES

The figures below show the interface of modules introduced above.

Figure 2-7 Login interface of TCP

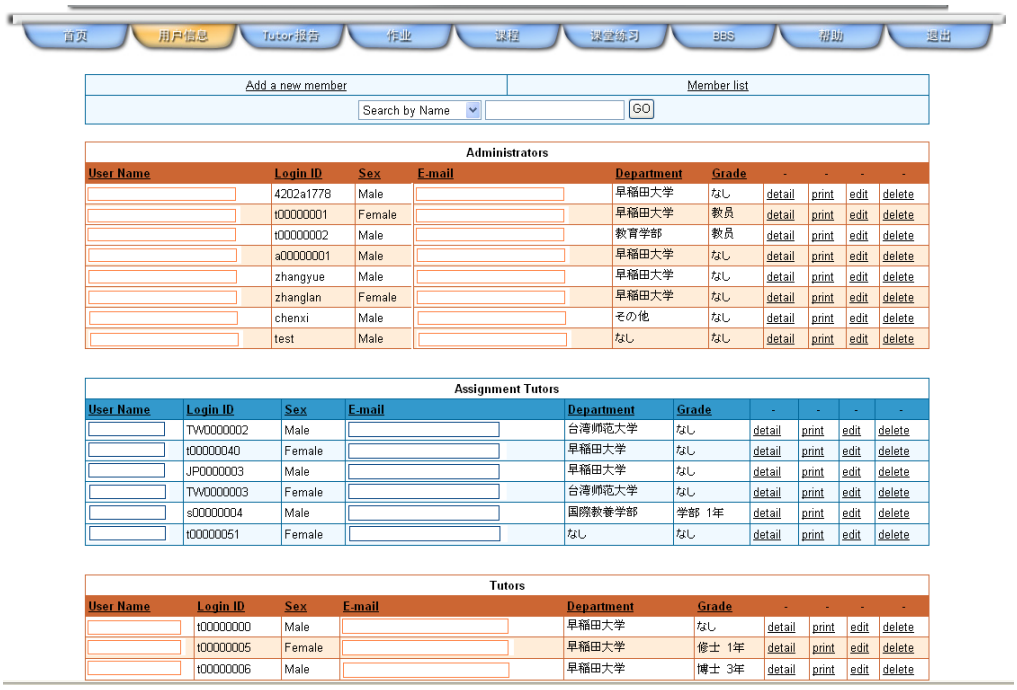


Figure 2-8 User management module interface

Figure 2-9 Lesson report module interface

Figure 2-10 Schedule & class management module interface

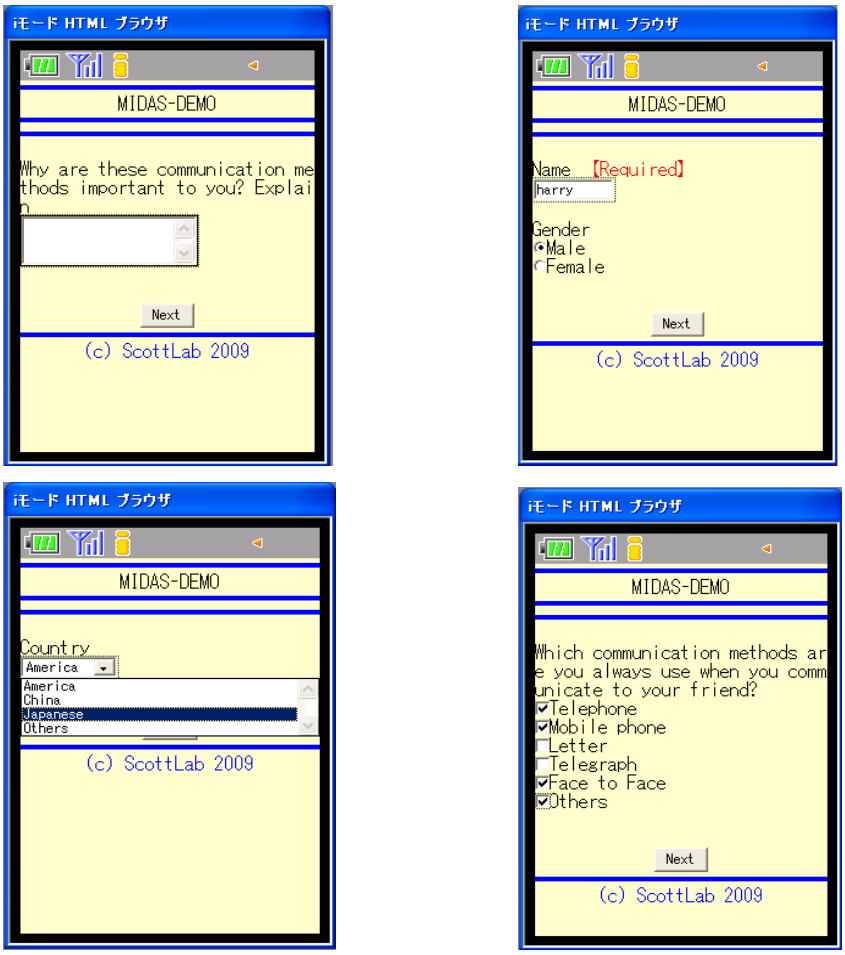


Figure 2-11　Web test / survey module interface

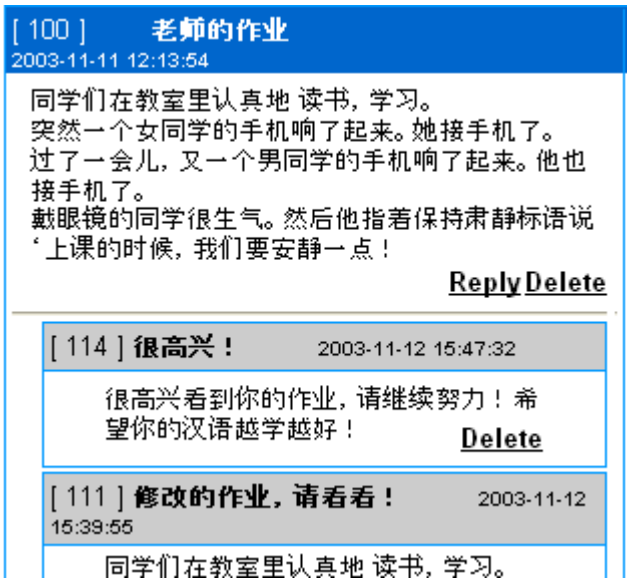


Figure 2-12　Online discussion module interface

2.5 EVALUATION OF TCP AND TUTORIAL CHINESE

The TCP was first implemented in 2004. There have been more than 200 students and 50 tutorial classes every year. The current state of the TC program is shown in Table 2-7.

Table 2-7 Tutorial Chinese Operation

TUTORIAL CHINESE PROGRAM	
Object	All students in Waseda University
Subjects Name	Chinese Conversation Practice
Class Style	1 Tutor 1 −4 Students
Class Quantity	10 weeks (twice a week, 20 lessons in total)
Curriculum	1.5 hour for one class + Self study
Credit	4 credits
Teacher and Tutor	2 teachers in JP, 4 teachers outside of JP 5 tutors from University of Peking 5 tutors from Capital Normal University 10 tutors from National Taiwan Normal University 15 tutors from Waseda University

At the end of 2004, students were surveyed about the TC program and system implementation status. The number of effective answers was 80. The evaluation concerns

1. Do you think you have enough chance to speak Chinese during the course?
   [A] have enough chance
   [B] have some chance
   [C] have chance occasionally
   [D] have a little chance
   [E] have no chance

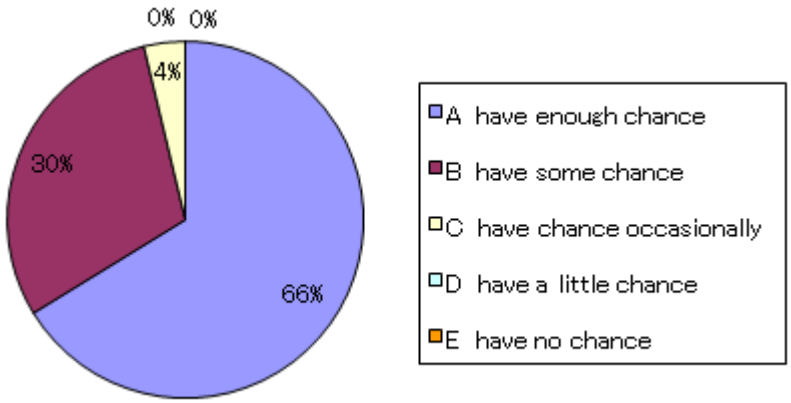


Figure 2-13　Result of evaluation of TC (1)

2. Do you think the discourse themes are appropriate for improving your communicational skill?

[A] appropriate

[B] nearly appropriate

[C] not very appropriate

[D] not appropriate

[E] have no idea

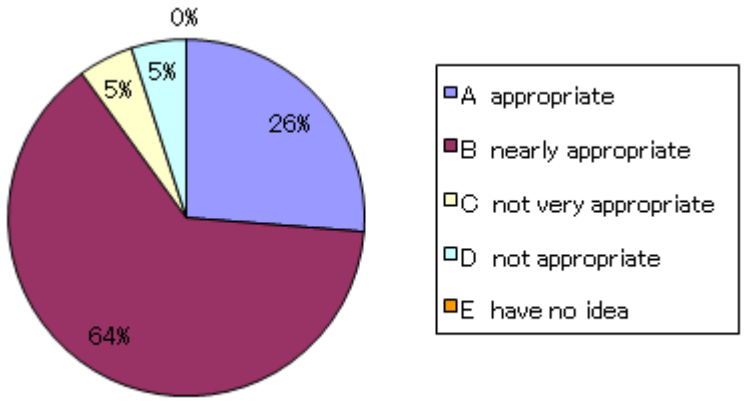


Figure 2-14　Result of evaluation of TC (2)

3. Do you think the level of current class is appropriate to your ability?
   [A] too lower than my ability
   [B] a little lower than my ability
   [C] just appropriate
   [D] a little higher than my ability
   [E] too higher than my ability

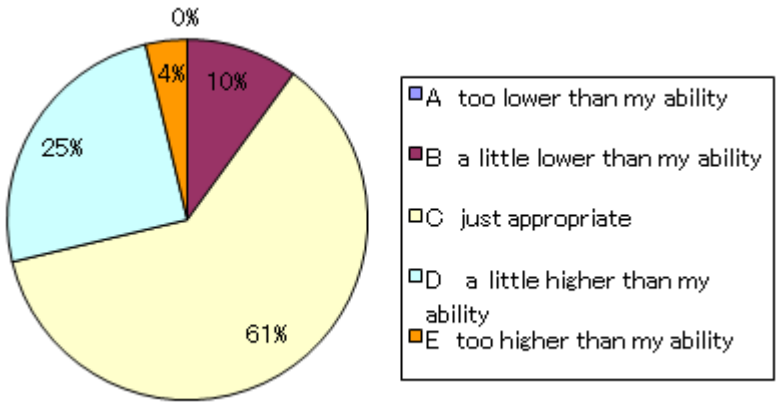


Figure 2-15 Result of evaluation of TC (3)

4. Do you mind the time delay through video conferencing in distance TC?
   [A] not at all
   [B] almost no problem
   [C] been used to that
   [D] a little mind
   [E] be intolerable

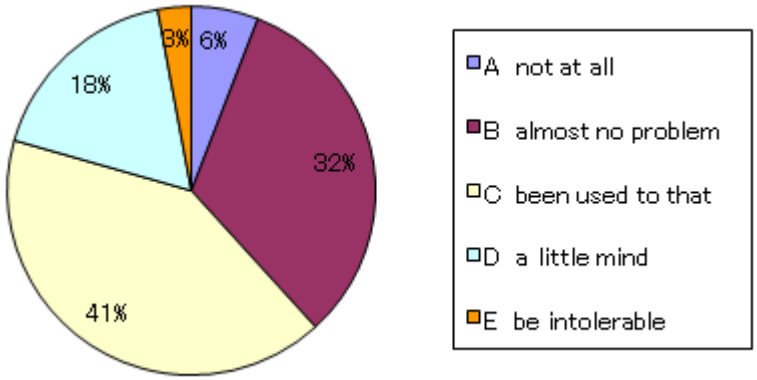


Figure 2-16 Result of evaluation of TC (4)

5. How long have you made use of Tutorial Chinese Platform when you preview and review?

[A] more than 1.5 hour

[B] 1 hour~1.5 hour

[C] 0.5 hour~1 hour

[D] less than 0.5 hour

[E] do not use at all

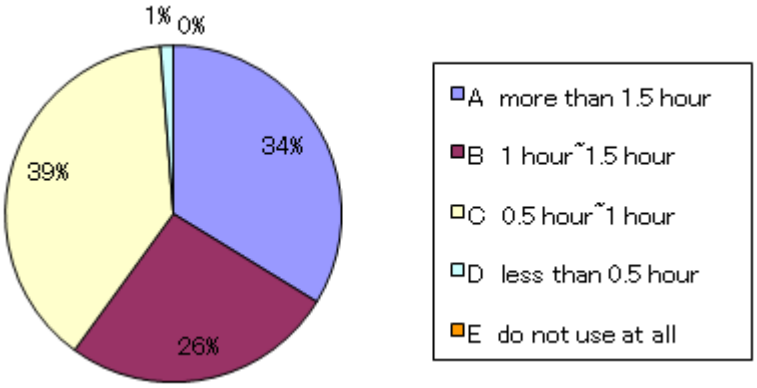


Figure 2-17 Result of evaluation of TC (5)

6. Do you feel confident of your Chinese language ability through Tutorial Chinese course?

[A] very confident

[B] some confident

[C] no change

[D] a little diffident

[E] very diffident

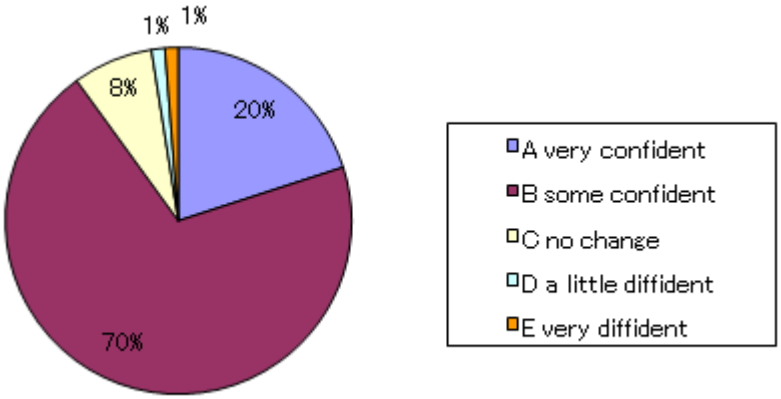


Figure 2-18 Result of evaluation of TC (6)

About 99% (Q5) of students have spent more than half an hour to make use of the TCP platform to preview or review a class. With respect to the time delay of video conferencing, 21% (Q4) of students have been not so satisfied, which means that we should improve this part of the program in the next step. When asked about the goal of the education model (to enhance communication ability), 66% (Q1) responded that they think they have enough opportunities to speak Chinese during class. Among the students, 90% (Q2) confirm that the discourse themes are appropriate for improving their communication ability. Further, 61% (Q3) are satisfied with how the level of the current class corresponds to their language ability, while 90% feel confident of their Chinese language ability.

The results of the questionnaire show that students are generally satisfied with this new language education model. Through the course, most students have received sufficient opportunities to improve their communication ability and feel confident of their Chinese language ability.

CHAPTER 3.
CALL DRILL SYSTEM

3.1 INTRODUCTION

The education in Tutorial Chinese is based on the concept that to study a second language in the environment of this target language. The education in TC is conducted in conversational style and all grammar and vocabulary are integrated around topics of discourse. In this kind of educational environment, the progressing of collaborative live lesson among students and native speaking tutors cannot be fully predicted and the teaching of vocabulary only based on static textbook as normal language education is no longer enough. Many vocabulary and grammar items appeared in classroom are far beyond the contents of static textbook and these items are valuable to be utilized in pedagogy. In this research, a mechanism that collecting this information through tutor lesson report module of the TCP has been constructed. Every week, tutors will access the TCP immediately after class and report pedagogical information to teacher through the interface shown in figure 2-9. There are three types of information are included in lesson reports.

- Information about lesson itself, such as class name, date, period, type of class (distance or face-to-face), textbook lesson number, tutor's information, etc.
- Educational information generated in classroom, such as the new words appeared, hot topics in classroom, common errors and advices to this class students, comments to hardware environment, etc.
- Evaluation information to individual student, such as the attendance, status of preview, status of speech, grammar, vocabulary, and advice to this student, etc.

The information gathered in Lesson_New_Words field, after analyzed by teacher,

will enhance vocabulary corpus in WTCC and be sent to CALL drill authoring module to make exercises.

Table 3-1 and Table 3-2 show the difference between vocabulary used in actual classrooms of TC and the vocabulary of static textbook. The conversation topic of this week is “accident and disaster” (In Chinese is “出事”), which is lesson 15 in textbook. Just before that week, there was heavy rainfall due to typhoon and earthquake just occurred in Nigata. Meanwhile, in Taiwan and China mainland flood damage was just reported. Therefore the classes were focus on the topic of disaster and theme-based vocabulary was prominent [Sunaoka, 2005].

Table 3-1 Vocabulary in lesson 15 of textbook

安排	从	会	美国	塞车	通知	因为
安全	打	火灾	迷路	三级	统计	银行
安全帽	打算	给予	灭火	伤亡	突然	应该
把	大	技术	灭火器	上学	推销员	有
办	大家	交通工具	摩托车	生命	腿	遇到
保护	大学	交通警察	哪里	什么	脱离	原来
报道	逮捕	交通事故	那里	声音	妥善	原因
报警	到底	结果	哪儿	失事	外面	援助
爆炸	的	今年	呢	时候	危险	院子
被	得到	警察局	你	事	问题	在
比较	地点	救火	浓烟	是	我	造成
不是	地铁	卡车	朋友	事故	我们	怎么
不要	地震	考试	平时	事件	现场	着火
不要紧	都	可怕	破坏	事情	相信	找
布告栏	堵车	可以	骑	受害人	消防队员	这里
不可	对	恐怖	起火	受伤	消息	政府
不少	对方	恐怖主义分子	企业家	受害者	小偷	知道
财产	多	快	汽车	树	小心	重地
仓库	而	那么	钱包	数字	新生	重要
曾经	发生	垃圾	前面	水灾	信用卡	住宿
常	飞机	里面	清楚	说	幸好	注意
常常	服务生	立刻	情况	死亡	学生	撞车

长江流域	刚才	练习	请	损失	学校	状态
超过	给	乱	请问	所以	严重	自行车
车祸	根据	轮船	确定	他	要求	最
车子	工作	麻烦	人民	她	要	最近
出	官方	马虎	人数	台风	要是	做
出来	关于	马上	人员	特别	一定	作弊
出事	还	满意	任务	天干物燥	以后	
出现	很	没	日本	天灾	已经	
处理	后来	没有	如果	听说	一点儿	

Table 3-2　Vocabulary generated from lesson reports

Class Name	Tutor	Level	Vocabulary
F-F R	A	Elementary	滑，瘫痪，直升飞机，劫机，海啸
F-F L	B	Elementary	干燥，燃烧，倒塌，道路，预告，根据，震灾，备用，坠毁，空难，裂开，情味，扒手，酷暑，出海，洪水，敞篷车，面包车，小卧车，小货车，猜测，波浪，海啸，劫机
F-F E	C	Middle	摇，倒，窄，行人，斑马线，流氓，强盗，扒手，犯罪，医药费
F-F F	D	Middle	摇醒，流血，墙壁，缝（了五）针，瓦斯，躲到，手枪，派出所，电脑中毒，凶
F-F G	E	Middle	爬到树上，强盗，一艘轮船，倒塌（房屋倒塌了），劫持，乘客，燃烧（火燃烧起来了），报警，架（一架飞机），～级（日本发生了 6 级地震），人质（把乘客作为人质），救命啊！，损失，察觉（他察觉到有人在偷他的钱包）
F-F K	F	Middle	洪水，劫持，告示牌，人质，失火，空难，生命危险，遇难
F-F N	G	Middle	楼房，倒了，倒塌，裂了，地裂，跑，逃，体育，男的，女的，冲走了，防震棚，卧轨，卧室，飞行员，枪，恐怖分子，吵嘴，打架，爬，爬树，掉下来，飞机失事，坠落
F-F M	H	Advanced	坠毁，劫机，空难，规划，海啸，扒手，天干物燥，天灾人祸，趁火打劫，雪上加霜，雪中送炭，倾斜，倾倒，便道，非机动车

F-F A	I	Advanced	倒塌，冲走，洪涝灾害，天灾人祸，逃生，吃一堑长一智，私了，黑车，瞌睡
F-F Q	J	Advanced	统计，妥善，搜索，损失，数字，信号，逮捕，援助，地铁，一连串，一则，老年人，赔偿，无法，协商，解决，案件，讹诈
D(Taipei) B	K	Elementary	地震，發生，出事，警察，立刻，安全，危險，失火，失事，救火，滅火，防火，受災，人數，統計，馬虎，爆炸，恐怖，小偷
D(Taipei) C	L	Middle	搖晃，電綫杆，洪水，意外，消防隊員，勇敢，技術，水桶
D(Taipei) D	M	Advanced	地震，火災，放鞭炮，原來，失火，水災
D(Beijing) G	N	Elementary	地震，发生，出事，警察，立刻，安全，危险，妥善，失事，失火，救火，灭火，受灾，人数，统计，水灾，丢，马虎，注意，爆炸，恐怖，出现，小偷，车祸，处理，事故，骑，撞，曾经
D(Beijing) L	O	Elementary	曾经，手表，硕士，研究生，利害，体育馆，藏，跟，一样

The vocabulary generated in each class is distributed, basically around topic of conversation but also extended. This kind of extension is according to tutor and students' logical thinking in classroom, therefore it will be very easy to be remembered if there are some functions to help students recall those words timely. In this research, the mechanism has been constructed through CALL drill system. These kinds of live educational information have been collected into corpus named Waseda Tutorial Chinese Corpus (WTCC) and utilized to make queries of drill.

3.2 SYSTEM CONSTRUCTION

CALL drill system has corpus-based, flexible CALL-drill authoring tool for teachers and ubiquitous language drill environment for students. It has been designed with the following features [Liu et al., 2004: a]:

- Web based interface and transmission mechanisms
- Integrated Waseda Tutorial Chinese Corpus (WTCC)
- Mechanism to save word corpus, problems and exercise papers in different layers, to make the authoring tool more flexible and reusable
- Ubiquitous language drill environment through mobile telephone technology

There are four function modules designed in CALL drill system: corpus management, drill queries authoring, exercise paper authoring and drill & administration interface. A mechanism has been constructed that words corpus, questions and exercise papers in different layers. Therefore, when teachers want to construct an exercise paper, they can construct new questions or just select questions that already exist. The same question can be utilized repeatedly in different papers to save time and enhance review. Based on this mechanism, the authoring tool is more flexible and reusable.

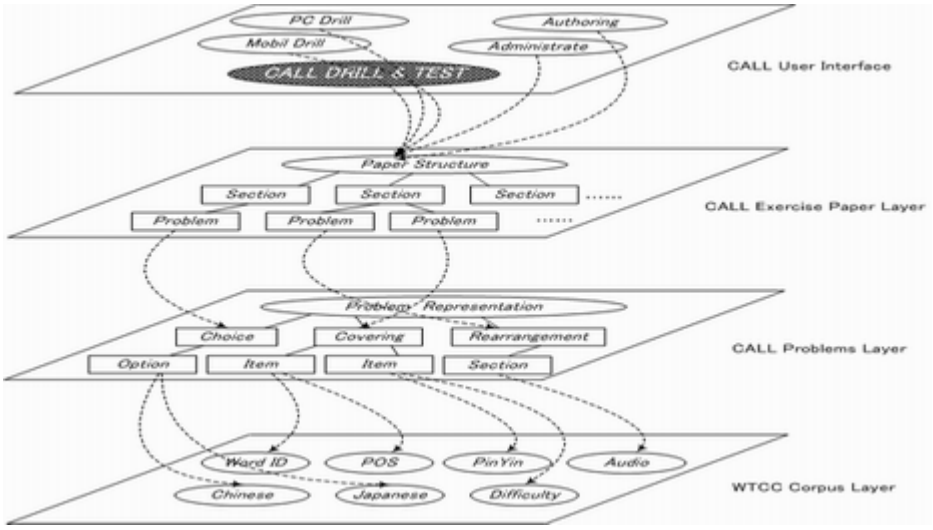


Figure 3-1 CALL drill layer mechanism

Figure 3-2 shows the dynamic working cycle that combines CALL drill to TC classroom pedagogy activities.

