



HOKKAIDO UNIVERSITY

Title	A Study on English Language Tutoring System Using Code-Switching Based Second Language Vocabulary Acquisition Method
Author(s)	Mazur, Michal
Degree Grantor	北海道大学
Degree Name	博士(情報科学)
Dissertation Number	甲第12195号
Issue Date	2016-03-24
DOI	https://doi.org/10.14943/doctoral.k12195
Doc URL	https://hdl.handle.net/2115/61833
Type	doctoral thesis
File Information	Mazur_Michal.pdf



学位論文

Doctoral Dissertation

**A Study on English Language Tutoring System Using
Code-Switching Based Second Language Vocabulary
Acquisition Method**

(コードスイッチングに基づく第二言語習得手法を用いた英語チュータ
リングシステムに関する研究)



Michal MAZUR

ミハウ・マズル

Graduate School of Information Science and Technology

北海道大学大学院情報科学研究科

Division of Media and Network Technologies

Hokkaido University

January 2016

Thesis Abstract

These are challenging times for education, because of the increased demands of the global society to acquire new skills. Knowing second language is one of the abilities we have to master in order to freely communicate with the global society. However, the current educational establishments are not always ready to prepare students for the challenges of the modern world. Recent technological innovations are slowly changing this situation, and bring some motivation for finding new ways of teaching and enhancing existing learning methods. Vocabulary is considered to be the most important part of second language education.

This thesis approaches the problem of second language vocabulary acquisition and presents a work on a method, and then its implementation in a computer system to improve vocabulary acquisition for second language students. Co-MIX Japanese, a teaching system using code-switching method for English vocabulary acquisition, is a research project aimed at creating such a second language vocabulary teaching system. The background of the research and the underlying theories are described, along with the summary of the progress and some implications on how they can be used to further advance the project. Each part of the system is described and the way how the presented method can be applied to facilitate second language vocabulary acquisition is explained. Also, various ways of increasing users' engagement are presented.

To the author's best knowledge there exists no robust methodology or teaching application for code-switching-based vocabulary acquisition in e-learning. Therefore, investigating how code-switching could be utilized in second language vocabulary acquisition and utilized in e-learning is a central contribution of this thesis. In

addition, the code-switching effect on second language vocabulary acquisition has not yet been implemented in the field of computer science.

The significance of this research comes from the fact that a code-switching-based vocabulary acquisition method is presented for the first time in the context of an e-learning system, with evidence to support the effectiveness of the presented method. Experiments aimed to evaluate system's performance showed that Co-MiX Japanese system with code-switching method was scored higher than the baseline that lacks the proposed method. Additionally, the preliminary evaluation of a second system, designed for English-speaking users who want to study Japanese vocabulary, was performed and the experiment results confirmed that presented method can also be adapted in other languages. These results emphasize both the usefulness and need for this system and provide ideas for future improvements.

学位論文内容の要旨

Japanese Abstract

教育は現在、グローバル化した社会の新しいスキルを求めるニーズの拡大に伴い、試練の時を迎えている。世界中の人々と意思疎通するために、語学力は重要な能力の一つである。一方、現在の教育制度はこの要求に常に対応できている訳ではなく、学習者は練習問題の反復を余儀なくされる。しかし、近年の技術革新は、この状況を徐々に変えつつある。そこで、著者は、このような状況から本学位論文において、これまでの学習方法を改善するための新たな手法の提案を行う。外国語学習の最も重要な要素とは、語彙習得である。本学位論文は外国語の語彙習得という問