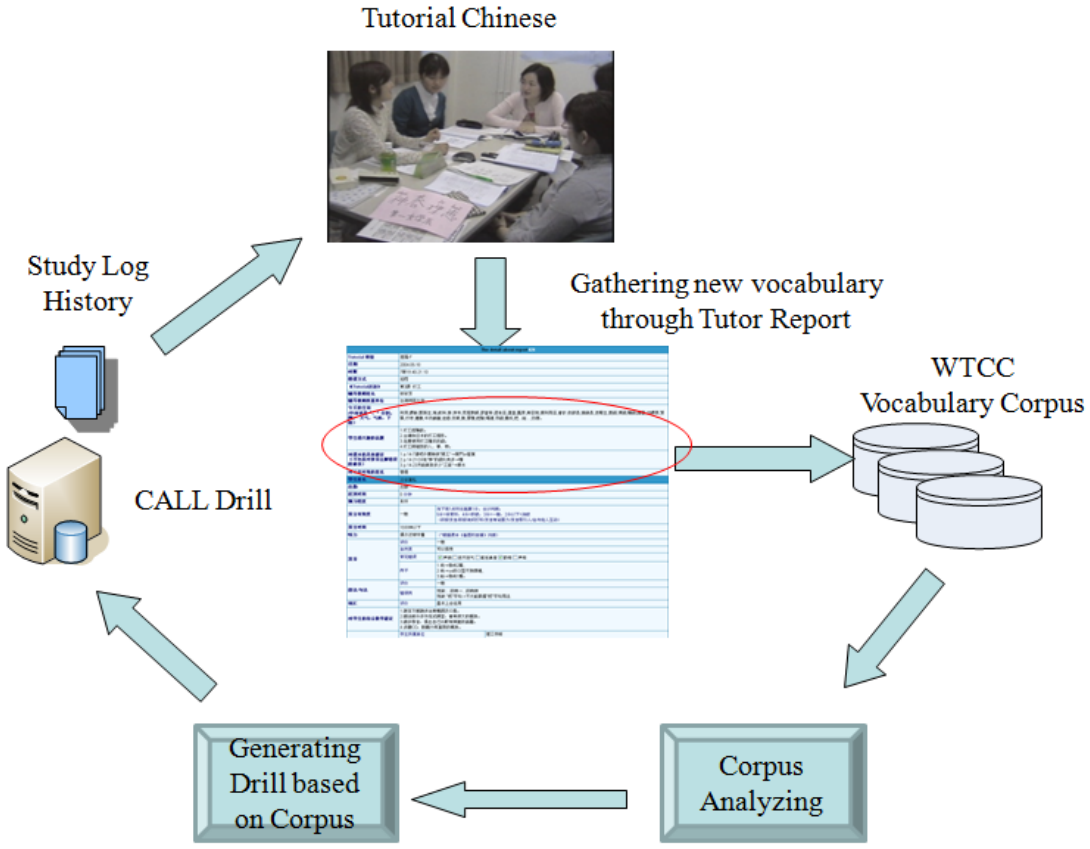


Figure 3-2 Working cycle of CALL drill system

Firstly, the new vocabulary is gathered from tutor lesson reports. This operation is split into the following steps:

(1) Split new vocabulary field of TC report into words and morpheme by word segmentation parser.

(2) Extract POS, Pinyin spell and difficulty level information of each word by searching the existed dictionary such as Beijing University Corpus and HSK corpus.

(3) Select the target word from polyphonic or multi-meaning words based on the TC textbook.

(4) Find Japanese meaning of words using auto translation tools and manual confirmation.

(5) Add other meta information such as textbook lesson number, date, etc.

(6) Construct vocabulary corpus into WTCC.

The vocabulary corpus of WTCC has the attributes as follows:

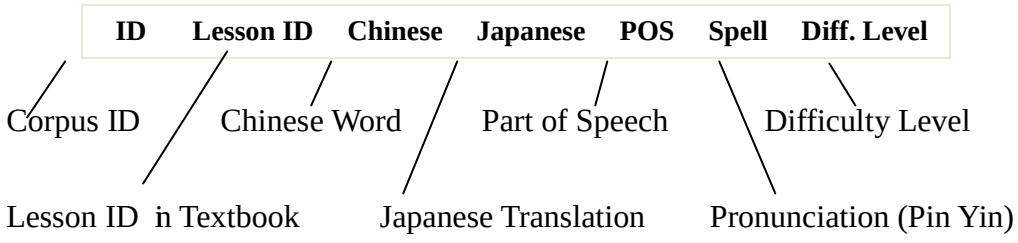
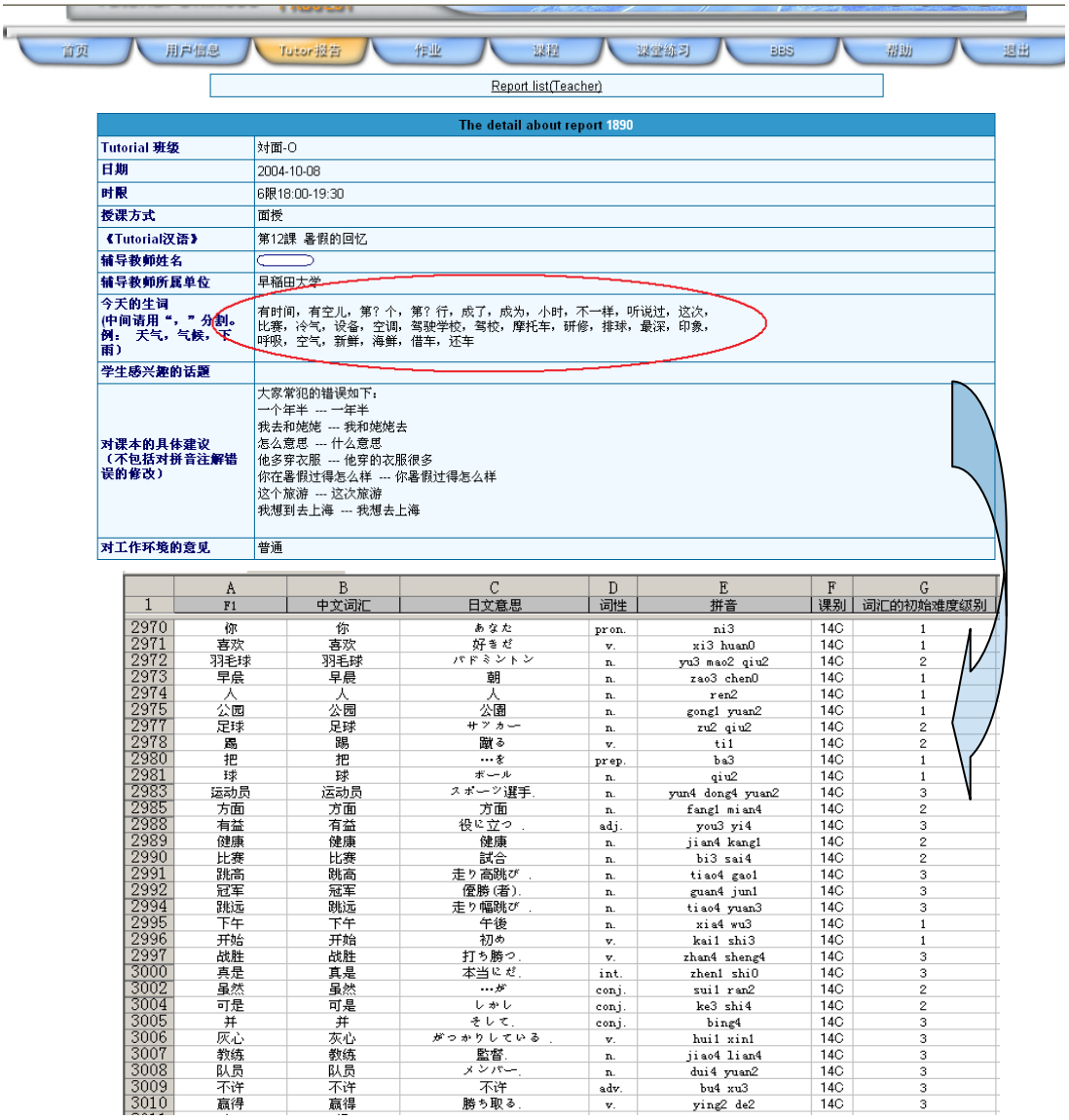


Figure 3-3 Attribute of vocabulary corpus in WTCC



	A	B	C	D	E	F	G
1	F1	中文词汇	日文意思	词性	拼音	课别	词汇的初始难度级别
2970	你	你	あなた	pron.	ni3	14C	1
2971	喜欢	喜欢	好きだ	v.	xi3 huan0	14C	1
2972	羽毛球	羽毛球	バドミントン	n.	yu3 mao2 qiu2	14C	2
2973	早晨	早晨	朝	n.	zao3 chen0	14C	1
2974	人	人	人	n.	ren2	14C	1
2975	公园	公园	公園	n.	gong1 yuan2	14C	1
2977	足球	足球	サッカー	n.	zu2 qiu2	14C	2
2978	踢	踢	蹴る	v.	ti1	14C	2
2980	把	把	…を	prep.	ba3	14C	1
2981	球	球	ボール	n.	qiu2	14C	1
2983	运动员	运动员	スポーツ選手.	n.	yun4 dong4 yuan2	14C	3
2985	方面	方面	方面	n.	fang1 mian4	14C	2
2988	有益	有益	役に立つ .	adj.	you3 yi4	14C	3
2989	健康	健康	健康	n.	jian4 kang1	14C	2
2990	比赛	比赛	試合	n.	bi3 sai4	14C	2
2991	跳高	跳高	走り高跳び .	n.	tiao4 gao1	14C	3
2992	冠军	冠军	優勝(者).	n.	guan4 jun1	14C	3
2994	跳远	跳远	走り幅跳び .	n.	tiao4 yuan3	14C	3
2995	下午	下午	午後	n.	xia4 wu3	14C	1
2996	开始	开始	初め	v.	kai1 shi3	14C	1
2997	战胜	战胜	打ち勝つ.	v.	zhan4 sheng4	14C	3
3000	真是	真是	本当にだ.	int.	zhen1 shi0	14C	3
3002	虽然	虽然	…が	conj.	sui1 ran2	14C	2
3004	可是	可是	しかし	conj.	ke3 shi4	14C	2
3005	并	并	そして.	conj.	bing4	14C	3
3006	灰心	灰心	がっかりしている .	v.	hui1 xin1	14C	3
3007	教练	教练	監督.	n.	jiao4 lian4	14C	3
3008	队员	队员	メンバー.	n.	dui4 yuan2	14C	3
3009	不许	不许	不许	adv.	bu4 xu3	14C	3
3010	赢得	赢得	勝ち取る.	v.	ying2 de2	14C	3

Figure 3-4 Construction of WTCC vocabulary corpus

Then the vocabulary corpus is sent to corpus management module of CALL drill. Based on the frequency of occurrence and existing situation in HSK word list, the vocabulary will be sorted and listed to teacher for analyzing. After this step, the selected words will be saved into drill queries authoring module for generating drill queries. System has implemented three types of drill questions: choice question, filling question and rearrangement question. The flow of question authoring is as follows.

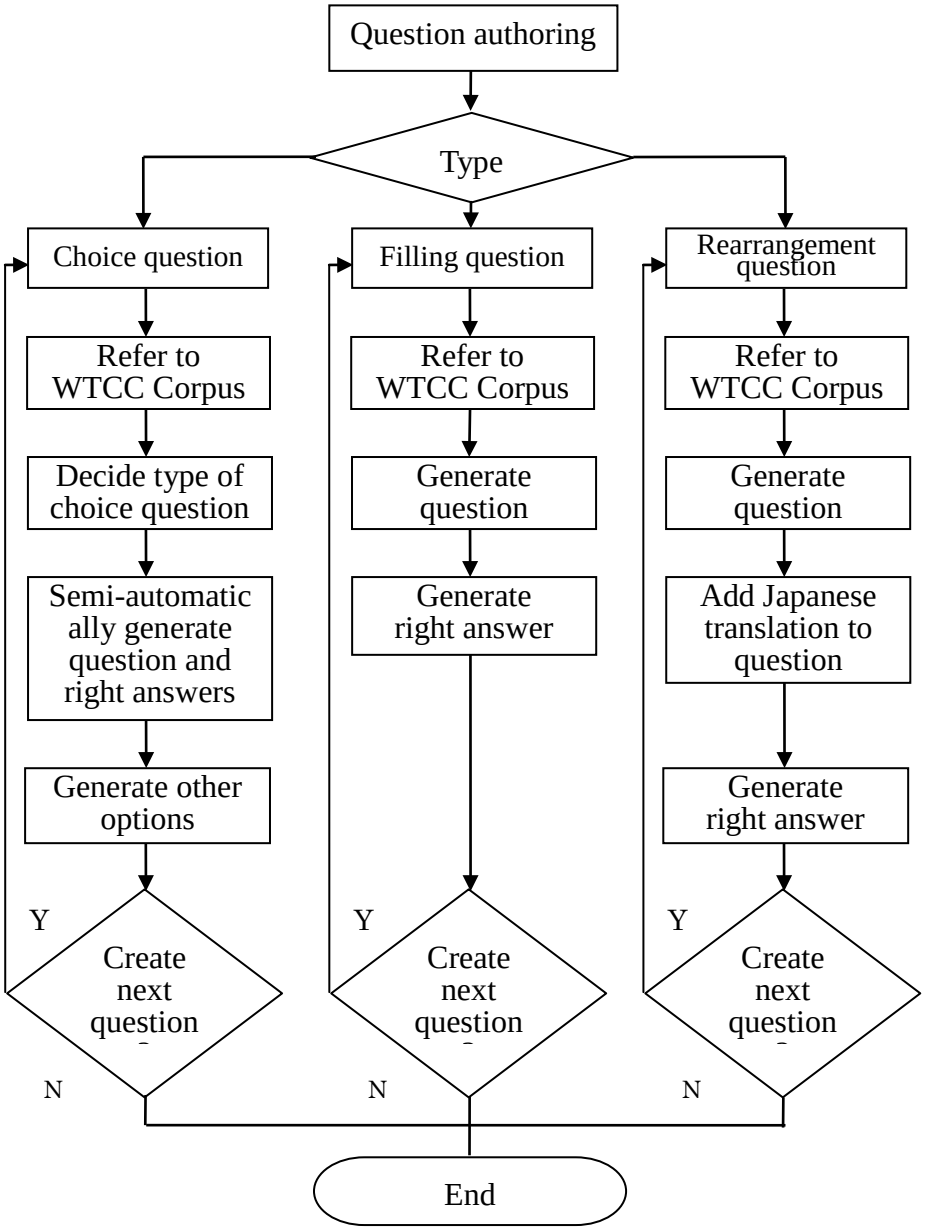


Figure 3-5 CALL drill question authoring flow chart

Currently the choice question is semi-automatically created with vocabulary corpus. There are three types of choice question based on different aspects of words.

- “Select its Japanese meaning according to Chinese word”
  (e.g. 签证 A. ビザ B　手続き　C パスポート)
- “Select its pronunciation (Pinyin) according to Chinese word”
  (e.g. 选举 A.　xuan3ju3　　B　suan3ju3　　C　xian3qu3)
- “Select its Japanese meaning according to pronunciation (Pinyin)”
  (e.g. xuan3ju3　A　选举 B 仙居 C 选取)

Teacher just needs to select appropriate words from vocabulary corpus for this lesson and the specific type of question, then the query, right answer and other options will automatically filled by system. System will select the dummy options from the other words with same difficulty level and in the same textbook lesson, because these words are in the same conversation topic and with relatively close logic distance.

The filling question is also created with the support of vocabulary corpus. Firstly, teacher makes query sentence based on appropriate words of vocabulary corpus. Then word segmentation parser of system will split the sentence into sequence of words and add automatically query space instead of the selected appropriate vocabulary (e.g. selected vocabulary is “如果...的话”, the query sentence generated will be “（ ）明天下雨（ ），比赛就中止”. Students should input the vocabulary into those spaces).

When teacher creates rearrangement questions, the query sentence, respective parts of sentence need be rearranged and the right answers of sequence need be input by hand (e.g. query sentence is　“我投民主党的票”; the sentence will be separated into sector1: 我, sector2:投, sector3:民主党, sector4:的, sector5:票; the right answers of sequence will be both 12345 and 12543).

After generating drill queries based on WTCC corpus, all the queries will be stored into question database and system will make drill for students. Currently this system implemented ubiquitous drill environment for both computer and mobile phone. Because most mobile phones in Japan do not support full Unicode, simplified Chinese character, traditional Chinese character and Pinyin character cannot be displayed in Japanese mobile phone. It needs to construct a mechanism of changing font of different languages into graph and displaying in mobile phone. In this system, GD (Graphics Draw) Engine has been utilized to change fonts to graph file. The scheme is shown

below [Liu et al., 2005c], [Zhang et al., 2004]. After image compression, 16 X 16 pixel monochrome character (normal character size displayed in mobile phone) will be shrank to about 40-60 bytes, approximately 1/8-1/10 of bitmap size. Relative to transmission rate of today's mobile phone (normally 2.1 MB/Sec), displaying of this kind of graph file will not cause any inconvenience. Meanwhile, many other characters besides Chinese can also be displayed in this mechanism therefore it is possible to be widely applied to other foreign language education.

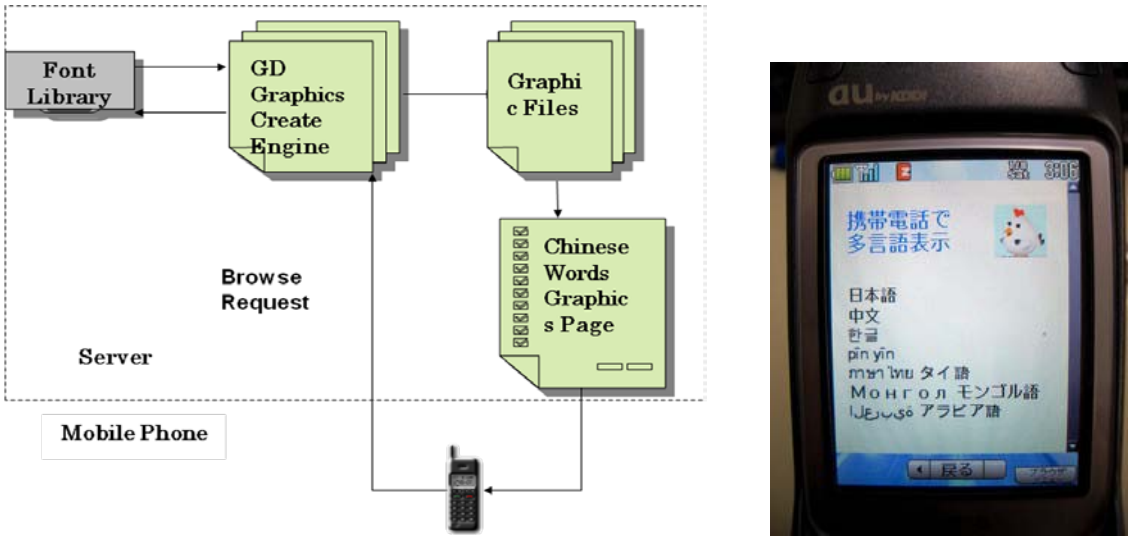


Figure 3-6 Scheme of multi-language font display in mobile phone

3.3 SYSTEM IMPLEMENTATION

CALL drill system was proposed and constructed in May 2004, operated as collaboration with TCP since September 2004 for over two hundred students. In the end of 2004, comparison was done between students with normal classroom Chinese language education and students who took the course of TC and practiced with CALL drill through the result of Test of Ability for Chinese Communication (TACC) [Sunaoka, 2005] to measure their ability of vocabulary. All the monitors were divided into three different groups; C I represented those students that had taken normal classroom Chinese language course for one year, C II represented those students with two year of normal classroom Chinese language education, TC represented those students that had taken TC elementary course for one year. The learning time of every course was 120 hours. As the result, in 100 point scale, the average score of C I was 41.17, C II was 49.5 and TC was 76.02. Compare to normal classroom Chinese language education, the model of TC and CALL drill has significantly increased students' vocabulary ability. The following figures show screenshots of drill queries authoring and drill interface.

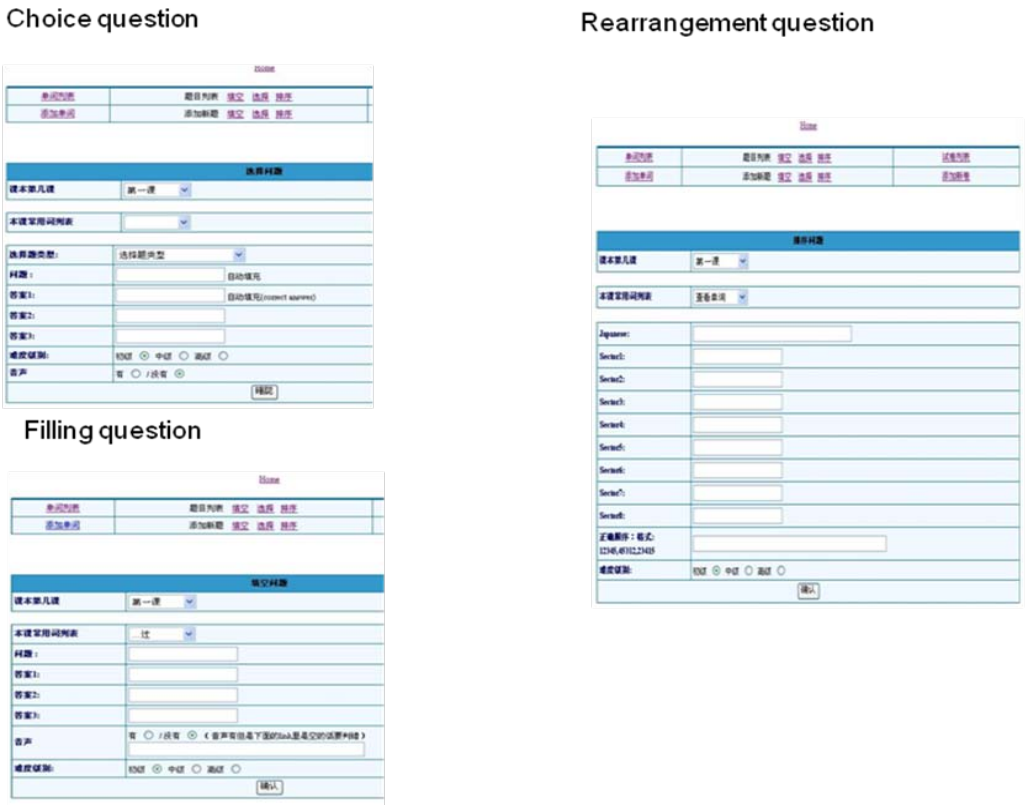


Figure 3-7 Screenshot of drill queries authoring

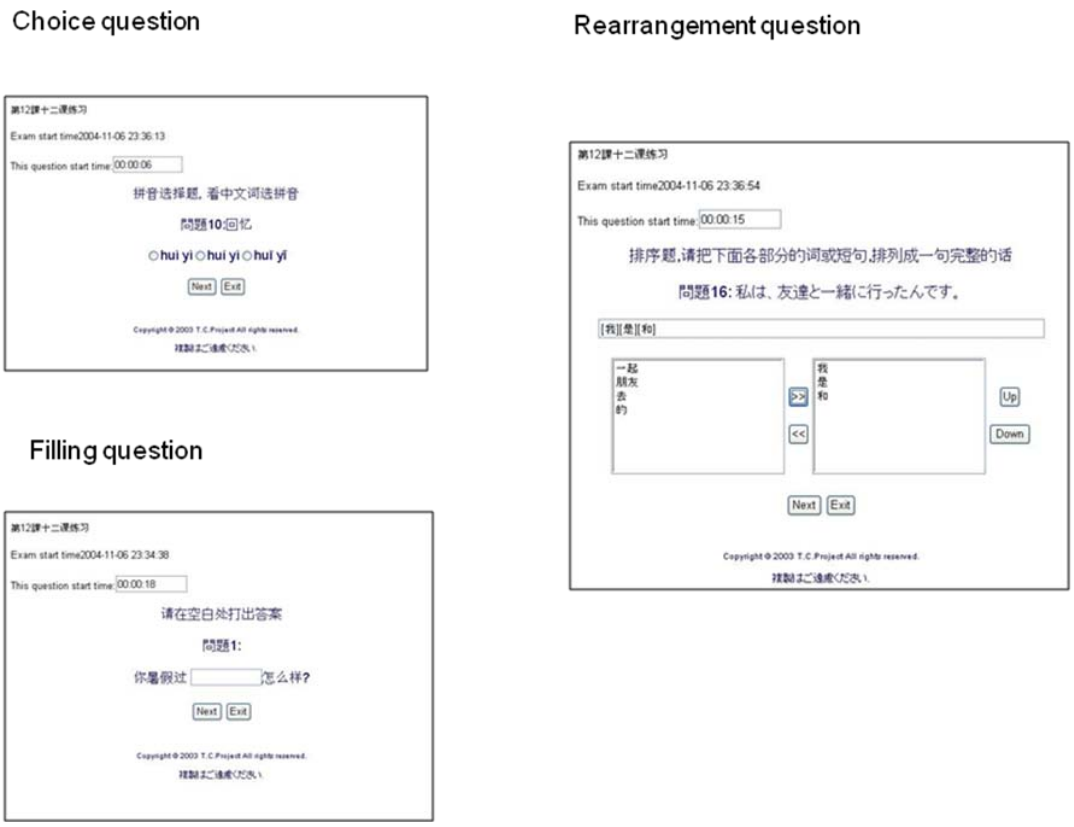


Figure 3-8 Screenshot of drill on computer

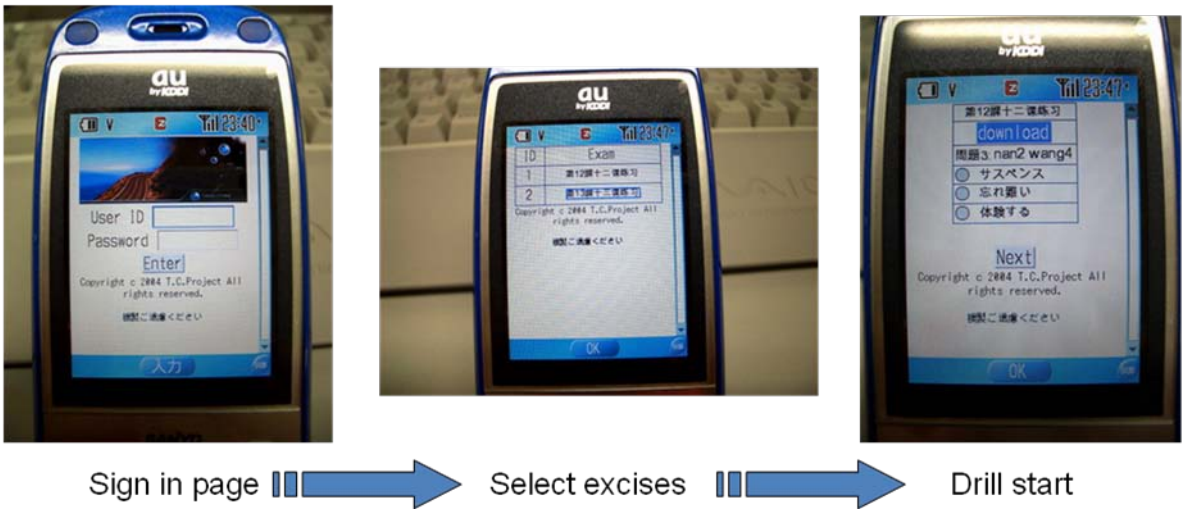


Figure 3-9 Screenshot of drill on mobile phone

3.4 CONCLUSION

In TC conversation-style language education, many components of language such as new vocabulary and sentences taught in the classroom go far beyond the contents of a static textbook and are valuable to be utilized in pedagogy. CALL drill system has been constructed to collect new vocabulary into WTCC corpus through tutor lesson report module of the TCP. The mechanism of corpus analyzing and semi-automatically drill queries creating has also been implemented. Meanwhile, this system has many advantages which distinguish it from conventional software. Among these advantages are: (a) web-based user interface for students and teachers to facilitate system access, (b) integrated with WTCC corpus to make drills more effective, (c) layer architecture to make authoring flexible, and (d) mobile phone-based ubiquitous environment to make the drills accessible to users anytime and anywhere. In this proposed mechanism, teachers, tutors, and students are involved in the process of generating, gathering, analyzing, and utilizing of language educational information generated in classroom. This process works as a dynamic cycle with valuable vocabulary corpus being stored and applied.

In the next step, text-style corpus generated from composition/correction system described in Chapter 4 will also be embedded into system and utilize the sentences with common errors to create filling questions and rearrangement questions.

CHAPTER 4. COMPOSITION/CORRECTION SYSTEM WITH CORPUS RETRIEVAL FUNCTION

4.1 INTRODUCTION

Practice and research in the composition education using computer and network have been more and more active. While for the normal correction functions in many composition systems, student's learning stage is usually not considered, and there is not enough correction tools supplied to teachers corresponding to a variety of misuse. The composition and correction information has often been accumulated respectively and not easily been retrieved. Despite great effort of teachers, the feedback effect of correction is not satisfied enough. More serious problem is that when the number of students becoming larger and larger, variation of evaluation and scoring will be caused. The education is lacking in stability and objectivity [Urasaki and kogo, 1998] [Kawai, 1996].

A web-based composition/correction system has been constructed as part of Tutorial Chinese Platform (TCP) based on the concept of collection and application of corpus. The system will not only suitable for learners in writing composition, but also support teachers in correcting and making comments on composition, as well as in retrieving and detecting learners' common problems. Moreover, the composition from students, together with the correction information from teachers, has been structured and constructed automatically into learner corpus, and the retrieval function has also been developed. This is also useful for researchers to analyze language resource from the real educational environment.

This web based Chinese language composition/correction system has been carried out in Waseda University since 2003. By utilizing the standardized “Grammatical Correction Mark”, previous problems such as non-standardization of evaluation criterion and non-standardization of correction description have been solved. Meanwhile, the working efficiency of correction has been dramatically increased. In this chapter, the development issues, usability evaluation and misuse status analyzed from learner corpus will be described.

4.2 CONCEPT AND FEATURES OF SYSTEM

The main concept of this web based composition/correction system is taking advantage of Internet, implementing online composition correction functions similar to paper-based working environment, and through the correction process, simultaneously storing the compositions and correction information into learner corpus.

This system takes requirements of teachers, students and researchers in to consideration. Before constructing system, the requirements from teachers, students and researchers have been investigated. For teachers, they need convenient way to manage all the compositions, easily usable correction tools preferably the same as working on paper, various retrieval methods for composition and correction information with different classification ways, such as long term information for an individual student, the group of students who have written compositions with same topic, or all students who have made the same kind of errors. Teachers also need to remove duplication of works involved in correcting on common misuses. For students, they need timely feedback from teacher and comprehensible comment on their writing. Students also need methods to discuss with teacher and other peers about their composition. For researchers, they need convenient method to access large amounts of learner corpus and retrieval tools for analyzing these data.

In order to answer these requirements, composition/correction system has been designed with the following features.

1. Convenient online correction tools

Teachers are familiar to correction works on paper. In order to reduce the discomfort of online operation, this system has supplied template-based coding rich-text tools, such as underline, strikethrough, bold mark, etc. With the help of these tools, paper based operation such as insert, replace, delete and comment can also be fulfilled online.

2. Unifying of misuse mark

In the correction of free composition, misuse will range from superficial errors such as spelling miss, or wrong use of vocabulary, to ambiguous errors concerning to concept or ideas. If not basing on some kind of correction standard, education effect will

not be obvious despite the great effort. In this research, the misuses have been classified into several error types referring to statistical result and utilized “grammatical correction mark” to label errors with standardized criteria. These unified misuse information and comments will be automatically granted XML-style tags and stored into learner corpus.

3. To stimulate the enthusiasm by setting up appropriate topics and deadline

An example of one composition topic is shown below. The system is working together with lesson report module of the TCP. When assigning composition topics for each lesson, teachers will first refer to lesson reports to get the important or interested topics. Then teachers from Beijing and Taiwan will decide two topics for elementary and advanced level respectively, as well as preferred sentence patterns, grammar points and keywords. These efforts will limit the free composition to some extent and enhance the effect of correction.

Figure 4-1 Example of composition topic

Because there are two topics for one lesson for elementary and advanced level respectively from teachers of Beijing and Taiwan, the students can select one of them freely based on their ability and cultural differences. This flexibility will stimulate their enthusiasm for learning. Moreover, system will decide the submission deadline based on date of lesson and date of assigning homework. The educational benefits can be expected to maintain the tension of students.

4. Listing up compositions by urgency to reduce the time lag of feedback

Students always look forward to timely feedback from teacher after they submit composition. In order to maintain students’ learning enthusiasm, the submitted

compositions will be listed up to teacher by the sequence of submitting time. The higher position a composition is in the list, the urgency is higher.

5. Praise mechanism to promote the reconfirmation of comment

In order to promote students to reconfirm the comment on their compositions, praise mechanism has been designed that excellent compositions with high score will be public in excellent composition column to all students. Students can reconfirm compositions of themselves through the comparing with excellent composition.

6. Discussion function to support collaborative learning

Besides studying by themselves, students also need method to discuss with teachers or other peers about their composition. Discussion spaces have been designed for each composition topic to the unit of class to support collaborative learning.

7. Automatic storage and retrieval function for learner corpus

Researchers need convenient method to access large amounts of learner corpus and retrieval tools for analyzing the data. The composition and correction information can be automatically saved as corpus through the process of correction. Retrieval editor has also been designed to analyze the information of both composition corpus and other learners' profile information stored in TC_User table.

4.3 SYSTEM CONSTRUCTION

The user of this composition/correction system can be classified into three types.

[Student] will utilize system to write, modify and submit composition. The possible operation of students include: (1) Write new composition or edit composition that still not be corrected. (2) Submit compositions to teacher. (3) Browse and check their compositions. (4) Check the public excellent compositions. (5) Confirm the comment on their compositions. (6) Search for all correction information about their compositions (7) Post in the discussion space.

[Teacher] will utilize system to make correction. The possible operation of teachers include: (1) Assign the composition topics (2) Make comment on composition of the class in their charge. (3) Search students' composition and correction information. (4)Check students' other educational information in corpus. (5) Post in the discussion space.

[Researcher] will use retrieval and analyzing tools to access learner corpus and do statistical analysis.

The workflow of composition/correction system in Tutorial Chinese education process is as follows.

1. Assign topics: Teachers from Taiwan and Beijing create two topics per-lesson respectively for elementary and advanced level. The topics can be chosen freely by students based on their favorite culture and language skill.

2. Write composition: Students write compositions with online editor tools.

3. Submit composition: The submitted compositions will be stored into source data corpus. Simultaneously, the compositions will be shown in the list of waiting correction to teachers and students.

4. Make correction: Teachers make comment and correction on compositions with help of online editor tools. The correction marks, together with other commenting marks will be transferred into XML tags automatically.

5. Submit correction result: The submitted correction information will be stored

into correction data corpus. Simultaneously, the compositions will be shown in the list of finished correction to both teachers and students.

6. Reconfirm correction: Students will reconfirm the result of correction and check their misuses. If they have questions, the questions can be posted into discussion space and discussed with other students or teachers.

7. Analyze the studying status: Teacher will analyze the entire status of students to get reference data for next step education.

4.3.1 CONSTRUCTION OF SYSTEM FUNCTION

The structure of system functions is shown as the following graph. There are mainly four modules divided based on different functions: Browse (displaying function of compositions in list); Edit/Correct (editing and correcting function); Retrieve (searching and analyzing function); Discuss (posting and viewing threads in discussion space)

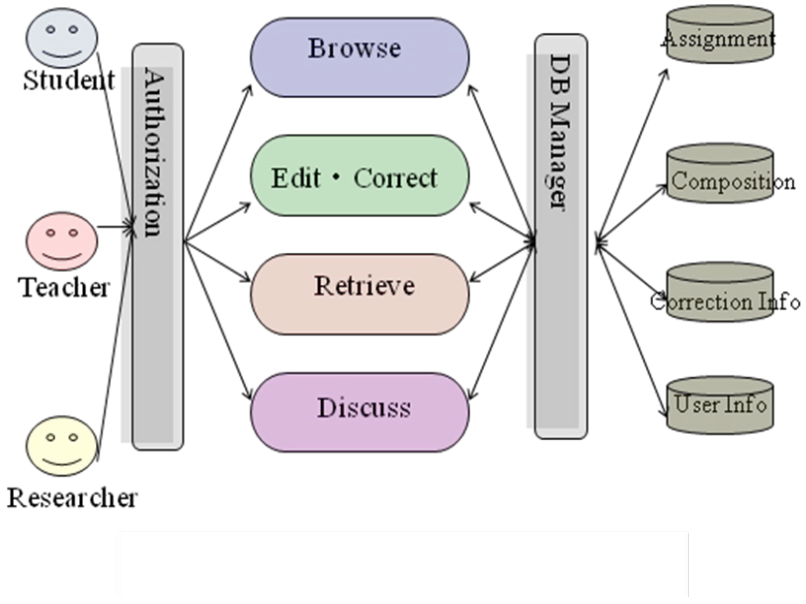


Figure 4-2 Composition/correction system functions' structure

1. Browse (displaying function of compositions in list)

The module of browse is shown as figures below. There are three lists in figure. The list in top part is for excellent compositions with high score (more than 8 points, 10 points is perfect score), after removing personal information. The middle list is for compositions waiting for correction and the list in bottom is for compositions finished correction. The excellent compositions' list is public for all users. In other two lists, teachers can browse all compositions of students in their class; students can only view their own writings. All compositions are listed up by urgency of deadline to maintain the tension of both students and teachers. Through this module, students can not only review correction information of their own compositions, but also compare with excellent compositions to promote active learning by themselves.

优秀作业

Text	Level	Title	Point	Edit
第22課	初級	我最爱看的新聞類型	8	view
第22課	初級	日本的媒体现状	8	view
第22課	中高級	新聞媒體的社會責任	8	view
第22課	初級	我最爱看的新聞類型	8	view
第22課	中高級	新聞媒體的社會責任	8	view

等待评分作业

Student	Text	Level	Title	Teacher	Submit Time	Edit
	第18課	中高級	我了解的医疗保险制度		2005-02-10 20:18:38	view
	第22課	初級	日本的媒体现状		2005-02-04 17:41:53	view
	第21課	初級	我对战争的看法		2005-02-04 17:35:28	view
	第20課	初級	关于"年金"改革的看法		2005-02-04 17:25:19	view
	第19課	初級	我关心的政治问题		2005-02-04 17:19:58	view
	第19課	初級	我关心的政治问题		2005-01-24 15:34:12	view

已评分作业

Student	Text	Level	Title	Remark by	Remark Time	Edit
	第22課	初級	我最爱看的新聞類型		2005-01-13 08:44:20	view
	第22課	初級	日本的媒体现状		2005-01-22 13:12:13	view
	第22課	中高級	新聞媒體的社會責任		2005-01-16 10:20:45	view
	第22課	初級	我最爱看的新聞類型		2005-01-13 22:29:47	view
	第22課	中高級	新闻媒体的可信度		2005-01-24 01:00:49	view

Figure 4-3 Interface of browse module

2. Edit/Correct (editing and correcting function)

The module of edit/correct is shown as figures below. In this module, template-based coding rich-text tools, such as underline, strikethrough, bold mark, etc, have been constructed to make teachers feel convenient in online operation. With the help of these tools, paper based operation such as insert, replace, delete and comment

can also be fulfilled online. All the comments and correction information will be transferred into XML tags automatically and be stored into correction data corpus after submission.

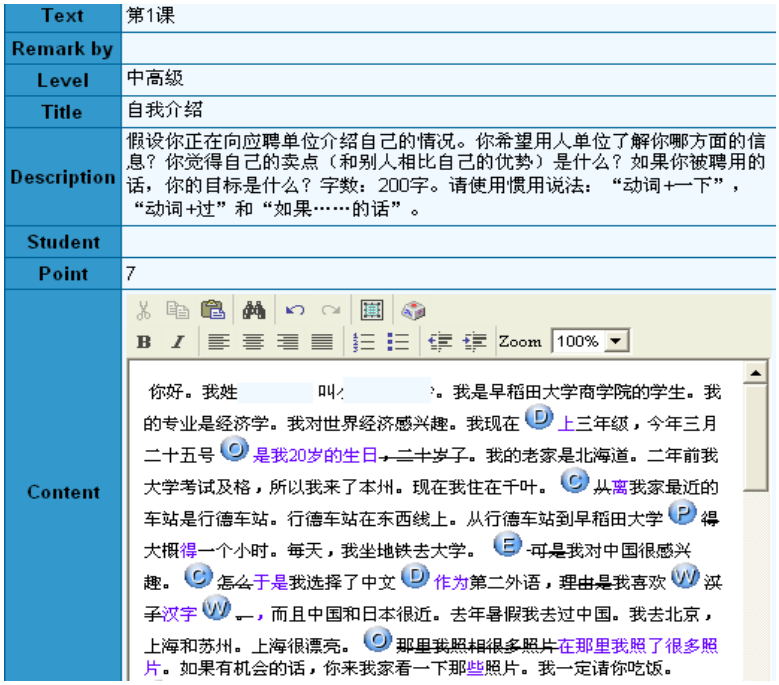


Figure 4-4 Interface of edit/correct module

3. Retrieve (searching and analyzing function)

In traditional paper based correction, it is extremely difficult for teachers to make analysis of correction information or track individual student's typical errors. With the module of retrieve, limitations have been broken. This module is shown as figures below. All the comments and correction information have already been transferred into XML tags and stored into correction data corpus. The combination of correction data corpus and learners' profile information stored in TC_User table can be extracted and listed as search results in the unit of sentence.

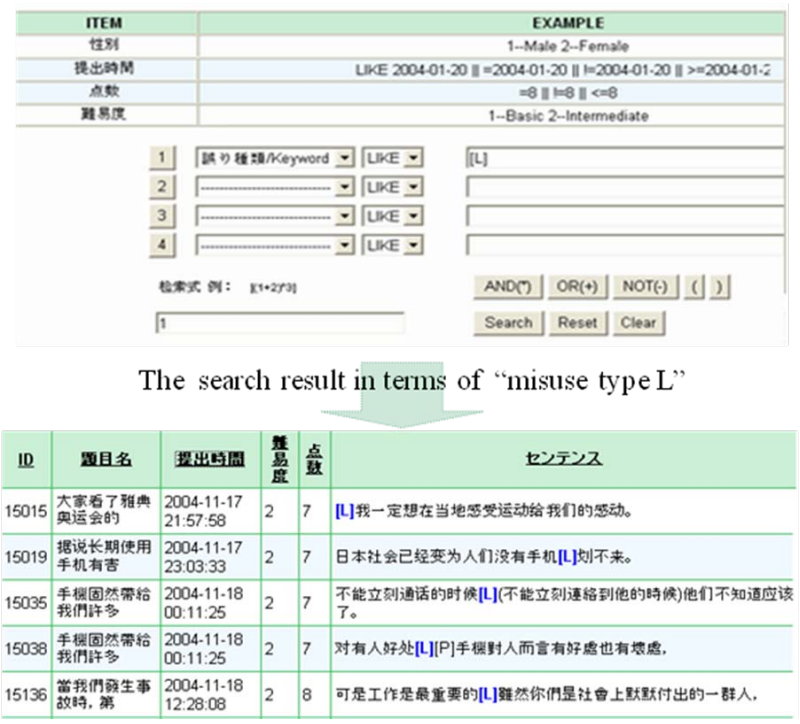


Figure 4-5 Interface of retrieve module

In detail, the components about composition can be retrieved include: <error types>, <keywords in composition>, <assignment titles>, <difficulty of assignment>, <score of composition>, <submission date>, and <correction date>. The components about learner profile can be retrieved include: <students ID>, <name>, <gender>, <age>, <department in university>, <grade in university>, and <class>. The retrieval operators include: "=", "!=", ">=", "<=", and "LIKE" to implement prefix match, backward match and partial match.

With support of this module, teachers can view studying status for one entire class or for individual student. This information can be utilized to detect which areas should be focused and emphasized in future education, and to make guidance for individual student.

4. Discuss (Posting and viewing threads in discussion space)

Not only public discussion space for all students and teachers, but also discussion

spaces for each composition topic in the unit of class have been designed. After posting questions into discussion space, advices from teachers and classmates can be expected. Records of discussion will also be stored for future reference. With support of discussion space, the collaborative learning among students can be expected.

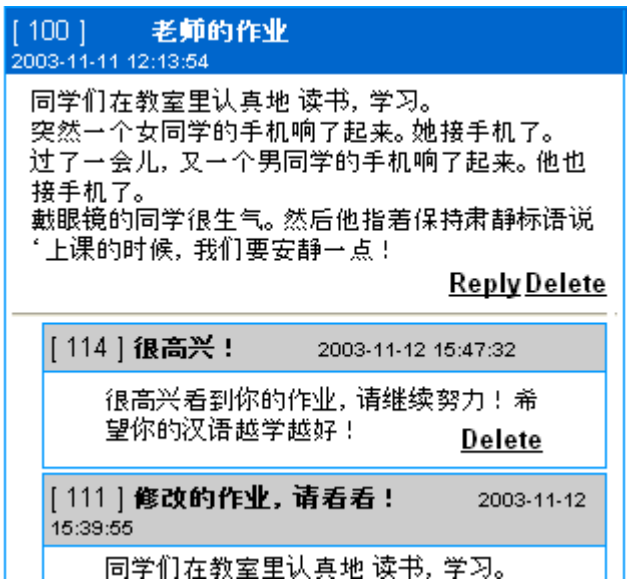


Figure 4-6 Interface of discussion space module

4.3.2 CONSTRUCTION OF LEARNERS' CORPUS

The learners' corpus in this system is divided into two parts, learners' profile information, and composition & correction information. Learners' profile information is stored in TC_User table of Tutorial Chinese Platform, which includes the fields of user id, name, gender, age, department in university, grade in university and class. Composition & correction information, which is stored with XML format, is an important feature of this system. The figure below shows correction information stored in learners' corpus. The upper part is commenting and correction interface of system, the lower part is XML tagged form stored in learners' corpus, the part enclosed in < > is XML tag.

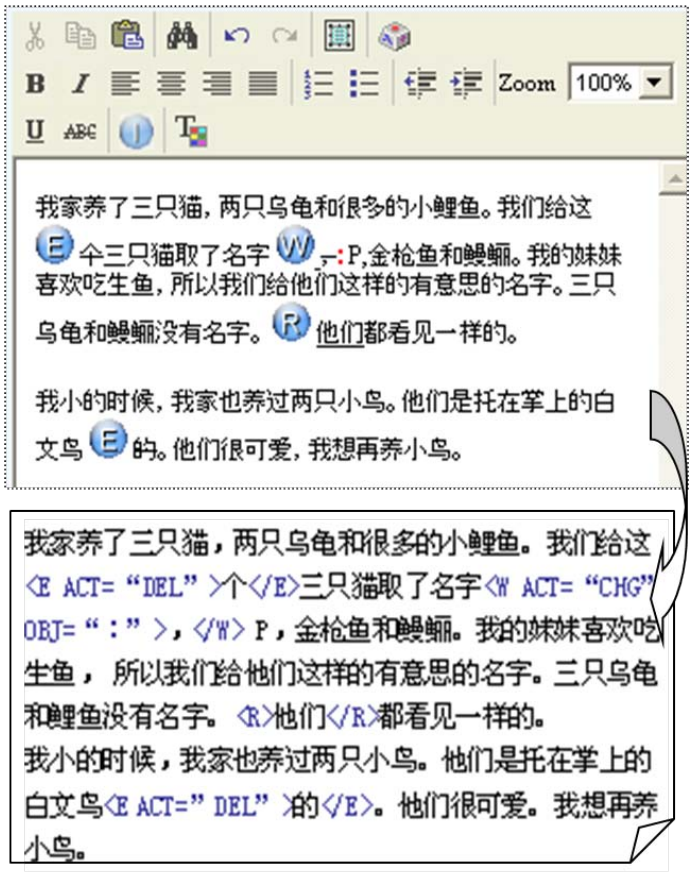


Figure 4-7　Correction information in corpus

The most common operations of correction: insert, change, delete and comment have been implemented online. In XML tags, the content in the label of “ACT” represents correction operations such as “Insert”, ”CHG” (change), ”DEL”(delete) and “Comment”.

The error classification is a difficult issue because it concerns about validity and reliability of correction. Carl James has stressed that should be at least two kinds of taxonomy (the term that James used for classification, which is defined in New Shorter Oxford English Dictionary as ‘the branch of science that deal with classification’) need be considered. (1) linguistic category classification and (2) target modification taxonomy. The former one carries out specification of errors in terms of linguistic categories. It will indicate that error is located on what level of language: in phonology, graphology, grammar, lexis, text or discourse level. The latter one relates the form diversity such as omission, addition, misformation, misordering, etc [James, 1998]. When deciding error classification, the taxonomy theory of James has been considered,

combined with teachers' pedagogical experience simultaneously.

The misuses have been classified into nine error types and "grammatical correction mark" has been utilized to label errors, which is list up as follows.

D: (Deformity), misuse about deformity and omission.
E: (Extra), misuse about extra of conjunctions, particles, subjects, etc.
C: (Correspondence), misuse about misordering, such as subject and predicate, verb and object, etc.
P: (Position), misuse about improper position of words.
W: (Wrong), misuse of characters or punctuation.
R: (Refer), misuse about unknown reference.
M: (Mix up), misuse about confusion usage, such as mix using of oral and written word.
L: (Logic), misuse about lack in logic.
O: (Others), the other kinds of misuse.

Combined with these nine error types, the four types of correction operations (insert, change, delete, and comment) has been implemented online. The samples of corresponding XML tags are shown as follows. The first capital letter in <> represents type of error such as <D> </D>. The content in the label of "ACT" represents correction operations such as "Insert", "CHG" (change), "DEL"(delete) and "Comment".

Insert: <D ACT="Insert" OBJ="content of insert"></D>
Change:<P ACT="CHG" OBJ="content after change">content before change</P>
<W ACT="CHG" OBJ="content after change">content before change</W>
<R ACT="CHG" OBJ="content after change">content before change</R>
<C ACT="CHG" OBJ="content after change">content before change</C>
<M ACT="CHG" OBJ="content after change">content before change</M>
Delete: <E ACT="DEL">content need to be deleted</E>
Comment:<L ACT="Comment">content of comment</L>
<O ACT="Comment">content of comment</O>

The XML tags make it very convenient to retrieve and analyze corpus. Moreover, after filtering tags by text analyzer, the original document will be recovered easily. This feature provides basis for design of retrieval and analysis functions.

4.4 OPERATION AND EVALUATION

4.4.1 OPERATION STATUS OF SYSTEM

This composition/correction system has been operated since October 2003. Every year, around 150 students make use of system to practice their writing ability. The table below shows sample of assignment topics of one school year. According to the important or interested topics in lesson reports, teachers from Beijing and Taiwan will assign two topics for elementary and advanced level respectively for each lesson. Students can choose one based on their language ability and favorite cultural.

Table 4-1 Assignment topics of one school year

第 1 課	臺灣	初级	自我介紹
		中高级	我的好朋友
	北京	初级	自我介绍
		中高级	我和我的家庭
第 2 課	臺灣	初级	釣魚
		中高级	我的興趣
	北京	初级	我喜欢……
		中高级	我为什么喜欢/不喜欢……
第 3 課	臺灣	初级	看圖說話/難忘的打工
		中高级	打工的好處和壞處
	北京	初级	难忘的一次打工经历
		中高级	打工与学习
第 4 課	臺灣	初级	看圖說話
		中高级	氣候的變化/比較
	北京	初级	家乡的气候
		中高级	梅雨
第 5 課	臺灣	初级	看圖說故事
		中高级	日本的節日
	北京	初级	新年
		中高级	家乡的节日
第 6 課	臺灣	初级	我家
		中高级	佈置自己的家

	北京	初级	我的家
		中高级	搬家/喜迁新居
第 7 課	臺灣	初级	國服
		中高级	我的衣服
	北京	初级	日本的传统服饰
		中高级	校服
第 8 課	臺灣	初级	旅遊
		中高级	旅遊
	北京	初级	难忘的一次旅游
		中高级	旅游的意义
第 9 課	臺灣	初级	逛街
		中高级	百貨公司
	北京	初级	买衣服
		中高级	在东京购物
第 10 課	臺灣	初级	拿手菜(日本)
		中高级	中國菜
	北京	初级	我最喜欢吃的菜
		中高级	晚饭
第 11 課	臺灣	初级	大學生活？
		中高级	最難忘的大學生活
	北京	初级	忙碌的学习生活
		中高级	我的社团生活
第 12 課	臺灣	初级	我的暑假
		中高级	最有意義的假期
	北京	初级	暑假回想
		中高级	我的悠长假期
第 13 課	臺灣	初级	我的理想
		中高级	40 年後的你
	北京	初级	毕业后的打算
		中高级	我的人生目标
第 14 課	臺灣	初级	我最喜歡的運動
		中高级	最難忘的運動會
	北京	初级	记一次比赛
		中高级	我看奥运比赛
第 15 課	臺灣	初级	第一次.....

		中高级	給救難人員的一封信
	北京	初级	为什么日本的小偷比较少?
		中高级	地震
第 16 課	臺灣	初级	手機與我
		中高级	手機的好處與壞處
	北京	初级	我的手机
		中高级	我为什么使用手机
第 17 課	臺灣	初级	我最喜歡的動物
		中高级	從動物飼養看動物保護
	北京	初级	我养过的宠物
		中高级	养宠物的经验
第 18 課	臺灣	初级	感冒的時候. . . . .
		中高级	中醫與西醫
	北京	初级	生病
		中高级	我了解的医疗保险制度
第 19 課	臺灣	初级	假如我是政治家
		中高级	我看台灣的選舉
	北京	初级	我关心的政治问题
		中高级	我看民主制度
第 20 課	臺灣	初级	假如我有一千萬元
		中高级	日本的經濟問題
	北京	初级	关于"年金"改革的看法
		中高级	中日经济交往中的问题
第 21 課	臺灣	初级	日本的恐怖攻擊事件
		中高级	我看日台關係
	北京	初级	我对战争的看法
		中高级	日本的国际责任与国际贡献
第 22 課	臺灣	初级	我最愛看的新聞類型
		中高级	新聞媒體的社會責任
	北京	初级	日本的媒体现状
		中高级	新闻媒体的可信度

During the period of October 2003 to November 2005, 1228 composition/correction documents have been stored into learners' corpus. The detail

information is shown in the following table. From this table we can find out that compared to compositions in elementary level, the compositions in advanced level have longer sentence, as well as more complex phraseology and richer vocabulary.

Table 4-2 Overview of stored learners' corpus

	Elementary Level	Advanced Level	All
Number of Pieces	678	550	1228
Number of Sentences	7748	9244	16992
Number of words	66554	95926	162480
Average length of sentences	8.6	10.4	9.6
Number of different words	4585	5594	7565

Through learners' corpus, the misuse status can be analyzed easily. The figure below shows distribution of types of misuse.

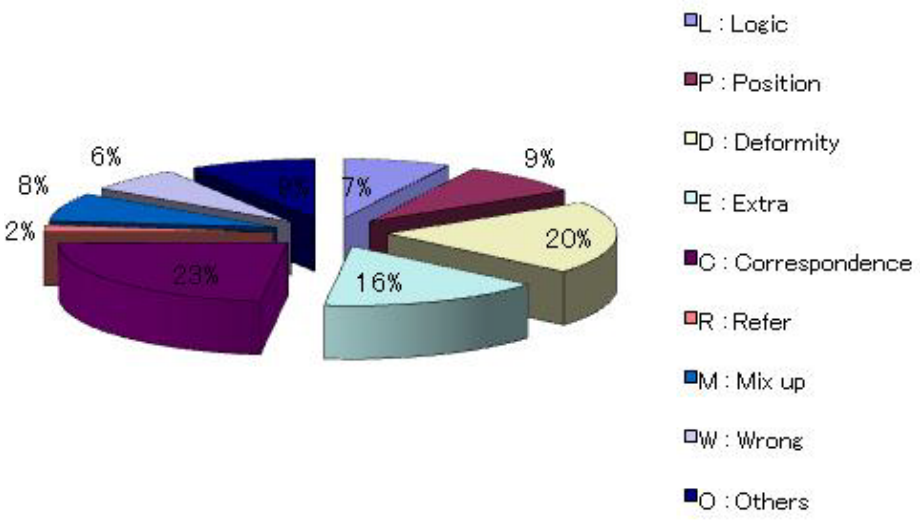


Figure 4-8 Frequency distribution of misuse

Correspondence (inadequate combination of subject and predicate, verb and object), Deformity (deformity and missing of predicate or verb status) and Extra (extra of

conjunctions, particles, subjects, etc) are the most three types of misuse. The occurrence rates are 22%, 19% and 16% respectively. Refer (unknown reference or numeral, quantifiers) has low frequency, only around 2%. The other types of misuse have almost the same frequency. This misuse distribution can be utilized in scientific analyses of future education and following assignment topics will be determined based on this statistics data.

4.4.2 USABILITY EVALUATION OF SYSTEM

In 2006, a survey about system usability evaluation has been conducted among six teachers. Each teacher has two years' experience of using system to correct composition. The evaluation concerns

(1) Operability of correction system
(2) Types of correction tool is enough or not
(3) Time consuming compared to paper correction
(4) Readability of correction result
(5) Effectiveness of misuse mark

Maximum scale of survey is 5 point. The result is shown as the following table.

Table 4-3 Result of usability evaluation for composition/correction system

Question Items	Average Score
(1) Operability of correction system	4.6
(2) Types of correction tool	4.8
(3) Time consuming compared to paper correction	4.8
(4) Readability of correction result	4.6
(5) Effectiveness of misuse mark	4.0

From the result of question (1) and question (2), we can see that system has good operability and types of correction tools are sufficient. The question (3) also gets a high score. In free answers space, we find out that with correction system, paper based

correction time in the past has almost been shorten to half now. About the question (4), there is comment that, because the familiar format on paper has been reproduced on the web, online correction result is similar to that on paper and it is very easy to read. The score of question (5) is slightly lower than other questions. In free answers, there are comments that, the “grammatical correction marks” are useful to recognize conception and categories of misuse, but there exists limitation in students’ understanding of these marks, detail explanation of misuse marks and joint use of comment is essential.

4.5 CONCLUSION

A web-based composition/correction system has been conducted as part of Tutorial Chinese Platform (TCP). This system is efficiently and effectively used not only for students in writing composition, but also for teachers in correcting and making comments in composition. Moreover, the compositions from students, together with the correction information has been structured and constructed automatically into learner corpus, and the retrieval function has also been developed, which is useful for researchers to analyze language resource from the real educational environment. System has many advanced features such as (1) supplying convenient online correction tools (2) unifying misuse marks to label errors with standardized criteria (3) stimulating enthusiasm by setting up appropriate topics and deadline (4) listing up compositions by urgency to reduce the time lag of feedback (5) constructing praise mechanism to promote students’ reconfirmation of comment (6) supplying discussion space to support collaborative learning (7) implementing automatic storage and analysis function for learner corpus.

This system has been used in Waseda University since 2003. The result of usability evaluation shows that composition/correction system has dramatically increased teachers’ working efficiency.

CHAPTER 5.
CAI SYSTEM FOR DISCRIMINATING CHINESE FOUR TONES

5.1 INTRODUCTION

Tone language has been defined as a kind of language that even the same phonemes consisted in one syllable; the dictionary meaning will be different because of the existing of high/low pitch, or relative rise/fall of pitch in this syllable [Chuang et al., 1975]. Chinese is tone language but not the only one. Some languages of Southeast Asia such as Vietnamese language, Burmese, Thai; Parts of Africa such as Western Ethiopia, Southern Sahara; and those of Latin American Indians are also tone languages. The samples of pairs of Chinese bi-syllabic words are shown below, which have the same phonemes but different meaning because of different tones.

	Chinese (Japanese)	Pinyin	Tone
Sample(1)	支持 （支持する）	zhi1 chi2	12
	智齿 （親知らず）	zhi4 chi3	43
Sample(2)	事件 （事件）	shi4 jian4	44
	时间 （時間）	shi2 jian1	21
Sample(3)	厂家 （製造業者）	chang3 jia1	31
	长假 （長期休暇）	chang2jia4	24
Sample(4)	厨房 （厨房）	chu2 fang2	22
	处方 （処方）	chu3fang1	31

Mandarin Chinese has four tones. These tones are neither absolute and fixed pitch

nor ratio of pitch to the average pitch of speaker, but just the relative difference between pitch of one syllable and pitch before or after that syllable. Tone-1 is flat pattern in mid-high level. Tone-2 is rising pattern from mid-low to mid-high. Tone-3 is long pattern with low flat beginning and rising end. Tone-4 is short falling pattern from high to low. There are tone symbols, which are written over the nuclear vowel in a syllable, to imitate these features, such as dì tiě (Subway), where the first syllable is tone-4 and the second syllable is tone-3; xīn wén (news), where the first syllable is tone-1 and the second syllable is tone-2; yóu yǒng (swimming), where the first syllable is tone-2 and the second syllable is tone-3. Besides these four tones, there is another type, sometimes called neutral tone, but not a real tone. Its pitch depends on the tones of preceding and following syllables. The syllables with neutral tone will be described by absence of tone, such as tù zi (rabbit), where the second syllable is neutral tone.

Because Japanese is not tone language, it is the first barrier for Japanese learners to understand features of tones when they study Chinese. Computer-Assisted Instruction (CAI) system for self-teaching of discriminating Chinese four tones has been developed, and been provided through the Internet to Japanese college students in beginner's course of Chinese language. According to the ability of each student, by selecting suitably among uses of screening, practicing and reviewing mode, and also among various bi-syllabic word lists with different degree of difficulty, the required learning time for achieving the goal can be reduced.

In order to give students more direct perception of Chinese tones, system presents visualized pitch pattern of tones and gives an effective guidance for tone discrimination. Moreover, precise acoustical data of the Chinese tones have been utilized for designing the CAI system. Inexperience in tonal perception by Japanese students has been taken into consideration in the method of compiling word lists from corpus. Characteristics of errors found in the results of examinations given to about 100 students have been analyzed, and have been referred to in constructing the CAI system.

In this chapter, the issues considered in design, construction and evaluation of this CAI system will be described.

5.2 PURPOSE OF THE STUDY

Because Japanese is not tone language, when Japanese students study Chinese language, especially in the early stage, it is very difficult to discriminate the difference among tones. The discrimination ability of individuals has great variation and many students are slow behind others. The purpose of this study is to design and construct a CAI system for self-teaching of discriminating Chinese four tones for elementary learners to study outside classroom and recover the delay.

In order to achieve this goal, this system gives a more direct perception about Chinese tones visually through illustrations of different tone's pitch contours, and gives an effective guidance for tones discrimination. System has been designed and constructed based on the analyses of characteristics of mistakes and errors made by Japanese students during their Chinese tones discriminating. There are two kinds of studying course in system, standard course and intensive course. In intensive course, the studying route selection algorithm has been embedded to enable different level students to study by themselves according to their own listening proficiency. This CAI system makes independent learning possible by providing target-oriented studying course to students to practice spontaneously as many times as possible. Moreover, because all testing results have been recorded into learner corpus in digital format, it is easier for teachers to analyze how the syllable positions affect tones discriminating and the characteristics of errors occurring in four tones. Learners can also take advantage of this information to help them improve their tones discrimination ability.

5.3 STANDARD PITCH PATTEN OF CHINESE FOUR TONES

The visualization characteristic of acoustic analysis has already been utilized in the field of special education for deaf children for long while, such as voice guidance [Hiki, 1977]. It has recently spread into general education of foreign language [Sakamoto, 2003]. In Chinese language education, the visual presentation using voice pitch pattern has also been tried [Takagi et al., 2005] [IFLYTEK, 2003]. However, in the education and training of discriminating tones, there is still no effective tool using voice pitch pattern designed based on the analysis of learners' characteristics.

In this CAI system, visualized pitch pattern of tones plays very important role, and should be expressed as easy as possible to learners. For this purpose, from the precise acoustic analysis results[Chuang et al., 1975], standard visualized pitch patterns of mono-syllabic word and bi-syllabic word have been derived and based on these standard visualized pitch patterns, guidance manual for tones discrimination has also been made.

5.3.1 STANDARD PITCH PATTERN OF MONO-SYLLABIC WORD

Standard pitch pattern of mono-syllabic Chinese word is shown in the graph below. Solid straight line is the fundamental part and dotted curve line is part of accompanying change. Six diatonic (whole tone) has been used in the notation [Hiki et al., 2005]. The change in fundamental part of pitch pattern shows the features of each kind of tone. Tone-1 is flat pattern in mid-high level. Tone-2 is rising pattern from mid-low to mid-high. Tone-3 is long pattern with low flat beginning and rising end. Tone-4 is short falling pattern from high to low.

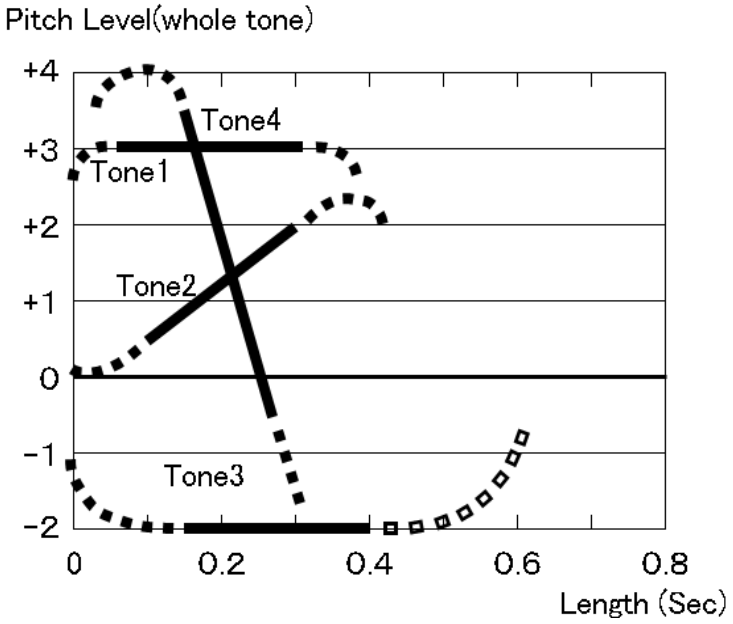


Figure 5-1 Standard pitch pattern of Chinese mono-syllabic word

5.3.2 STANDARD PITCH PATTERN OF BI-SYLLABIC WORD

In the case of bi-syllabic words, there are fifteen combinations of four kinds of tones in the first and second syllable. (One combination is missing because the tone changes from tone-3 + tone-3 to tone-2 + tone-3). The pitch pattern of combination of four tones in bi-syllabic word is represented in the following figure. In each small graph of left side, there is same tone in the first syllable and four different tone combinations in the second syllable. In right side, there are four different combinations with four kinds of tone in the first syllable and same tone in the second syllable.

In bi-syllabic words, the changes of voice pitch in each of the two syllables are concatenated, shown in dotted curve in the figure. But, the substantial portion of tones features observed in mono-syllabic word is remained unchanged. Because the two syllables will be phonated together, there is a little changes compare with the utterance of mono-syllable word. For example, height of pitch pattern of the second syllable is lower than that of the first syllable by about semitone; length of pitch pattern of the first syllable is slightly shorter than that of the second syllable. Moreover, tone-4 in the second syllable is wider than in the first syllable, and the slope of falling is also steeper.

Tone-2 in the second syllable becomes narrower than in the first syllable, and the rising becomes gently.

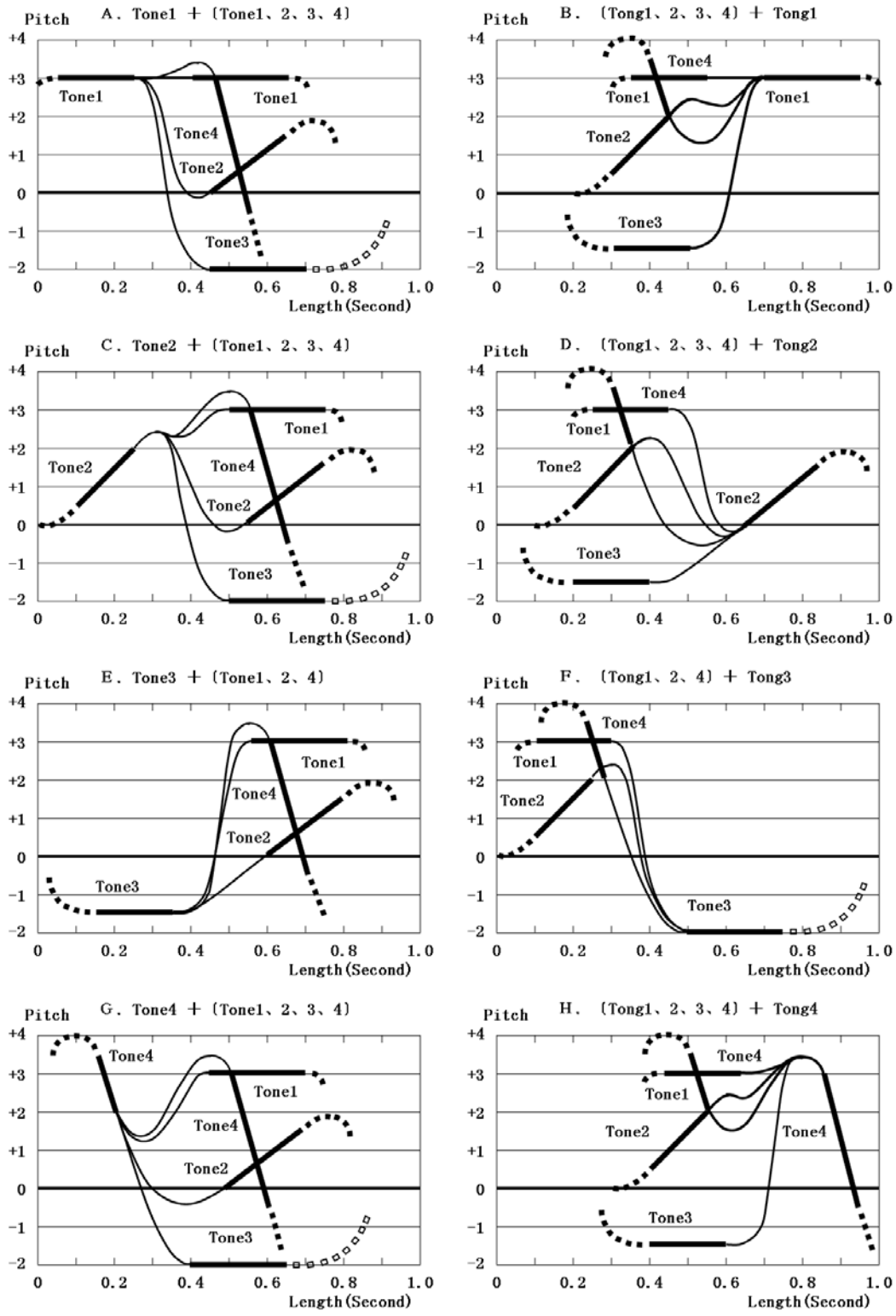


Figure 5-2 Standard pitch pattern of combination of four tones

in Chinese bi-syllabic word

5.3.3 CONTRAST OF CHINESE TONE AND JAPANESE ACCENT

Japanese also has the feature of pitch accent. It also uses phonemic tone, but only one or two syllables in a word can be phonemically marked for tone, and many words are not marked for tone at all. Japanese has fewer complexes than fully tonal language as Chinese. For comparison, the pitch pattern of accent of Japanese words is shown in the right part of figure below. It is an example of two words' accent, 商人 [So ù nin]("Merchant"), with accent, and 証人 [So ù nin] ("witness"), without accent. From the graph we can see the difference between Chinese tone and Japanese accent.

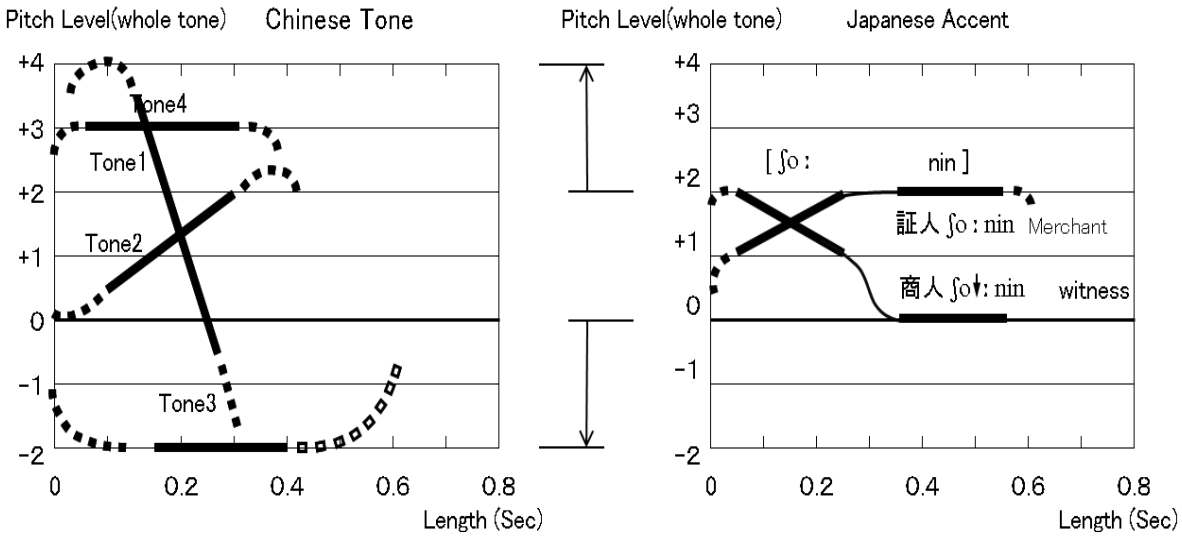


Figure 5-3 Voice pitch pattern of Chinese tone and Japanese accent

Japanese accent has only one kind of change in the type of voice pitch—steep fall. However, in Chinese, there are flat, low, rising, falling, and even the combination of these change patterns. The pattern of change is more complex in Chinese than in Japanese.

About the scope of change, the voice pitch difference between low and high contrast in syllabic segment associated with Japanese word accent is two whole tones in diatonic scale, while in Chinese, the pitch difference has covered six whole tones in diatonic scale (one octave). Although in Japanese pitch accent, there is change of steep fall, because the scope of change is too different, it is still difficult for Japanese native speaker to discriminate the falling pitch pattern in Chinese mandarin.

From the figure, we can also see that the length of Japanese syllables is fairly stable, even accent being added, the length only grows not more than 20%. While in the case of Chinese, tone-3 is clearly longer that tone-1 and tone-2, and tone-4 is shorter that tone-1 and tone-2. The length of tone-3 is two times longer than that of tone-4. For Japanese native speaker, how to master the length difference among tones is also difficult.

Because of the above difference between Japanese and Chinese, in the elementary stage of studying, Japanese learners will encounter many difficulties in discriminating of Chinese tones. In this background, this CAI system has been designed for Japanese students for self-teaching of discriminating Chinese four tones outside classroom.

5.4 WORD LISTS IN CAI SYSTEM

5.4.1 WORD LISTS SELECTED FROM CORPUS

The words list utilized in CAI system are selected from the vocabulary corpus of Waseda Tutorial Chinese Corpus (WTCC) introduced in Chapter 2. Because the targets of this CAI system are students in elementary stage, words of low difficulty level have mainly been selected.

In CAI system, practice is carried out in unit of tables. Each table has been constructed with fifteen bi-syllabic words. These fifteen bi-syllabic words represent all combinations of four tones in the first and second syllable (One is missing because the tone combination changes from tone-3 + tone-3 to tone-2 + tone-3). The figure below shows an example of one table. In the figure, for the words in one line, tone of the first syllable is in order of tone-1, tone-2, tone-3 and tone-4. For the words in one column, the tone of second syllable is in order of tone-1, tone-2, tone-3 and tone-4. The information embedded in word list includes: tone combination, Chinese character, Japanese translation, Pinyin (displayed by both tone numbers and tone symbols), visualized pitch pattern and speech waveform. For each word, the information displayed in the sequence of tones' combination, simplified Chinese character, Japanese translation, Pinyin (with tone numbers), Pinyin (with tone symbols), visualized pitch pattern and audio waveform diagram, from top to bottom.

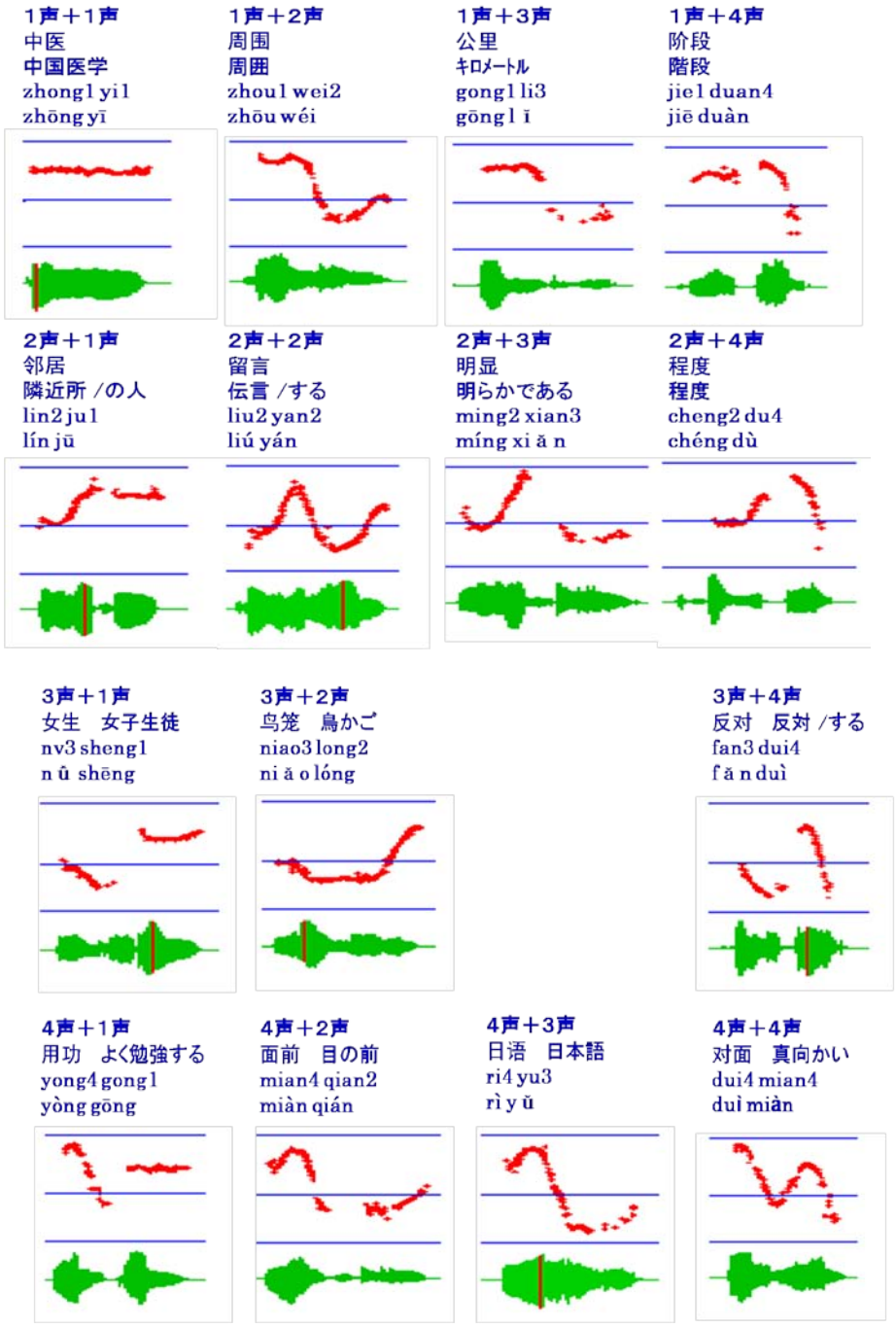


Figure 5-4　Sample of one table in CAI system

Words have been chosen based on the type of tones' combination, as well as on other acoustic features of words such as trying to cover different types of vowels and consonants. Moreover, preliminary hearing test has been carried out with all optional words. Based on the overall results, the words with relatively high error rate, so called relatively difficult words have been constructed in four tables, and the words with

relatively low error rate, so called relatively easy words have been constructed in the other four tables. Both male and female vocal recordings have been taken for all words.

The tables utilized in CAI system are listed as follows:

Relatively easy tables
Female voice: A1, A2, A3, A4; Male voice: A1, A2, A3, A4

Relatively difficult tables
Female voice: B1, B2, B3, B4 Male voice: B1, B2, B3, B4

According to the result of preliminary hearing test, error rates of words with only tone-1 and tone-4 (include tone-1 + tone-1, tone-1 + tone4, tone-4 + tone-1 and tone-4 + tone-4) are particularly low. In order to improve the efficiency of practice for those learners who have already mastered discrimination of tone-1 and tone-4, tables without the above four combinations have been constructed, only eleven combinations involved tone-2 and tone-3.

From table A1, A2, A3, A4, to construct
Female voice: C1, C2, C3, C4; Male voice: C1, C2, C3, C4

From table B1, B2, B3, B4, to construct
Female voice: D1, D2, D3, D4; Male voice: D1, D2, D3, D4

The contents in all tables will be described in the appendix.

5.4.2 PROCESS OF CONSTRUCTION OF PITCH PATTERN

As introduced above, through the display of visualized pitch pattern, the utilization of visual images to reinforce the auditory discrimination is a significant feature of this CAI system. Therefore, the accuracy in constructing of visualized pitch pattern is very important. This process consists of the following steps.

1. To get pitch contour with WaveSurfer

In order to get visualized pitch pattern of each word, free software named WaveSurfer have been used [WaveSurfer HP], which is an open source tool for sound processing and visualization. As the figure below shows, the panels from top to bottom are respectively waveform, spectrogram, pitch contour and time axis. By setting up parameters of pitch tracking algorithm according to speaker's sound characteristics, the pitch contour will be analyzed and constructed.

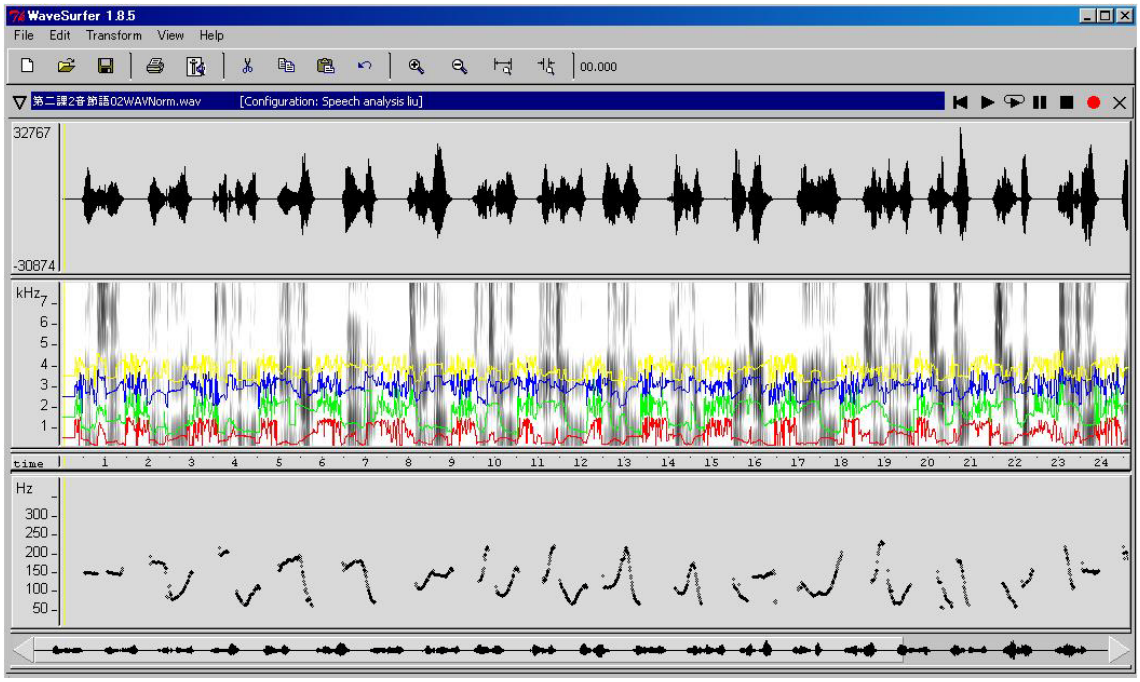


Figure 5-5 Construction for pitch pattern

2. To get data of pitch contour

With WaveSurfer, the fundamental frequency (F0) of pitch contour can be saved as a group of data in text file. Each point of data represents frequency (Hz) of sound in a moment, where the time difference between adjacent two points is to 10ms of time units.

3. To process raw data with excel

Because people's hearing sense to sound is proportional to the logarithm of fundamental frequency, macro program in excel has been written to calculate the

logarithm of F0. Then, for those singularities running out of frequency curve, appropriate average value from values before and after the points is calculated to smooth pitch contour. After calculation for value of vertical axis for each time unit, line chart is constructed based on the calculation result. That is shown as figure below.

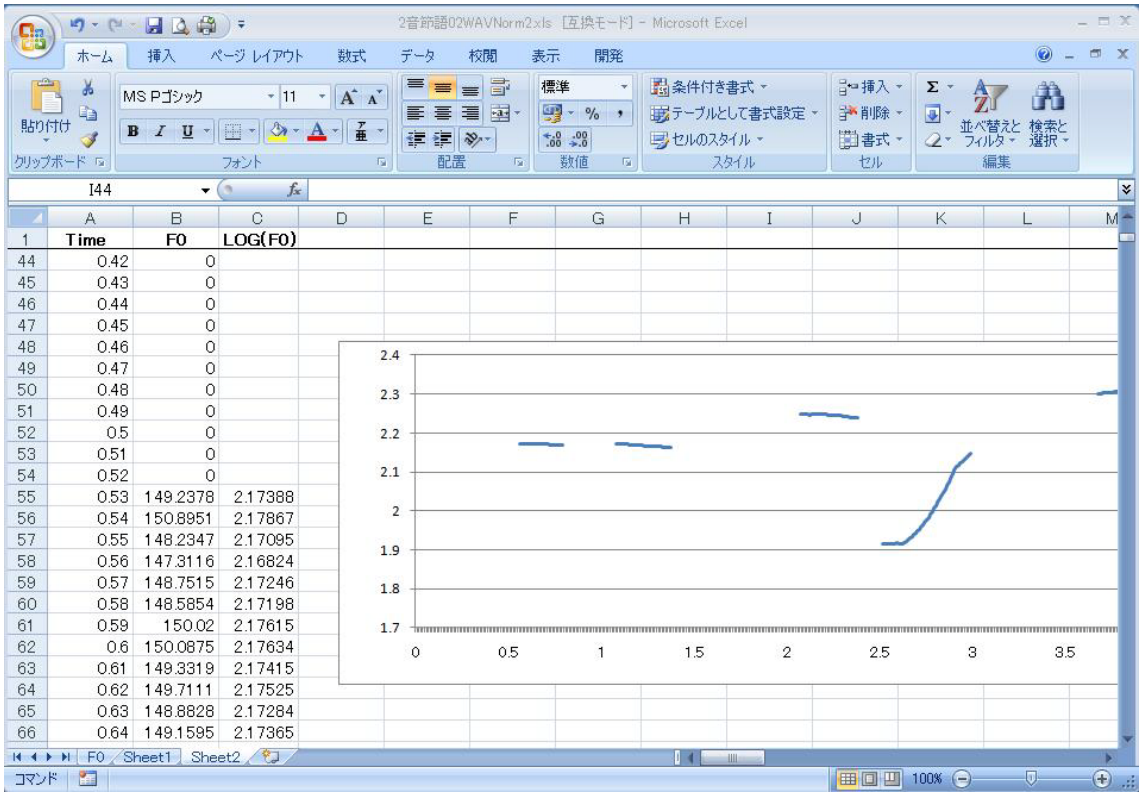


Figure 5-6 Analyzing result of F0 in excel

From the analyzing result, we can see that, the pitch pattern of speech samples of actual individual words is not complete same as the standard pitch pattern. In the educational process, we have given advice to teachers that, firstly, to make use of standard pitch pattern in classroom to deepen understanding and feeling of tone; then, to utilize this system to train students' discrimination ability.

5.5 IMPLEMENTATION OF CAI SYSTEM

CAI system has been carried out through Internet, so learners can make use of system to practice their discrimination ability in anytime and anywhere. In order to record learners' studying history, it is necessary for users to input their ID when login the system.

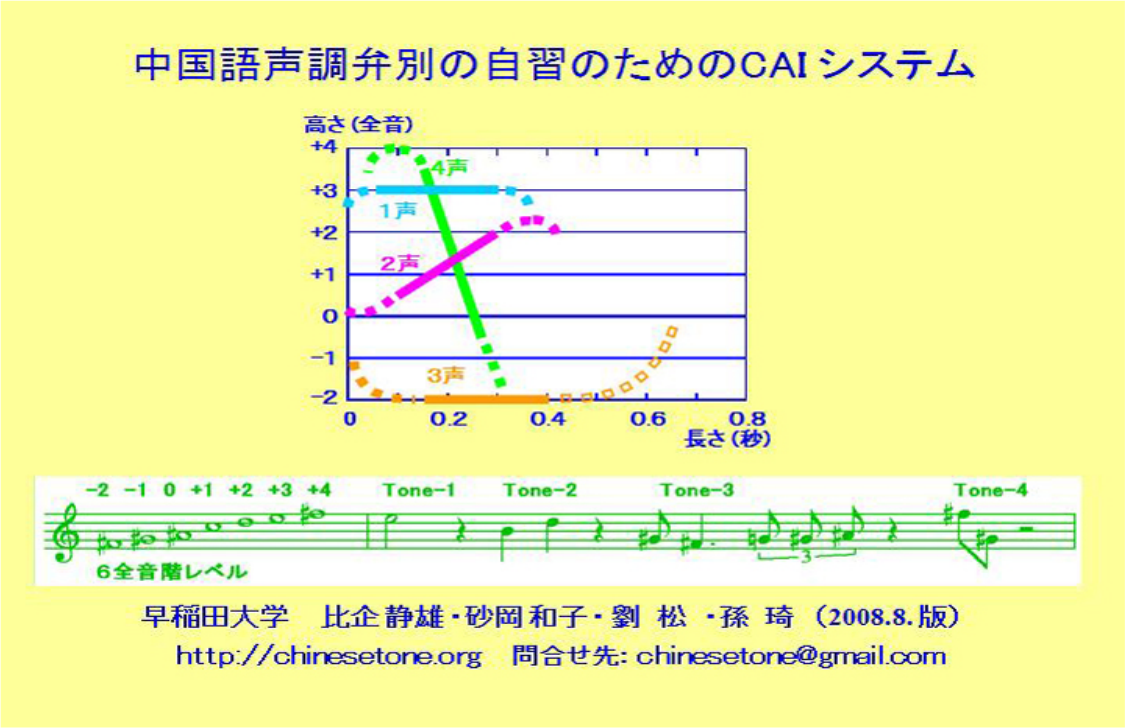


Figure 5-7　Login interface of CAI system

Figure 5-8 Table selecting interface of CAI system

5.5.1 OVERVIEW OF STUDY COURSE DESIGN

There are two kinds of course carried out in CAI system: standard course and intensive course. The main difference between two courses is that, in standard course, the sequence of presented word tables is step by step based on level of difficulty. While in intensive course, the sequence of presented word tables is according to the students' learning result and appropriate word table will be presented to them based on their practice result. Therefore, it will not waste learners' time to practice tables too easy or too difficult for them. Of course, even in intensive course, learners can also choose word table they want to practice freely, apart from the intensive sequence.

5.5.2 STANDARD COURSE

In standard course, the sequence of presented word tables is based on level of difficulty, which is female voice A, male voice A, female voice B, male voice B, female voice C, male voice C, female voice D, male voice D. Moreover, CAI system provides

three modes according to the learning process.

In screening mode, voice of question word will be presented only once. After listening to the voice, learners are supposed to click their mouse to choose the corresponding tone symbols. Then system proceeds to next question. After finishing one table, the errata list with Chinese character, Pinyin and tone symbol of all words will be displayed to serve the reviewing purpose and the errata list can also be printed out. Through the screening mode, learners will find out which combinations of tone are difficult points to them.

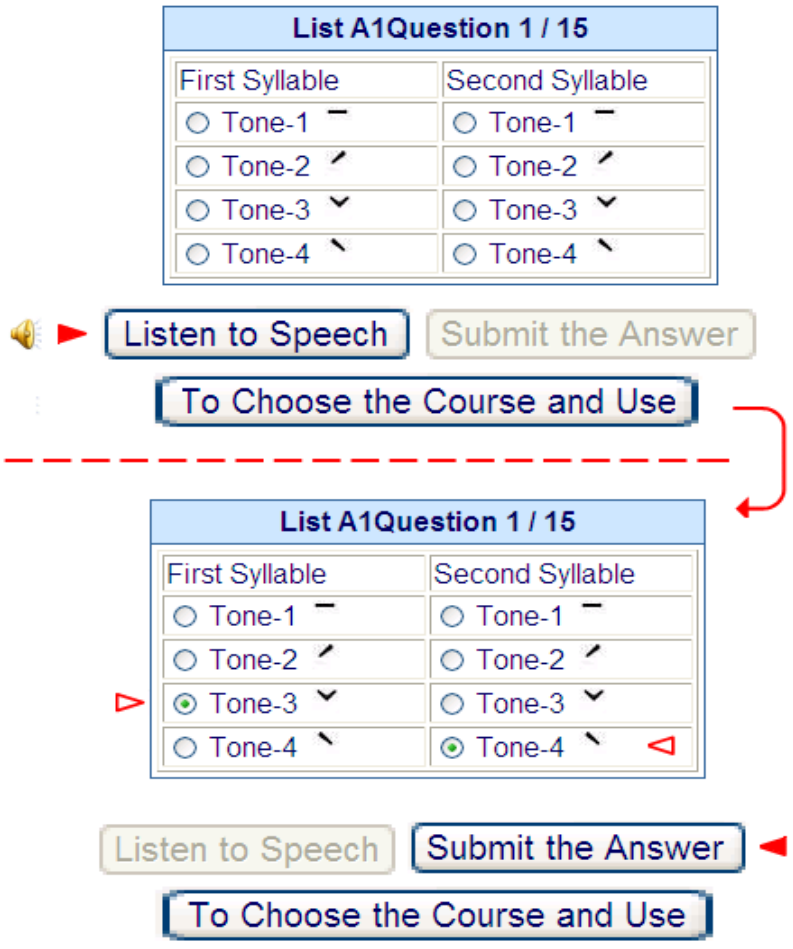
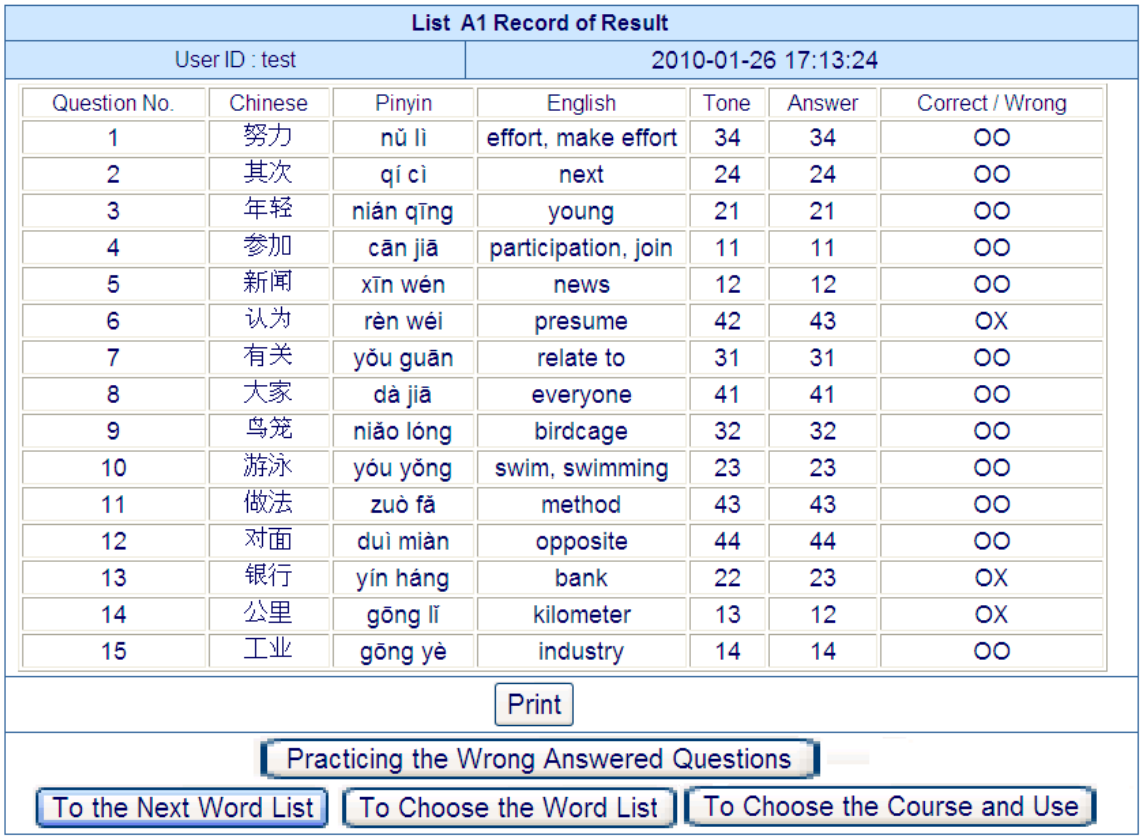


Figure 5-9 Interface of screening mode of CAI system



List A1 Record of Result

User ID : test | 2010-01-26 17:13:24

Question No.	Chinese	Pinyin	English	Tone	Answer	Correct / Wrong
1	努力	nǔ lì	effort, make effort	34	34	OO
2	其次	qí cì	next	24	24	OO
3	年轻	nián qīng	young	21	21	OO
4	参加	cān jiā	participation, join	11	11	OO
5	新闻	xīn wén	news	12	12	OO
6	认为	rèn wéi	presume	42	43	OX
7	有关	yǒu guān	relate to	31	31	OO
8	大家	dà jiā	everyone	41	41	OO
9	鸟笼	niǎo lóng	birdcage	32	32	OO
10	游泳	yóu yǒng	swim, swimming	23	23	OO
11	做法	zuò fǎ	method	43	43	OO
12	对面	duì miàn	opposite	44	44	OO
13	银行	yín háng	bank	22	23	OX
14	公里	gōng lǐ	kilometer	13	12	OX
15	工业	gōng yè	industry	14	14	OO

Print

Practicing the Wrong Answered Questions

To the Next Word List | To Choose the Word List | To Choose the Course and Use

Figure 5-10 Result list of one table

In practicing mode, firstly, voice of question word will be presented once, after choosing the corresponding tone symbols, Chinese character, Pinyin and tone symbol will be displayed, as well as the answer that learners chose and the correction information. If desired, learners can repeat listening to voice of question word and referring to the visualized pitch pattern. Learners can keep practicing until they comprehend the question word. In the upper part of figure 5-11, the radio buttons in dark blue color show the answers that learner chose (in this example, tone-1 and tone-2 are chosen), and the areas in light blue show the right answers (the right answer is tone1 and tone 1 in this example). In the lower part of figure 5-11, all the information of question is shown in lower left. If learners click the button “Listen to Speech”, voce of question word will be played repeatedly. The figure in upper right shows standard pitch pattern and the figure in lower right shows actual pitch pattern of word.

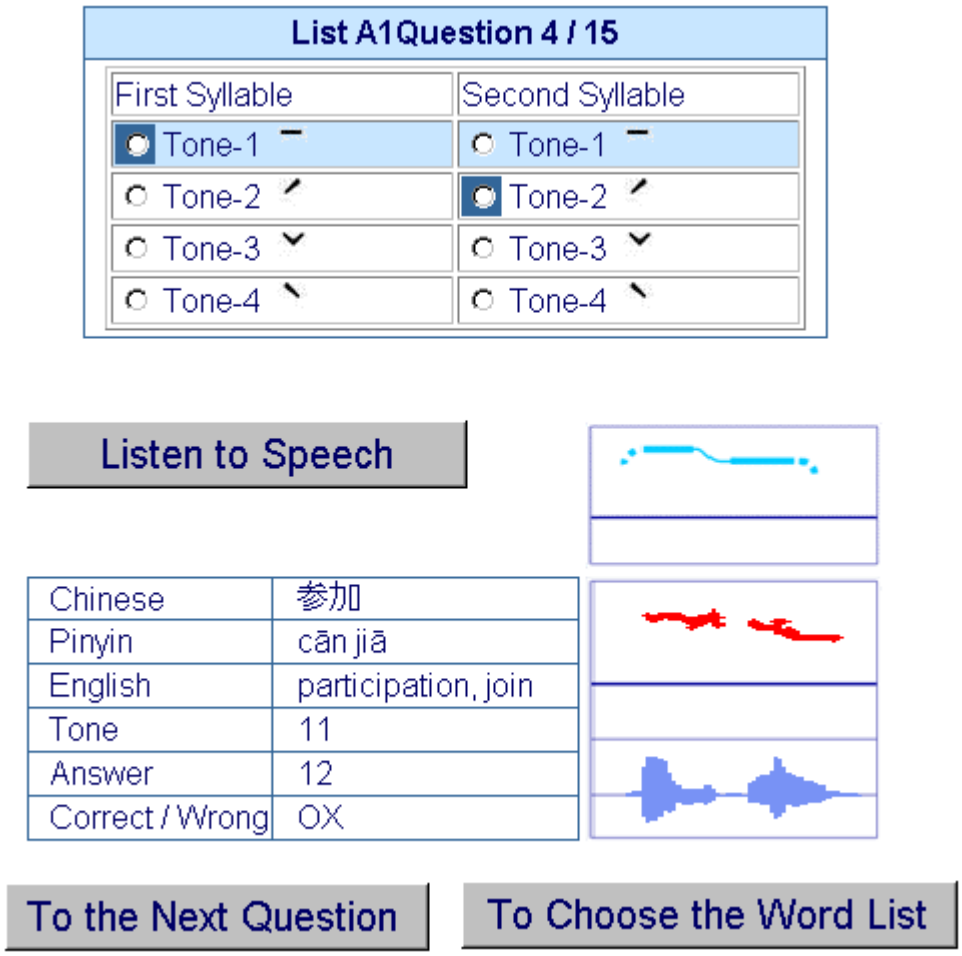


Figure 5-11 Interface of practicing mode of CAI system

In reviewing mode, Chinese character, Pinyin, tone symbol and visualized pitch pattern of question word will be displayed at the same time with the presentation of voice and no need to choose answers. Learners can confirm all the information simultaneously one by one, to achieve the purpose of review.

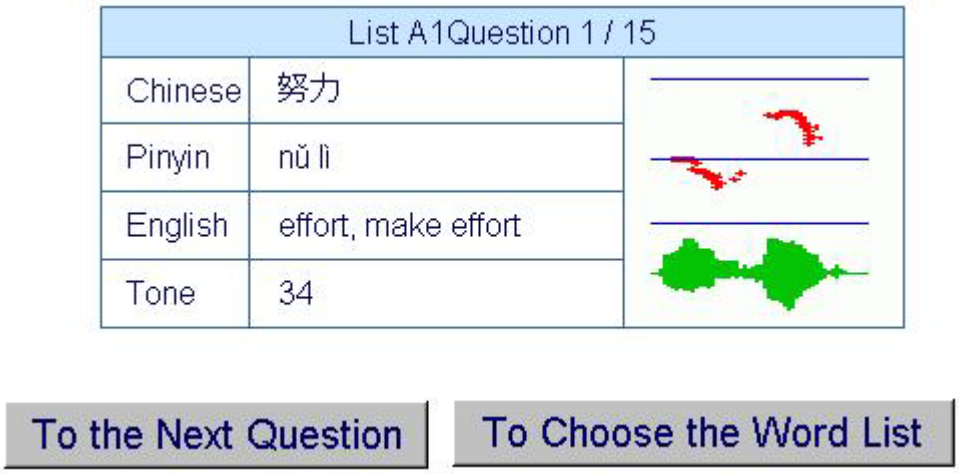


Figure 5-12 Interface of reviewing mode of CAI system

5.5.3 INTENSIVE COURSE

In intensive course, CAI system provides two modes: screening mode and reviewing mode.

The screening mode of intensive course is basically same to that of standard course. In order to improve training efficiency, after practicing one word list, the questions with wrong answer will be remembered by system, and can be practiced again if desired.

In reviewing mode, only the word lists practiced in screening mode will be presented.

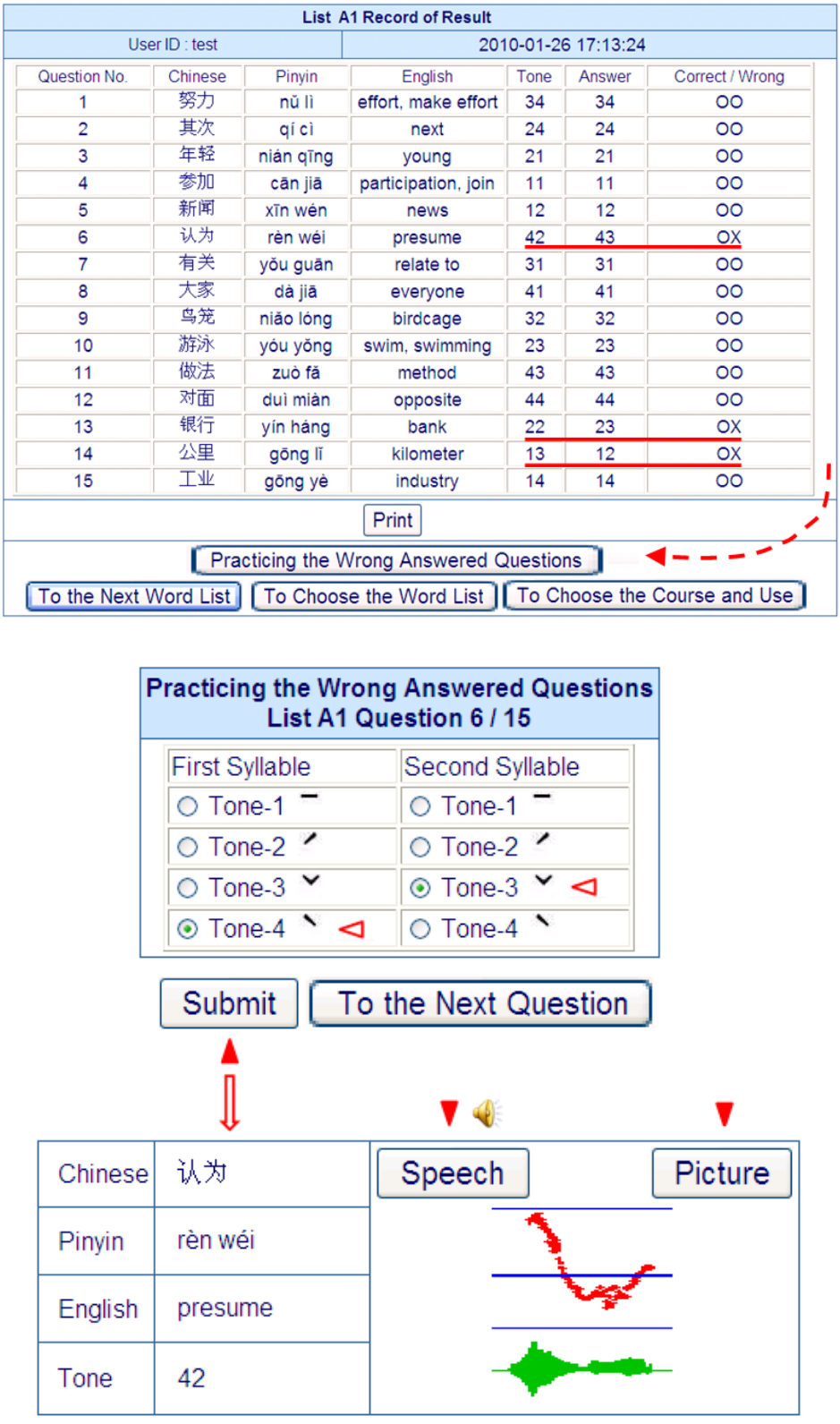


Figure 5-13 Interface of repeat practicing of questions with wrong answer in intensive course

Intensive studying route selection algorithm has been designed in order to correspond to different ability levels of learners. In this studying route selection algorithm, the presented sequence of word lists will be flexible according to the studying results of learners. Appropriate word list will be presented to learners based on their ability to avoid unnecessary wasting of time. The route is shown in the following figure.

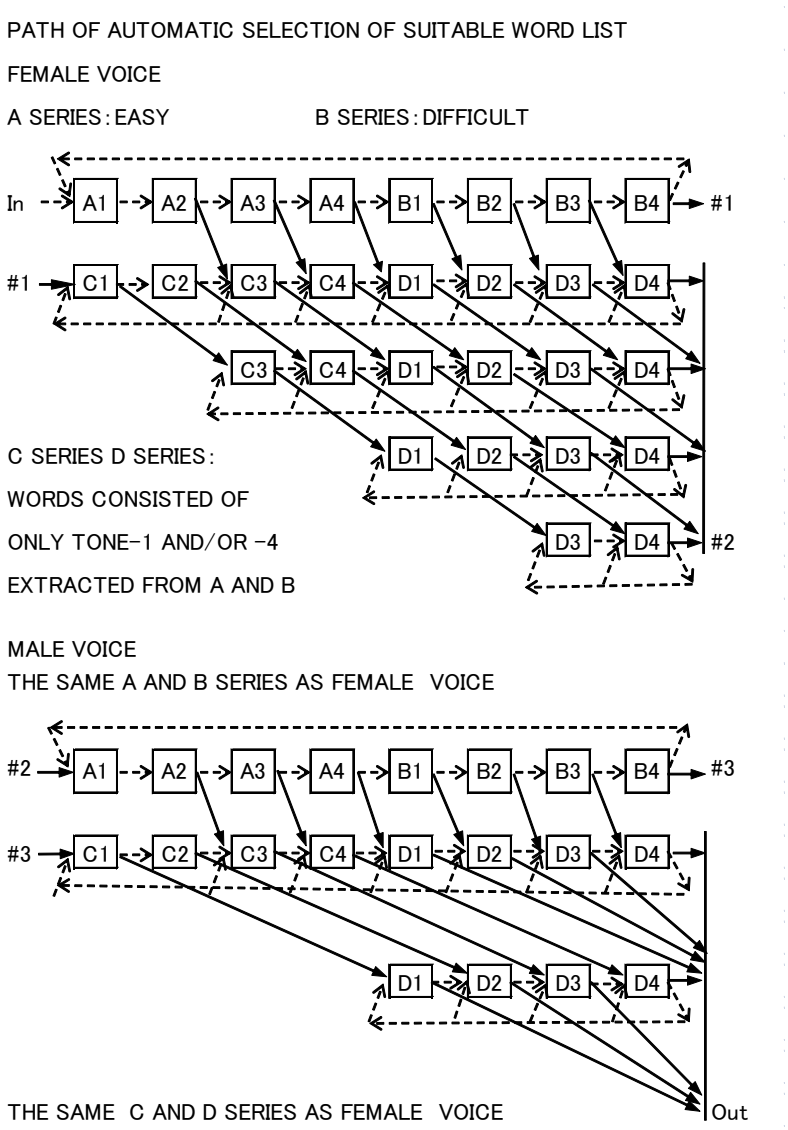


Figure 5-14 Automatic selection of suitable word lists in intensive studying route

As introduced above, eight word lists have been designed and constructed from WTCC vocabulary corpus. Based on the result of preliminary hearing test, all the optional words have been distributed into these eight tables to construct four relatively

easy ones: A1, A2, A3, A4; and four relatively difficult ones: B1, B2, B3, B4. Because the words with only tone-1 and tone-4 are particularly easy for learners to distinguish, the other eight tables without combination of tone-1 and tone-4 have been constructed: C1, C2, C3, C4 based on A1, A2, A3, A4; and D1, D2, D3, D4 based on B1, B2, B3, B4. Both male and female vocal recordings have been taken for all lists. Therefore, there are altogether thirty-two word lists embedded in the system.

The automatic selection algorithm in intensive studying route is as follows:

(1) Starting point is one table of A series with female voice, which is the most elementary table in system(dashed arrows in the figure). If learners gives correct answers to the words with all combinations of tone-1 and tone-4 in two consecutive tables, then in the next step, system will present the following table in sequence in C or D series(diagonal solid arrows in the figure). Otherwise, system will present the following table in original sequence(dashed arrows in the figure). According to this route, for those learners who have already mastered the ability of discrimiting tone-1 and tone-4, the number of practicing words in the following route will be reduced to about 3/4 (need not take the practice of four words with tone-1 and tone-4).

(2) If learners still give wrong answer to words with tone combinations of tone-1 and tone-4 even in the last table of B series, system will present the first table of A series again to let learners reinforce practicing (the dashed arrows above A and B series in the figure).

(3) In C or D series, if learners give perfect correct answer to one table, system will present the next next table in original series. So the next step in learners' study route will be jumping to the next of next table (diagonal solid arrows in the figure). Otherwise, system will present the following table in original sequence(dashed arrows in the figure). According to this route, for those learners who have given perfect correct answer for one table in C or D series, the number of practicing words in the following route will be reduced to 1/2.

(4) If learners still give wrong answers even in the last table of D series, system will present the first table of original C or D series again to let learners reinforce practicing (the dashed arrows under C and D series in the figure).

(5) If learnerd give perfect correct answer to the final table of female voice, system will present the first table of A or B series of male voice (#2 in the bottom half of figure).

(6) In the bottom half of figure, it has same words in tables with only different voice, the purpose of this part is just to confirm the practicing effect once again through words with male voice. In A and B series, the route is as same as that described in (1) and (2). In C and D series, if learners give perfect correct answer to one table, the next step in learners' practicing route will be jumping to the fourth table behind (diagonal solid arrows in the figure). The number of practicing words in the following route will be reduced to 1/4.

(7) If learners give perfect correct answer to the final table of female voice, the practice will be finished. Otherwise, system will present the first table of original C or D series again to let learners reinforce practicing.

Based on this intensive studying route, learners with high ability will not waste too much time to do the standard practice step by step. For example, those learners who give perfect correct answer to all the tables, it needs only nine tables, about 1/4 of that number in the standard course route to finish all practices.

5.6 EVALUATION OF CAI SYSTEM

In order to verify the effectiveness of CAI system, a testing site has been designed. There are four tables utilized in testing site, two tables are recorded by male voice and the other two tables are recorded by female voice. The vocabularies in four tables are different respectively and different to the sixteen tables in CAI training system. In testing site, learners can only know their result of correct numbers; other information like visualized pitch pattern, correct answer, etc, will not be notified. Therefore learners can login testing site repeatedly and verify their improvement. By gathering and analyzing the results of different examinations with the same login ID, learner's tendency and characteristic of wrong answer can be analyzed.

Experiment was carried out in May 2008 to examine the effect of improvement in discriminating ability of Chinese tone. The experimental targets are freshmen students in Waseda University, who have learnt Chinese for less than 3 months. The process of experiment is:

1. Students login CAI testing site in classroom and take the examination of four tables of bi-syllabic words.

2. In the following two weeks, students are asked to utilize CAI training site to take practice freely in their homes or on campus.

3. After two weeks, students login the same testing site again in classroom and take the same examination with two weeks ago.

Forty freshmen students have taken all the three steps. As comparison, test result of the other group of sixteen students only taken step 1 and step 3 is also considered.

The results of two examinations and answer records of training site, as well as the login histories recorded by CAI system have been analyzed to get the improvement of results and answers characteristics.

5.6.1 TRAINING EFFECT OF CAI SYSTEM

There are fifteen bi-syllabic words in one table, so there is a total of 120 syllables

in four tables of testing site. According to their answering accuracies of examination, the examinees are divided into three levels:

High level: correct rate is higher than 75% (more than 90 syllables are correct out of 120 syllables).

Middle level: correct rate is between 50% and 75% (60-90 syllables are correct out of 120 syllables).

Low level: correct rate is less than 50% (less than 60 syllables are correct out of 120 syllables).

1. Changes in frequency distribution of correct answers

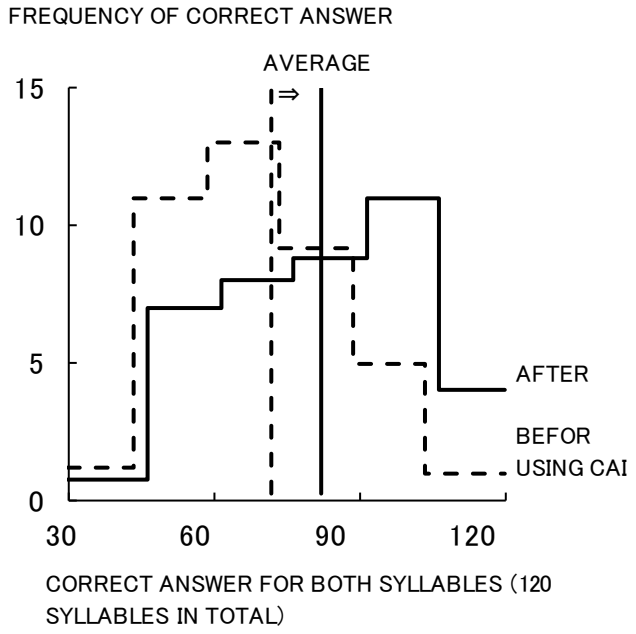


Figure 5-15　Frequency distribution of correctness

The contrast of correct answers' frequency distribution of two examinations is shown in the figure above. The horizontal axis is distribution of correct answers (by the unit of syllable) and the vertical axis is the degree of correct answer (here is number of examinees).

In the first time examination before taking practice of CAI system, the number of low and middle level examinees is high (dotted line), while in the second time examination after taking practice of CAI system, the number of high level examinees has increased obviously (solid line). In average, the number of correct answers has increased 10 (from 73 to 83) between two examinations.

2. Increase in the correct rate of each student

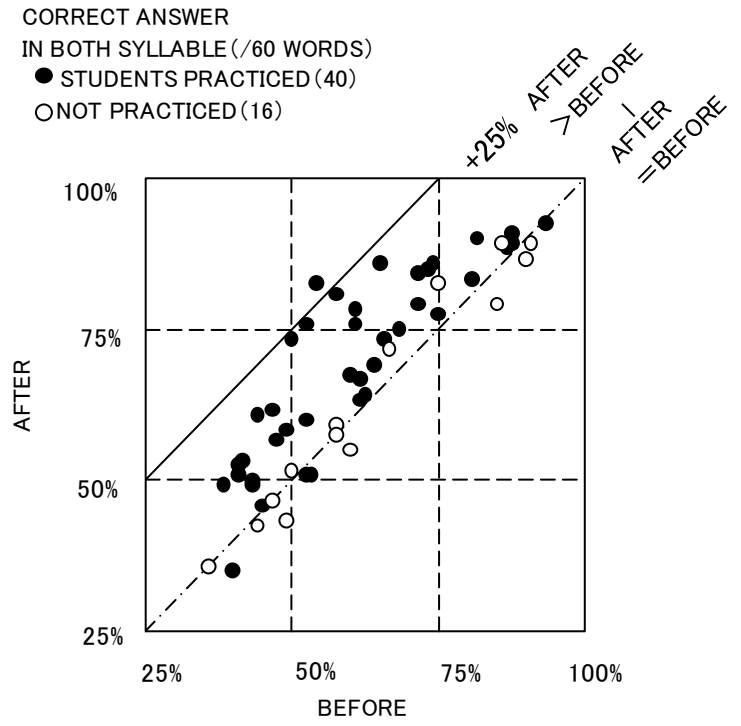


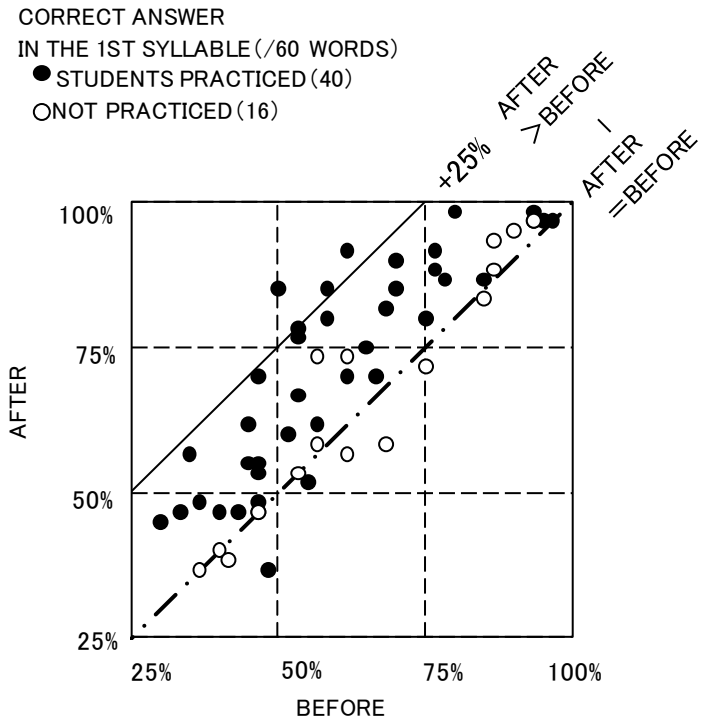
Figure 5-16 Increase in the correct rate of each student

The increase in the correct rate for each student is shown in the figure above. Each circle in the figure represents an examinee. Filled circles represent examinees who have taken practice between two examinations. Open circles represent examinees who have not taken the practice. The horizontal axis is result of the first examination before practicing and the vertical axis is result of the second examination after practicing. The diagonal dotted line in the middle means the same score between two examinations.

For those examinees who have taken practice between two examinations, most of them are above the diagonal dotted line in the middle, which means for most students who have taken practice the correct rate in the second examination are higher than in the first examination. The scope of improvement is different among different levels of examinees. For those high level students, there is less room for improvement. While an obvious improvement has been achieved for middle level students, some of them increased more than 20%. For low level students, there is also much improvement. The result of low level and middle level students has increased an average of 10%.

In contrast, for those examinees who have not taken practice, there is little change between the results of two examinations (open circles are almost along the diagonal dotted line).

3. Increase in the correct rate of each student according to syllable position



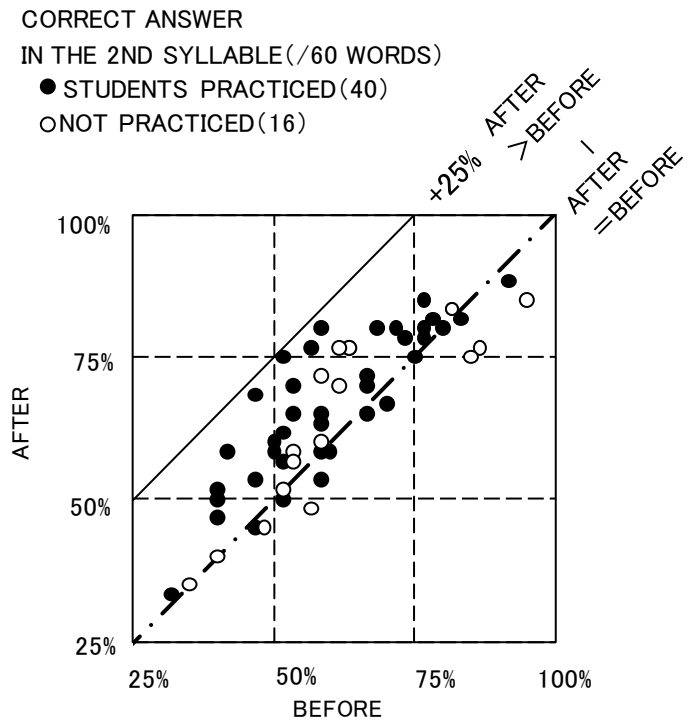


Figure 5-17 Increase in the correct rate of each student according to syllable position

The increase in correct rate by syllable position is shown in the figure above. Each circle in the figure represents an examinee. The horizontal axis is result of the first examination before practicing and the vertical axis is result of the second examination after practicing. The diagonal dotted line in the middle means the same score between two examinations.

From the figure we can see that the improvement tendency is different according to syllable position. In the first syllable, students of all three levels have made obvious improvement. In the second syllable, students of low and middle level still have made obvious improvement, while the improvement of students of high level is not so obvious. The average improvement of correct rate of the first syllable is about 10%, in contrast to 5% for that of the second syllable. CAI system will be helpful to improve the discrimination correct rate, especially for the first syllable.

4. Times and table numbers in practice

The average login times of examinees that have taken practice is 3.5 times, taking 17 tables of practice by average. The total online time is 73 minutes in average. Along

with the increasing of times and tables taken in practice, the improvement of correct rate tends to increase. From the result above, we can confirm that if this CAI system is instructed to students to participate in appropriate stages of curriculum, effective use can be expected.

5.6.2 CHARACTERISTICS OF ANSWERS

In order to provide useful information to support teachers in their teaching process, more examination results need to be gathered and analyzed in various aspects. The characteristics of answers during learning process will also be utilized to improve system function.

Another about one hundred students' practice results of examination have been gathered. The content is four tables in CAI training site [Sunaoka and Hiki, 2005] and the high level, middle level, low level examinees have also been defined as:

High level: correct rate is higher than 75% (more than 90 syllables are correct out of 120 syllables).

Middle level: correct rate is between 50% and 75% (60-90 syllables are correct out of 120 syllables).

Low level: correct rate is less than 50% (less than 60 syllables are correct out of 120 syllables).

1. Relationship of correct rate of two syllables by learner's level

Relationship of correct rate of two syllables by learner's level is shown in below figure. Horizontal axis is correct rate for the first syllable and vertical axis is correct rate for the second syllable. The diagonal dotted line in the middle from lower left to upper right means the same correct rate for both syllables. Each circle in the figure represents an examinee. The filled circles represent those examinees whose correct rate of the

second syllable is higher than that of the first syllable. The open circles represent those examinees whose correct rate of the second syllable is lower than that of the first syllable.

As shown in the figure, for almost all examinees of middle level and low level (43 students), their results have a remarkable tendency that the correct rate of second syllable is higher than that of the first syllable. While for those examinees of high level, the number of students who have higher correct rate in the first syllables (open circles) is increased very much.

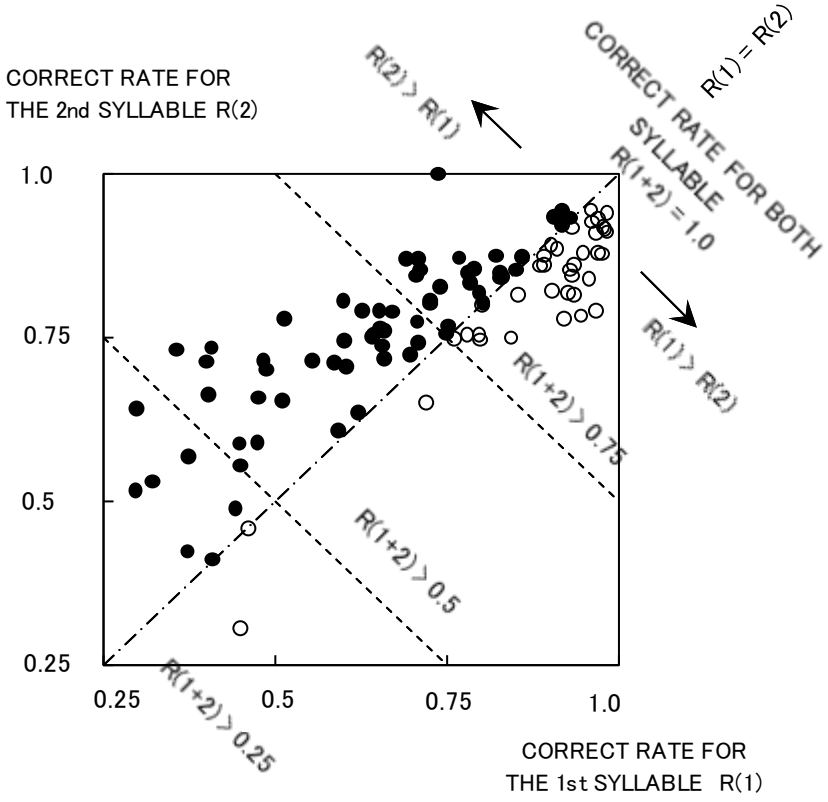


Figure 5-18 Relationship of correct rate of two syllables by learners' level

In the early stage of learning, students still have no enough ability to discriminating tones within a short time. Sound of the second syllable tone is relatively clearer in their temporary memories and this makes it easy to distinguish. According to the observing of teachers during examinations, the scene has often been observed that many students fill answer for the second syllable firstly. In higher stage with enough ability to recognize tones, students can correctly memorize tone patterns in spite of the

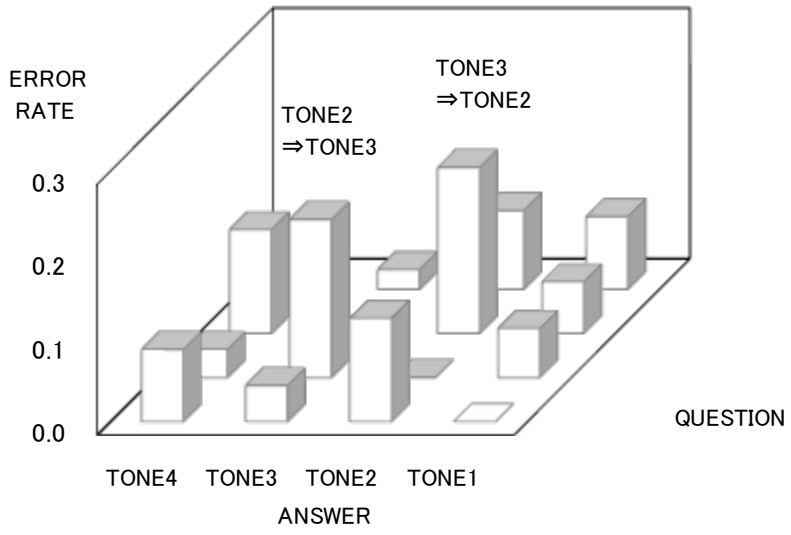
position. Both the first syllable and the second syllable can be distinguished in the same way [Sun et al., 2008b].

Moreover, voice pitch the first syllable is especially disturbed easily by its following syllable. This will make it difficult to distinguish. For those students in elementary stage, teacher needs to focus on improving their discrimination ability especially for the first syllable.

2. Relationship of question and wrong answer

Relationship of question and wrong answer is shown in the figure below. The vertical axis is error rate to syllable units. The horizontal axis is tone in answers and the axis in depth direction is tone in questions. The upper side of figure is data of the first syllable and lower side is data of the second syllable.

From the figure we can see that, in the first syllable, error rate of mistaking tone-2 for tone-3 and mistaking tone-3 for tone-2 are the highest, both are around 20%. Mistakes made in tone-1 and tone-4 are almost half less. In the second syllable, there is a little different. Error rate of mistaking tone-3 for tone-2 is about 20%, similar to that in the first syllable, while error rate of mistaking tone-2 for tone-3 has increased to about 30%, almost 1.5 times. And the mistakes made in tone-1 and tone-4 has reduced to half.



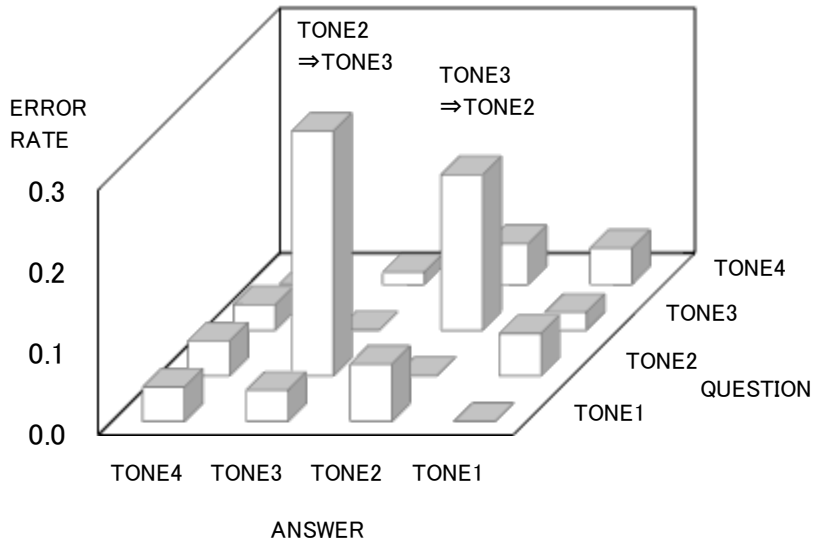


Figure 5-19 Relationship of question and wrong answer

The reason for this phenomenon is that, characteristics of pitch pattern of tone-2 and tone-3 both have the changing from relatively low frequency to high frequency. Comparing to the first syllable, the rising of tone-2 in the second syllable is not so noticeable and similar to that feature of tone-3, therefore it is quite easy to make mistakes from tone-2 to tone-3 in the second syllable [Sun et al., 2008a]. The characteristics of pitch pattern of tone-1 and tone-4 are relatively easy to distinguish. Moreover, the flat feature of tone-1 and falling feature of tone-4 is more obvious in the end of second syllable. In contrast, feature of tone-1 and tone-4 is easily to be disturbed in the transition from the first syllable to second syllable. This is the reason why error rates of tone-1 and tone-4 are relatively higher in the first syllable.

Based on the detailed analysis of examinees' data, we can get the conclusion that particular emphasis should be put on the discrimination of tone-2 and tone-3 in teaching process.

3. Error rate distribution of combination tones for bi-syllabic words

Error rate distribution of tones' combination of bi-syllabic words is shown in the figure below. The vertical axis is error rate to syllable units and the horizontal axis is

combination of tones of bi-syllabic words. The thickest solid line is average error rate of all the four tables.

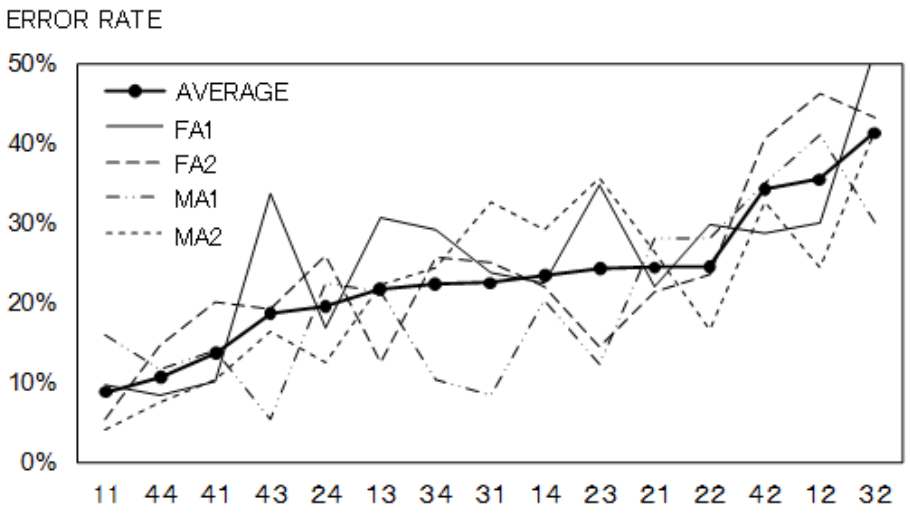


Figure 5-20 Error rate distribution of combination of tones for bi-syllabic words

The average error rates of all combinations are in the range of 10% to 40%, spread to about four times. Error rate of all combinations for each table is not so consistent to other tables (thin solid line and dotted line). Even for the same tone combination, error rate is various according to different words. The variation is about ± 10% from the average error rate.

As seen in figure, tone combinations with only tone-1 and tone-4 have lowest error rate and gathering in the left side. In the right side of high error rate, combinations of tone-2 in the second syllable are gathered together.

4. Relationship of tone discrimination and character studying

In order to figure out the relationship of tone discrimination and character studying (include Chinese character and Pinyin mark), an experiment was carried out in 2008. The targets of experiment were 64 students who had learnt Chinese for about 1 month.

The content was 22 bi-syllabic words that already been taught in elementary textbook. Result of this experiment is shown in the following two figures.

Vertical axis of the first figure is correct rate for Chinese characters to syllable units and horizontal axis is correct rate for Pinyin to syllable units. Vertical axis of the second figure is correct rate for tone discrimination to syllable units and horizontal axis is average correct rate for Chinese character and Pinyin.

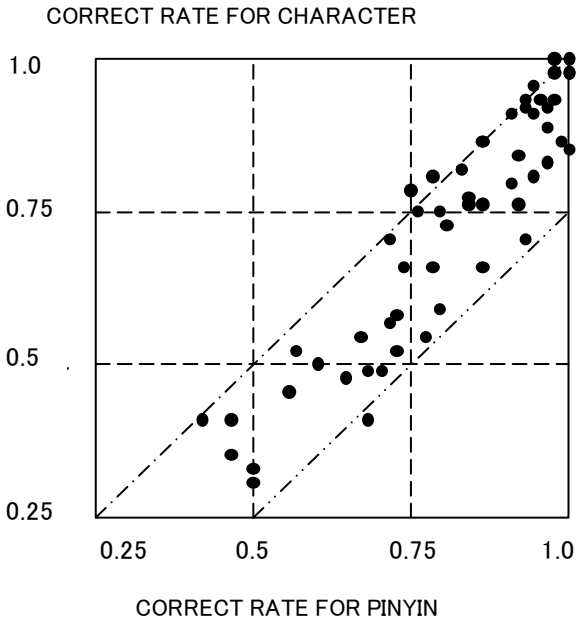


Figure 5-21 Relationship of Chinese character and pinyin

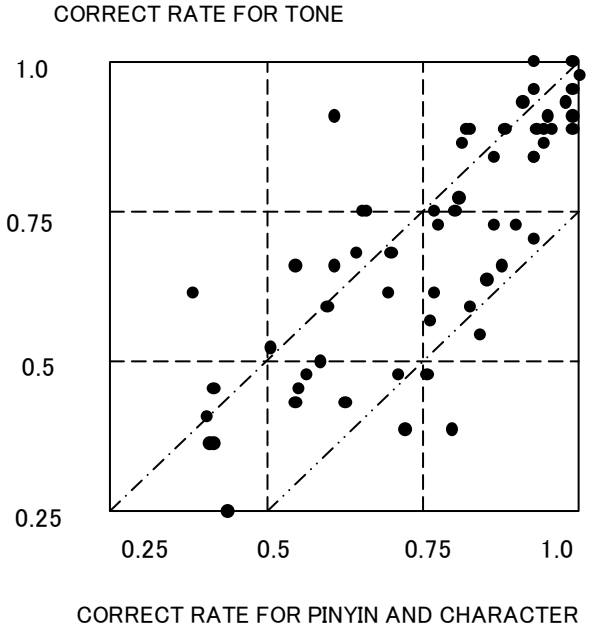


Figure 5-22 Relationship of tone discrimination and character studying

From the result, we can find out that the score of Chinese character, Pinyin and tone discrimination correlates to each other. However, the studying progress of tone is behind studying progress of other two character types. The sequence of studying progress can be illustrated as: tone discrimination < Chinese character < Pinyin. The detail analyzing result of answers shows that, even for those students of high level score (correct rate is higher than 75%), if they make only one type of error in these three types of characters, error rate of tone discrimination is the highest. From the second figure we can also find out that no matter in what levels, because individual differences in auditory sensitivity of sound, students' correct rate of tone discrimination have large variation. Because this CAI system is especially essential in intensively training of tone discrimination, it will serve as an effective supporting tool for teachers.

Several experiments have been carried out from 2005 to 2008. The above characteristics have been verified by a large quantity of testing data gathered by CAI system. These characteristics will be utilized in CAI system improvement and enhancement.

5.7 CONCLUSION

Tone is a great difficulty for Japanese students to study Chinese. How to guide students to master the discriminating skill of tones is one of key points of Chinese language education. In this research, a computer-assisted instruction system for self-teaching of discriminating Chinese four tones has been constructed. Precise acoustical data of Chinese tones have been utilized for designing the CAI system. Inexperience in tonal perception by Japanese students has been taken into consideration in the method of compiling word lists from corpus.

Through the result of examinations, the system’s effects on students have been confirmed. In the early stage of studying of Chinese tones, system will serve as an effective supporting technique of self-teaching for the retarded students to catch up on their study. Moreover, this system can also provide more complete and detailed information to teachers for making effective education plans.

CHAPTER 6.
CONCLUSION AND FUTURE WORKS

Tutorial Chinese has been carried out at Waseda University since 2002. The classes are small, usually consisting of one native speaking tutor and four students. One of the educational goals of this course is to provide more opportunities to Japanese students to learn Chinese with native speakers. The tutors in this course are located in Tokyo, Beijing, and Taipei and employ a conversational style in conducting lessons. In order to support this education model, Tutorial Chinese Platform (TCP) has been constructed. Reasonable corpora have been utilized to construct system functions for more effective language education. With the support of e-Learning system in this research, teachers, tutors, and students are involved in the process of generating, gathering, analyzing, and utilizing of language educational information. This process works as a dynamic cycle with valuable learner corpora being stored and applied.

In this dissertation, I introduced related research fields, including e-Learning, learner corpus, and the current situation of Chinese education at Waseda University in Chapter 1 as the research background. The importance and necessity of the TCP and the objectives of the research was presented.

Chapter 2 described the research proposal. The working mechanism that how to integrate construction of learner corpora and their applications to daily pedagogical activities, and the corpus designed in this research named WTCC was described.

Chapter 3 proposed the Computer Assisted Language Learning (CALL) drill sub-system, which combined lesson report module of TCP with WTCC to collect new

vocabulary used in the classroom. Meanwhile the mechanism of analyzing the corpus and automatically creating exercises had also been implemented. This system had features such as: (a) web-based user interface for students and teachers to facilitate system access, (b) integration with WTCC corpus to make drills more effective, (c) layer architecture to make authoring flexible, and (d) mobile phone-based ubiquitous environment to make the drills accessible to users anytime and anywhere. The mechanism implemented in the CALL drill sub-system had been operated in Tutorial Chinese since 2004 and achieved satisfactory results.

In Chapter 4, the composition/correction sub-system with a corpus retrieval function was described. The concept of this sub-system was to take advantage of the Internet, implement online composition/correction functions similar to that in a paper-based work environment, and simultaneously store the compositions and correction information into the learner corpus through the correction process. The result of the usability evaluation showed that composition/correction sub-system had dramatically increased teachers' working efficiency.

Chapter 5 cited a typical example named the CAI sub-system for self-teaching of discriminating Chinese four tones. Precise acoustical data of the Chinese tones had been utilized for designing the system. Characteristics of the errors found in the examination had been analyzed, and these characteristics had been carefully considered while constructing the system. Examination of the improvement in the students' scores after practicing with the CAI system had clearly confirmed the effectiveness of the system. Moreover, the result of the analysis of the students' answers had provided valuable insights for teachers to make effective education plans.

In the stage of TCP design, there were two objectives need to achieve.

- To integrate construction of learner corpus to daily pedagogical activities of TC efficiently.
- To apply the constructed learner corpus to TC's pedagogy itself effectively.

The TCP had been operating since 2003 and the vitality of its role in Chinese language education at Waseda University has been confirmed by the improvement in students' language skills and the positive results of its usability evaluation. Moreover, teachers had analyzed the constructed learner corpora to understand students' learning aptitude and give appropriate advice to individual students. On the basis of the results

mentioned above, it can be concluded that this research had achieved success in the fields of ICT-based Chinese language learning.

There remain many topics for further research. In the CAI system, the selection of a word corpus is mainly based on teachers' experience. In order to further leverage this system, it is necessary to automate the word corpus selection process. Rules that predict the difficulty of discriminating Chinese words have already been examined. Through observation of the characteristics of incorrect answers by Japanese students, the proposed rules of prediction include 1) properties of confusion among voice pitch patterns in bi-syllabic tones, 2) ill effects of kinds of consonants and vowels involved in each syllable, 3) unclearness of the utterance of speech presented in the system, and 4) degree of unfamiliarity of the words. In the next step, our objective is to make use of these rules to achieve the function of automatically selecting appropriate words for students. Additionally, the composition/correction system should be combined with the CALL drill system to generate sentence-style exercises based on the common errors identified in student compositions.

In this study, the application of the mechanism of e-Learning systems that helps construct and utilize learner corpora for language education has been proposed, which is a new attempt to integrate both the dynamic construction of learner corpora and their applications to daily pedagogical activities through e-Learning systems. The application is just an initial attempt; the long-term goal of this research is to provide a personalized and adaptive learning environment where the system will guide the learner by suggesting relevant learning materials correspond to the learner's proficiency level. I hope that this study will inspire new ideas for further research. More studies are expected in the near future.

APPENDIX

Table list used in CAI system for self-teaching of discriminating Chinese four tones

表Ａ１

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字)	ﾋﾟﾝｲﾝ（声調記号)	第１音節	第２音節	声調	和訳
a10411	参加	can1 jia1	cān jiā	can1	jia1	11	参加/する
a10512	新闻	xin1 wen2	xīn wén	xin1	wen2	12	ニュース
a11413	公里	gong1 li3	gōng l ǐ	gong1	li3	13	キロメートル
a11514	工业	gong1 ye4	gōng yè	gong1	ye4	14	工業
a10321	年轻	nian2 qing1	nián qīng	nian2	qing1	21	若い
a11322	银行	yin2 hang2	yín háng	yin2	hang2	22	銀行
a11023	游泳	you2 yong3	yóu y ǒng	you2	yong3	23	泳ぐ，水泳
a10224	其次	qi2 ci4	qí cì	qi2	ci4	24	次
a10731	有关	you3 guan1	y ǒu guān	you3	guan1	31	関係がある
a10932	鸟笼	niao3 long2	ni ǎ o lóng	niao3	long2	32	鳥かご
a10134	努力	nu3 li4	n ǔ lì	nu3	li4	34	努力/する
a10841	大家	da4 jia1	dà jiā	da4	jia1	41	みんな
a10642	认为	ren4 wei2	rèn wéi	ren4	wei2	42	～と考える
a11143	做法	zuo4 fa3	zuò f ǎ	zuo4	fa3	43	やり方
a11244	对面	dui4 mian4	duì miàn	dui4	mian4	44	真向かい

表Ａ２

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字)	ﾋﾟﾝｲﾝ（声調記号)	第１音節	第２音節	声調	和訳
a20111	初期	chu1 qi1	chū qī	chu1	qi1	11	初期
a20412	周围	zhou1 wei2	zhōu wéi	zhou1	wei2	12	周囲
a21413	西北	xi1 bei3	xī běi	xi1	bei3	13	西北
a20214	阶段	jie1 duan4	jiē duàn	jie1	duan4	14	段階
a20621	原因	yuan2 yin1	yuán yīn	yuan2	yin1	21	原因

a21322	湖南	hu2 nan2	hú nán	hu2	nan2	22	湖南(地名)
a21023	梅雨	mei2 yu3	méi yǔ	mei2	yu3	23	梅雨
a20524	程度	cheng2 du4	chéng dù	cheng2	du4	24	程度
a20731	满分	man3 fen1	mǎn fēn	man3	fen1	31	满点
a20932	假如	jia3 ru2	jiǎ rú	jia3	ru2	32	もしも～なら
a20334	采用	cai3 yong4	cǎi yòng	cai3	yong4	34	採用/する
a20841	目标	mu4 biao1	mù biāo	mu4	biao1	41	目標
a21542	面前	mian4 qian2	miàn qián	mian4	qian2	42	目の前
a21143	地铁	di4 tie3	dì tiě	di4	tie3	43	地下鉄
a21244	附近	fu4 jin4	fù jìn	fu4	jin4	44	付近

表Ａ３

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
a31411	突出	tu1 chu1	tū chū	tu1	chu1	11	特に目立つ
a31212	刚才	gang1 cai2	gāng cái	gang1	cai2	12	先ほど
a30113	欧美	ou1 mei3	ōu měi	ou1	mei3	13	欧米
a31314	发现	fa1 xian4	fā xiàn	fa1	xian4	14	発見/する
a30321	国家	guo2 jia1	guó jiā	guo2	jia1	21	国
a31522	留言	liu2 yan2	liú yán	liu2	yan2	22	伝言/する
a30223	合理	he2 li3	hé lǐ	he2	li3	23	合理的な
a30524	活动	huo2 dong4	huó dòng	huo2	dong4	24	活動/する
a31031	女生	nv3 sheng1	nǚ shēng	nv3	sheng1	31	女子生徒
a30632	否则	fou3 ze2	fǒu zé	fou3	ze2	32	～でないと
a30934	有效	you3 xiao4	yǒu xiào	you3	xiao4	34	有効な
a30841	特殊	te4 shu1	tè shū	te4	shu1	41	特殊な
a30742	内容	nei4 rong2	nèi róng	nei4	rong2	42	内容
a31143	跳舞	tiao4 wu3	tiào wǔ	tiao4	wu3	43	ダンス
a30444	气象	qi4 xiang4	qì xiàng	qi4	xiang4	44	気象

表Ａ４

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
a40311	资金	zi1 jin1	zī jīn	zi1	jin1	11	資金
a41312	支持	zhi1 chi2	zhī chí	zhi1	chi2	12	支持/する
a40113	观点	guan1 dian3	guān diǎn	guan1	dian3	13	見方
a41214	趋势	qu1 shi4	qū shì	qu1	shi4	14	趨勢

a40221	直接	zhi2 jie1	zhí jiē	zhi2	jie1	21	直接
a41422	环节	huan2 jie2	huán jié	huan2	jie2	22	一環
a40723	明显	ming2 xian3	míng xiǎn	ming2	xian3	23	明らかである
a40624	逐渐	zhu2 jian4	zhú jiàn	zhu2	jian4	24	しだいに
a40831	好听	hao3 ting1	hǎo tīng	hao3	ting1	31	聞こえがいい
a41132	举行	ju3 xing2	jǔ xíng	ju3	xing2	32	挙行する
a40934	改变	gai3 bian4	gǎi biàn	gai3	bian4	34	変わる
a40541	用功	yong4 gong1	yòng gōng	yong4	gong1	41	よく勉強する
a41042	问题	wen4 ti2	wèn tí	wen4	ti2	42	問題
a40443	政府	zheng4 fu3	zhèng fǔ	zheng4	fu3	43	政府
a41544	社会	she4 hui4	shè huì	she4	hui4	44	社会
表Ｂ１							
表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
b11511	西安	xi1 an1	xī ān	xi1	an1	11	西安(地名)
b10712	非常	fei1 chang2	fēi cháng	fei1	chang2	12	非常に
b11413	污染	wu1 ran3	wū rǎn	wu1	ran3	13	汚染/する
b11314	失恋	shi1 lian4	shī liàn	shi1	lian4	14	失恋/する
b10421	长江	chang2 jiang1	cháng jiāng	chang2	jiang1	21	長江
b10322	长城	chang2 cheng2	cháng chéng	chang2	cheng2	22	万里の長城
b11223	来往	lai2 wang3	lái wǎng	lai2	wang3	23	往来/する
b10824	服务	fu2 wu4	fú wù	fu2	wu4	24	サービス/する
b11031	烤鸭	kao3 ya1	kǎo yā	kao3	ya1	31	北京ダック
b11132	可怜	ke3 lian2	kě lián	ke3	lian2	32	哀れ/む
b10534	法律	fa3 lu4	fǎ lǜ	fa3	lv4	34	法律
b10141	菜单	cai4 dan1	cài dān	cai4	dan1	41	メニュー
b10242	菜名	cai4 ming2	cài míng	cai4	ming2	42	料理名
b10943	父母	fu4 mu3	fù mǔ	fu4	mu3	43	父母
b10644	饭店	fan4 dian4	fàn diàn	fan4	dian4	44	ホテル
表Ｂ２							
表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
b21411	中医	zhong1 yi1	zhōng yī	zhong1	yi1	11	中国医学
b20212	高明	gao1 ming2	gāo míng	gao1	ming2	12	優れている

b21313	香港	xiang1 gang3	xiāng g ǎ ng	xiang1	gang3	13	香港(地名)
b21014	托运	tuo1 yun4	tuō yùn	tuo1	yun4	14	託送する
b20621	聊天	liao2 tian1	liáo tiān	liao2	tian1	21	雑談する
b20122	潮流	chao2 liu2	cháo liú	chao2	liu2	22	潮流
b21523	昨晚	zuo2 wan3	zuó w ǎ n	zuo2	wan3	23	昨晚
b20324	决定	jue2 ding4	jué dìng	jue2	ding4	24	決定/する
b20731	普通	pu3 tong1	p ǔ tōng	pu3	tong1	31	普通な
b20432	理由	li3 you2	l ǐ yóu	li3	you2	32	理由
b21234	武力	wu3 li4	w ǔ lì	wu3	li4	34	武力
b21141	卫生	wei4 sheng1	wèi shēng	wei4	sheng1	41	衛生
b20942	日元	ri4 yuan2	rì yuán	ri4	yuan2	42	日本円
b20843	日语	Ri4 yu3	Rì y ǔ	Ri4	yu3	43	日本語
b20544	利益	li4 yi4	lì yì	li4	yi4	44	利益
表Ｂ３							
表問声調	簡体字	ﾋﾟﾝｲﾝ(声調数字)	ﾋﾟﾝｲﾝ(声調記号)	第１音節	第２音節	声調	和訳
b30111	餐厅	can1 ting1	cān tīng	can1	ting1	11	レストラン
b30212	超级	chao1 ji2	chāo jí	chao1	ji2	12	スーパー
b31513	生产	sheng1 chan3	shēng ch ǎ n	sheng1	chan3	13	生産/する
b31414	沙漠	sha1 mo4	shā mò	sha1	mo4	14	砂漠
b31321	邻居	lin2 ju1	lín jū	lin2	ju1	21	隣近所/の人
b31122	篮球	lan2 qiu2	lán qiú	lan2	qiu2	22	バスケットボール
b31023	国语	guo2 yu3	guó y ǔ	guo2	yu3	23	国語
b31224	蓝色	lan2 se4	lán sè	lan2	se4	24	青色
b30831	广播	guang3 bo1	gu ǎ ng bō	guang3	bo1	31	放送/する
b30432	改革	gai3 ge2	g ǎ i gé	gai3	ge2	32	改革/する
b30334	反对	fan3 dui4	f ǎ n duì	fan3	dui4	34	反対/する
b30641	故宫	gu4 gong1	gù gōng	gu4	gong1	41	故宮
b30942	桂林	gui4 lin2	guì lín	gui4	lin2	42	桂林(地名)
b30543	各种	ge4 zhong3	gè zh ǒ ng	ge4	zhong3	43	各種の
b30744	挂号	gua4 hao4	guà hào	gua4	hao4	44	登録する
表Ｂ４							
表問声調	簡体字	ﾋﾟﾝｲﾝ(声調数字)	ﾋﾟﾝｲﾝ(声調記号)	第１音節	第２音節	声調	和訳

b40111	飞机	fei1 ji1	fēi jī	fei1	ji1	11	飛行機
b40312	开学	kai1 xue2	kāi xué	kai1	xue2	12	始業/する
b41113	失火	shi1 huo3	shī huǒ	shi1	huo3	13	失火する
b41414	专业	zhuan1 ye4	zhuān yè	zhuan1	ye4	14	専門
b40921	其它	qi2 ta1	qí tā	qi2	ta1	21	ほかの
b40822	零钱	ling2 qian2	líng qián	ling2	qian2	22	小銭
b40723	良好	liang2 hao3	liáng hǎo	liang2	hao3	23	良好な
b41224	同意	tong2 yi4	tóng yì	tong2	yi4	24	同意/する
b40631	老师	lao3 shi1	lǎo shī	lao3	shi1	31	教師
b40232	感觉	gan3 jue2	gǎn jué	gan3	jue2	32	感じ/る
b40434	考试	kao3 shi4	kǎo shì	kao3	shi4	34	試験/する
b40541	辣椒	la4 jiao1	là jiāo	la4	jiao1	41	とうがらし
b41542	自由	zi4 you2	zì yóu	zi4	you2	42	自由な
b41343	下雪	xia4 xue3	xià xuě	xia4	xue3	43	雪が降る
b41044	散步	san4 bu4	sàn bù	san4	bu4	44	散歩/する

表Ｃ１

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
c10412	新闻	xin1 wen2	xīn wén	xin1	wen2	12	ニュース
c11113	公里	gong1 li3	gōng lǐ	gong1	li3	13	キロメートル
c10321	年轻	nian2 qing1	nián qīng	nian2	qing1	21	若い
c11022	银行	yin2 hang2	yín háng	yin2	hang2	22	銀行
c10823	游泳	you2 yong3	yóu yǒng	you2	yong3	23	泳ぐ，水泳
c10224	其次	qi2 ci4	qí cì	qi2	ci4	24	次
c10631	有关	you3 guan1	yǒu guān	you3	guan1	31	関係がある
c10732	鸟笼	niao3 long2	niǎo lóng	niao3	long2	32	鳥かご
c10134	努力	nu3 li4	nǔ lì	nu3	li4	34	努力/する
c10542	认为	ren4 wei2	rèn wéi	ren4	wei2	42	～と考える
c10943	做法	zuo4 fa3	zuò fǎ	zuo4	fa3	43	やり方

表Ｃ２

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
c20212	周围	zhou1 wei2	zhōu wéi	zhou1	wei2	12	周囲
c21013	西北	xi1 bei3	xī běi	xi1	bei3	13	西北
c20421	原因	yuan2 yin1	yuán yīn	yuan2	yin1	21	原因

c20922	湖南	hu2 nan2	hú nán	hu2	nan2	22	湖南(地名)
c20723	梅雨	mei2 yu3	méi yǔ	mei2	yu3	23	梅雨
c20324	程度	cheng2 du4	chéng dù	cheng2	du4	24	程度
c20531	满分	man3 fen1	mǎn fēn	man3	fen1	31	满点
c20632	假如	jia3 ru2	jiǎ rú	jia3	ru2	32	もしも～なら
c20134	采用	cai3 yong4	cǎi yòng	cai3	yong4	34	採用/する
c21142	面前	mian4 qian2	miàn qián	mian4	qian2	42	目の前
c20843	地铁	di4 tie3	dì tiě	di4	tie3	43	地下鉄

表Ｃ３

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
c31012	刚才	gang1 cai2	gāng cái	gang1	cai2	12	先ほど
c30113	欧美	ou1 mei3	ōu měi	ou1	mei3	13	欧米
c30321	国家	guo2 jia1	guó jiā	guo2	jia1	21	国
c31122	留言	liu2 yan2	liú yán	liu2	yan2	22	伝言/する
c30223	合理	he2 li3	hé lǐ	he2	li3	23	合理的な
c30424	活动	huo2 dong4	huó dòng	huo2	dong4	24	活動/する
c30831	女生	nv3 sheng1	nǚ shēng	nv3	sheng1	31	女子生徒
c30532	否则	fou3 ze2	fǒu zé	fou3	ze2	32	～でないと
c30734	有效	you3 xiao4	yǒu xiào	you3	xiao4	34	有効な
c30642	内容	nei4 rong2	nèi róng	nei4	rong2	42	内容
c30943	跳舞	tiao4 wu3	tiào wǔ	tiao4	wu3	43	ダンス

表Ｃ４

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
c41012	支持	zhi1 chi2	zhī chí	zhi1	chi2	12	支持/する
c40113	观点	guan1 dian3	guān diǎn	guan1	dian3	13	見方
c40221	直接	zhi2 jie1	zhí jiē	zhi2	jie1	21	直接
c41122	环节	huan2 jie2	huán jié	huan2	jie2	22	一環
c40523	明显	ming2 xian3	míng xiǎn	ming2	xian3	23	明らかである
c40424	逐渐	zhu2 jian4	zhú jiàn	zhu2	jian4	24	しだいに
c40631	好听	hao3 ting1	hǎo tīng	hao3	ting1	31	聞こえがいい
c40932	举行	ju3 xing2	jǔ xíng	ju3	xing2	32	挙行する
c40734	改变	gai3 bian4	gǎi biàn	gai3	bian4	34	変わる
c40842	问题	wen4 ti2	wèn tí	wen4	ti2	42	問題

c40343	政府	zheng4 fu3	zhèng fǔ	zheng4	fu3	43	政府
表Ｄ１							
表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
d10512	非常	fei1 chang2	fēi cháng	fei1	chang2	12	非常に
d11113	污染	wu1 ran3	wū rǎn	wu1	ran3	13	汚染/する
d10321	长江	chang2 jiang1	cháng jiāng	chang2	jiang1	21	長江
d10222	长城	chang2 cheng2	cháng chéng	chang2	cheng2	22	万里の長城
d11023	来往	lai2 wang3	lái wǎng	lai2	wang3	23	往来/する
d10624	服务	fu2 wu4	fú wù	fu2	wu4	24	サービス/する
d10831	烤鸭	kao3 ya1	kǎo yā	kao3	ya1	31	北京ダック
d10932	可怜	ke3 lian2	kě lián	ke3	lian2	32	哀れ/む
d10434	法律	fa3 lu4	fǎ lǜ	fa3	lv4	34	法律
d10142	菜名	cai4 ming2	cài míng	cai4	ming2	42	料理名
d10743	父母	fu4 mu3	fù mǔ	fu4	mu3	43	父母
表Ｄ２							
表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
d20212	高明	gao1 ming2	gāo míng	gao1	ming2	12	優れている
d21013	香港	xiang1 gang3	xiāng gǎng	xiang1	gang3	13	香港(地名)
d20521	聊天	liao2 tian1	liáo tiān	liao2	tian1	21	雑談する
d20122	潮流	chao2 liu2	cháo liú	chao2	liu2	22	潮流
d21123	昨晚	zuo2 wan3	zuó wǎn	zuo2	wan3	23	昨晩
d20324	决定	jue2 ding4	jué dìng	jue2	ding4	24	決定/する
d20631	普通	pu3 tong1	pǔ tōng	pu3	tong1	31	普通な
d20432	理由	li3 you2	lǐ yóu	li3	you2	32	理由
d20934	武力	wu3 li4	wǔ lì	wu3	li4	34	武力
d20842	日元	ri4 yuan2	rì yuán	ri4	yuan2	42	日本円
d20743	日语	Ri4 yu3	Rì yǔ	Ri4	yu3	43	日本語
表Ｄ３							
表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
d30112	超级	chao1 ji2	chāo jí	chao1	ji2	12	スーパー
d31113	生产	sheng1 chan3	shēng chǎn	sheng1	chan3	13	生産/する

d31021	邻居	lin2 ju1	lín jū	lin2	ju1	21	隣近所/の人
d30822	篮球	lan2 qiu2	lán qiú	lan2	qiu2	22	バスケットボール
d30723	国语	guo2 yu3	guó yǔ	guo2	yu3	23	国語
d30924	蓝色	lan2 se4	lán sè	lan2	se4	24	青色
d30531	广播	guang3 bo1	guǎng bō	guang3	bo1	31	放送/する
d30332	改革	gai3 ge2	gǎi gé	gai3	ge2	32	改革/する
d30234	反对	fan3 dui4	fǎn duì	fan3	dui4	34	反対/する
d30642	桂林	gui4 lin2	guì lín	gui4	lin2	42	桂林(地名)
d30443	各种	ge4 zhong3	gè zhǒng	ge4	zhong3	43	各種の

表Ｄ４

表問声調	簡体字	ﾋﾟﾝｲﾝ（声調数字）	ﾋﾟﾝｲﾝ（声調記号）	第１音節	第２音節	声調	和訳
d40212	开学	kai1 xue2	kāi xué	kai1	xue2	12	始業/する
d40813	失火	shi1 huo3	shī huǒ	shi1	huo3	13	失火する
d40721	其它	qi2 ta1	qí tā	qi2	ta1	21	ほかの
d40622	零钱	ling2 qian2	líng qián	ling2	qian2	22	小銭
d40523	良好	liang2 hao3	liáng hǎo	liang2	hao3	23	良好な
d40924	同意	tong2 yi4	tóng yì	tong2	yi4	24	同意/する
d40431	老师	lao3 shi1	lǎo shī	lao3	shi1	31	教師
d40132	感觉	gan3 jue2	gǎn jué	gan3	jue2	32	感じ/る
d40334	考试	kao3 shi4	kǎo shì	kao3	shi4	34	試験/する
d41142	自由	zi4 you2	zì yóu	zi4	you2	42	自由な
d41043	下雪	xia4 xue3	xià xuě	xia4	xue3	43	雪が降る

REFERENCES

[BNC] The British National Corpus　http://www.natcorp.ox.ac.uk/ (Retrieved 2011/03/01)

[Cross and Hamilton, 2002] Jay Cross and Ian Hamilton (2002), The DNA of e-learning, http://www.gtnspa.it/white paper/DNA of e-learning.pdf, 2002 Internet Time Group, (Retrieved 2011/03/01)

[Feng, 2001] Zhiwei Feng (2001), The History and Current status of Chinese Corpus Research, International Conference on Chinese Computing ICCC2001, pp. 1-15 (In Chinese)

[Granger, 2002] Sylviane Granger (2002), A Bird's view of learner corpus research, Computer Learner Corpora, Second Language Acquisition and Foreign language Teaching. John Benjamins Publishing Company.

[Granger, 2009] Sylviane Granger (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. Corpora and Language Teaching, John Benjamins Publishing Company.

[Hiki et al., 2006a] 比企静雄, 砂岡和子, 劉松(2006), 中国語四声弁別のための CAI による自習システム, PC Conference 2006, 188-192

[Hiki et al., 2006b] Shizuo Hiki, Kazuko Sunaoka, Song Liu(2006), A computer-assisted instruction system for self-teaching of discriminating Chinese tones based of their acoustical properties, 4th ASAJ joint meeting, Honolulu, Hawai, USA.　3168z

[HSK, 2001] 国家漢語水平考試中心（2001）漢語水平詞彙与漢字等級大綱（修訂本), 経済科学出版社

[IFLYTEK, 2003] 中国科学技術大学訊飛信息科技股份有限公司 普通話在線模擬測試和学習系統（標準語発声模擬試験と自習システム）http://www.iflytek.com/Html/cpfw/newyuyin/kypc/pthmn/(　参　照　日 2009.12.30)

[James, 1998] Carl James (1998), Errors in Language Learning and Use—Exploring error analysis, Longman London and New York

[Kawai, 1996] 河合和久(1996), ネットワーク環境を利用した作文教育, 情報処理学会研究報告, コンピュータと教育研究会報告, 96(113):9-16

[Kennedy, 1992] Graeme Kennedy (1992). "Preferred ways of putting things with implications

for language teaching", Jan Svartvik, Trends in Linguistics—Studies and Monographs *65*, MOUTON DE GRUYTER, 335-373

[Liu and Hiki, 2009] 劉 松，比企静雄（2009），CAI自習システムの単語表編集のための2音節中国語声調弁別の困難度の予測規則，電子情報通信学会技術研究報告〔教育工学〕，109:268，57-62

[Liu and Urano, 2003a] Song Liu, Yoshiyori Urano(2003)，Research on Interactive Chinese Language Education Platform Based On PHP and MySQL, International Conference of Internet Chinese Education, ICICE2003, Taiwan, 94-99

[Liu and Urano, 2003b] 劉松,浦野義頼(2003)，インタラクティブな中国語教育のコーパス構築及びサポート・プラットフォームに関する研究，日本中国語学会第５３回全国大会,102-106

[Liu, 2004] Song Liu (2004) Research on the Multimedia Corpus and Support-Platform for Interactive Chinese Language Distance Learning. Master thesis in GITS, Waseda University.

[Liu et al., 2004: a] Song Liu, Kazuko Sunaoka, Yoshiyori Urano(2004), Mobile CALL-drill and Web-test Tools for the Communicative Chinese Conversation Lessons, The 18th Pacific Asia Conference on Language, Information and Computation, Tokyo,101-106

[Liu et al., 2004: b] 劉松,張悦,倪萌,砂岡和子,村上公一,浦野義頼(2004)遠程網絡漢語口語教学管理平台的開発　第四回中国語電子化教学国際会議,数字化対外漢語教学理論与方法研究（清華大学出版社），北京,293-299

[Liu et al., 2004: c] 劉松,砂岡和子,浦野義頼(2004) 日中台遠隔中国語口語レッスン管理プラットフォームの開発-早稲田大学 Tutorial Chinese E-Learning Platform- PC Conference 2004, 142-145

[Liu et al., 2005a] 劉松，砂岡和子，浦野義頼(2005)，コーパスに基づいたCALLドリルシステムの開発,電子情報通信学会総合大会予稿集，D-15-37,180

[Liu et al., 2005b] 劉松 砂岡和子浦野義頼(2005)，與語料庫相結合的中文CALL練習系統的開發,第四屆全球華文網路教育国際会議 International Conference of Internet Chinese Education， ICICE, Taiwan, 272-279

[Liu et al., 2005c] 劉松 砂岡和子 浦野義頼(2005)，日本的手機在中文教育上應用的問題與實踐，二十一世紀中文項目營運策略國際教學研討會 International Symposium: Operational Strategies and Pedagogy for Chinese Language Programs in the 21st Century, Taiwan, 244-250

[Liu et al., 2006a] 劉松，砂岡和子，浦野義頼(2006)，誤用データ統計機能を備えるWEB中国語作文・添削支援システム 「コンピュータ＆エデュケーション」論文集VOL.20，74-79

[Liu et al., 2006b] 劉松，砂岡和子，浦野義頼(2006)，具備誤用統計功能的在線中文作文

批改系統的開発，第五回中国語電子化教学国際会議,数字化対外漢語教学理論与方法研究（清華大学出版社）,香港，256-262

[Liu et al., 2009: a] Song Liu, Douglass J. Scott, Yuuki Kato, & Shogo Kato (2009) Development of a Mobile Phone Research Support System, Japan Society For Educational Technology, Annual Conference, E2a-245-05, pp1009-1010

[Liu et al., 2009: b] Song Liu, Douglass J. Scott, Yuuki Kato, & Shogo Kato (2009) Development of a Mobile Phone-based Data Collection and Analysis System, Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2009 VA: AACE, October 2009, Vancouver Canada

[McEnery and Wilson, 2001] Tony McEnery, Andrew Wilson (2001), Corpus Linguistic 2nd edition. Edinburgh University Press Ltd. ISBN 0-7486-1165-7

[Milton, 1998] Milton, J. (1998), WORDPILOT: enabling learners to navigate lexical universes, In S. Granger and J. Hung (eds), 97-98

[Mukherjee and Rohrbach, 2006:228] Mukherjee, J. and Rohrbach, J. M. 2006, Rethinking applied corpus linguistics from a language pedagogical perspective: New departures in learner corpus research. In Planning, Painting and Gluing Corpora. Inside the Applied Corpus Linguist's Workshop, B. Kettemann & G.Marko(eds),205-232. Frankfurt:Lang.

[Murakami et al., 2003] 村上公一,砂岡和子,劉松(2003),中国語レベル分けのための Web 上での中国語能力診断テスト　日本中国語学会第 53 回全国大会，東京，196-200

[Murakami et al., 2004] 村上公一,砂岡和子,劉松(2004) 発話能力を測定するための中国語 Web テストの開発，日本中国語学会第 54 回全国大会 250-254

[Murakami et al., 2005] 村上公一,砂岡和子,劉松(2005) Computerized Adaptive Testing（CAT）方式的網路中文口語測驗的開發，　第四屆全球華文網路教育国際会議 International Conference of Internet Chinese Education, ICICE, Taiwan, 471-477

[MySQL] http://www.mysql.com (Retrieved 2011/03/01)

[N. Nesselhauf, 2004:125] Nadja Nesselhauf (2004). Learner corpora and their potential for language teaching. How to Use Corpora in Language Teaching. John Benjamins Publishing Company.

[PHP] http://www.php.net (Retrieved 2011/03/01)

[Redhat Linux] http://www.redhat.com (Retrieved 2011/03/01)

[Rosenberg, 2001] Marc Rosenberg 2001 e-Learning: Strategies for Delivering Knowledge in the Digital Age. New York: McGraw-Hill, 28.

[Rossett, A. & Sheldon, K., 2001] Allison Rossett 2001, Beyond the Podium: Delivering Training and Performance to a Digital World. San Francisco: Jossey-Bass/Pfeiffer,

274.

[Sakamoto, 2003] 坂本昂（2003），高等教育改革に資するマルチメディアの高度利用に関する研究，科学研究費特定領域研究 平成 14 年度研究成果報告書

[Shi, 2003] 史有為 （2003） 中国語教育に於ける語彙の位置付けー中国語教育研究明海大学大学院応用言語学研究科紀要『応用言語学研究』No5

[Smrz, 2004] Smrz, P. (2004), Integrating Natural Language Processing into e-learning-A case of Czech. Paper presented at COLING2004.

[Sunaoka, 2002] 砂岡和子(2002)、中国語チュートリアルプログラム開発と発話分析 教養諸学研究第 102 号、319-369

[Sunaoka and Murakami, 2003] 砂岡和子,村上公一（2003)、提升交際能力為目的之漢語教學以及水平測驗芻議-早稻田大學國際漢語遠距教學四年實踐見證 第三届全球華文網絡教育国際会議 台北 1-9

[Sunaoka et al., 2004a] 砂岡和子，李利津，王玉，劉敬華，Tutorial Chinese Text for Communication,(株)東方書店

[Sunaoka et al., 2004b] 砂岡和子,村上公一,浦野義頼,劉松,張悦,倪萌(2004)基于 Active Database 与師生課本互動的国際漢語口語教学， 第四届中文電化教学国際会議,数字化対外漢語教学理論与方法研究（清華大学出版社），北京，328-334

[Sunaoka, 2005] 砂岡和子（2005)，コーパス利用による中国語(2)―可変的語彙教育を支援する中国語コーパス構築とその応用 教養諸学研究 第 180 号 2005-1 147-169

[Sunaoka and Hiki, 2005] 砂岡和子，比企静雄（2005），中国語声調の音響的特性に基づいた四声弁別CAI自習システム，日本中国語学会第55回全国大会予稿集，276-280

[Sunaoka et al., 2005] 砂岡 和子，村上 公一，劉 松（2005)為口語能力提升的中文教學項目量化 二十一世紀中文項目營運策略國際教學研討會 International Symposium: Operational Strategies and Pedagogy for Chinese Language Programs in the 21st Century, Taiwan, 236-243

[Sunaoka and Liu, 2006] 砂岡和子,劉松(2006),誤用データ機能を備える WEB 中国語作文添削支援システム設計と開発，PC Conference 2006，252-256

[Sunaoka, 2008] 砂岡和子(2008)，早稲田大学における多言語遠隔交流とミス・コミュニケーション学習支援プログラムの開発、京都大学国際シンポジウム、大学における外国語教育の二つの挑戦、多言語教育・自律学習ワークショップ

[Sun et al., 2008a] 孫琦,比企静雄,砂岡和子（2008）基于<<CAI 漢語四声聴力自習系統>>進行的声調聴力測評,『数字化漢語教学進展与深化』第六回中国語電子化教学国際会議論文集：474-479

[Sun et al., 2008b] 孫琦，砂岡和子，比企静雄（2008）中国語四声弁別 CAI 自習システム

の利用効用の確認と学習過程の考察．日本中国語学会第 58 回全国大会予稿集，49-53

[Takagi et al., 2005] 都木徹，服部有希子，小宮恵，今井篤，岸憲史，伊藤崇之(2005)，韻律の視覚化及び矯正音声を用いた中国語学習システム，電子情報通信学会論文誌 88-D1:. 2 478-487

[Time Corpus] The Time Corpus, http://corpus.byu.edu/time/ (Retrieved 2011/03/01)

[Tono, 2003] Yukio Tono (2003), Learner corpora: design, development and applications, Proceedings of the Corpus Linguistics 2003 Conference (CL 2003), Lancaster University: University Centre for Computer Corpus Research on Language: 800-809.

[Tono, 2007] 投野由紀夫(2007)，日本人中高生 1 万人の英語コーパス”JEFLL Corpus”：中高生が描く英文の実態とその分析　東京：小学館

[Urasaki and kogo, 1998] 浦崎久美子，向後千春(1998)，データベースを利用した作文授業の実践と評価，電子情報通信学会技術研究報告，ET，教育工学，98(433)：97-104

[Waseda Univ. DCC]　Waseda University, Digital Campus Consortium http://www.waseda.jp/dcc/4th/index.html (Retrieved 2011/03/01)

[WaveSurfer HP]　http://www.speech.kth.se/wavesurfer　(Retrieved 2011/03/01)

[Wible et al., 2001] Wible, D., Huo, C.-H., Chien, F.-Y., Liu,A.&Tsao, N.-L. (2001), A web-based EFL writing environment: Integrating information for learners, teachers and researchers, Computers and Education 37: 297-315

[Yang et al., 2006] YANG Yi and Others (2006), Tentative Ideas of Constructing Chinese Learners' Spoken Corpus, Chinese Language Learning, 2006 vol 03

[Yu et al., 1998] 俞士汶，朱学鋒，王恵，張芸芸(1998) 現代漢語語法信息詞典詳解，清華大学出版社

[Zhang et al., 2004] LanZhang, SongLiu, JianmingWu, Yoshiyori Urano (2004),Chinese Language Learning Systems with 3G Mobile Phones, Asian Info-communications Council, AIC 2004,　ShenZhen, China. No. 91

[Zhang et al., 2004] 張宝林，崔希亮，任杰 (2004) 関于“HSK 動態作文語料庫”的建設構想，第三届全国語言文学応用学術研討会論文集

LIST OF PUBLICATIONS

	Title, Publication, Date, Authors
論文誌論文	○ 中国語四声弁別の自習のための CAI システム, 日本教育工学会論文誌, 34 巻－3 号, pp 223-233, 2010/12, 劉松, 浦野義頼, 比企静雄
論文誌論文	○ 誤用データ統計機能を備える WEB 中国語作文・添削支援システム, コンピュータ利用教育協議会論文誌, VOL.20, pp 74-79, 2006/06, 劉松, 砂岡和子, 浦野義頼
国際学会	○ Development of a Mobile Phone-based Data Collection and Analysis System Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2009 VA: AACE, pp2881-2886,October 2009, Vancouver Canada, Song LIU, D.J. SCOTT, Y. KATO, and S. KATO
国際学会	○ 具備誤用統計功能的在線中文作文批改系統的開発, 第五回中国語電子化教学国際会議 「数字化漢語教学理論与方法研究」, pp437-442, 香港, 2006/07, 劉松, 砂岡和子, 浦野義頼

国際学会	○ 与語料庫相結合的中文 CALL 練習系統的開発, The 4th International Conference of Internet Chinese Education, pp272-279, Taipei, 2005/06, 劉松, 砂岡和子, 浦野義頼
国際学会	与評奨和 Self-Editing 相結合進行的網絡華文写作教学, The 4th International Conference of Internet Chinese Education, pp058-065, Taipei, 2005/06, 砂岡和子, 村上公一, 劉松
国際学会	Computerized Adaptive Testing (CAT)方式的網路中文口語測験的開発, The 4th International Conference of Internet Chinese Education, pp471-477, Taipei, 2005/06, 村上公一, 砂岡和子, 劉松
国際学会	○ 日本的手機在中文教育上応用的問題与実践, Operational Strategies & Pedagogy for Chinese Language Programs in the 21st Century: An International Symposium, pp244-249, Taipei, 2005/06, 劉松, 砂岡和子, 浦野義頼
国際学会	為提昇口語能力的中文教学項目量化, Operational Strategies & Pedagogy for Chinese Language Programs in the 21st Century: An International Symposium, pp236-243, Taipei, 2005/06, 砂岡和子, 村上公一, 劉松
国際学会	○ Mobile CALL-drill and Web-test Tools for the Communicative Chinese Conversation Lesson, IWLeL 2004:An Interactive Workshop on Language e-Learning, The 18th Pacific Asia Conference on Language, Information and Computation, pp79-85, Tokyo, 2004/12, Song Liu, Kazuko SUNAOKA, Yoshiyori URANO

国際学会 査読なし	Chinese Language Learning Systems with 3G Mobile Phones, AIC, Asian Info-communications Council, Document No. 91, ShenZhen, China, 2004/11, Lan ZHANG, Song LIU, Jianming WU, Yoshiyori URANO
国際学会	○ 遠隔漢語口語教育管理平台的開発, 第四回中国語電子化教学国際会議 「数字化漢語教学理論与方法研究」, pp293-299, 北京, 2004/07, 劉松, 張悦, 倪萌, 砂岡和子, 村上公一, 浦野義頼
国際学会	基于 Active Database 与師生課本互動的国際漢語口語教学, 第四回中国語電子化教学国際会議 「数字化漢語教学理論与方法研究」, pp328-334, 北京, 2004/07, 砂岡和子, 村上公一, 劉松, 張悦, 倪萌, 浦野義頼
国内 研究会	○ CAI自習システムの単語表編集のための2音節中国語声調弁別の困難度の予測規則, 電子情報通信学会技術研究報告〔教育工学〕, 109:268, pp57-62, 2009/11, 劉松, 比企静雄
国内大会	Development of a Mobile Phone Research Support System, Japan Society For Educational Technology, Annual Conference, E2a-245-05, pp1009-1010, 2009/09, Song LIU, J. Scott DOUGLASS, Yuuki KATO, Shogo KATO
国内大会	誤用データ統計機能を備えるWEB中国語作文添削支援システム設計と開発, PC Conference 2006大会論文集, pp391-394, 2006/07, 砂岡和子, 劉松
国内大会	Corpusに基づいた多端末向けCALLドリルシステムの開発, PC Conference 2005大会論文集, pp299-300, 2005/08, 劉松, 砂岡和子, 浦野義頼

国内大会	コーパスに基づいたCALLドリルシステムの開発, 電子情報通信学会総合大会論文集D-15-37, pp180, 2005/03, 劉松, 砂岡和子, 浦野義頼
国内大会	日中台遠隔中国語口語レッスン管理プラットフォームの開発－ 早稲田大学Tutorial Chinese E-Learning Platform, PC Conference 2004大会論文集,pp142-145, 2004/08, 劉松, 砂岡和子, 浦野義頼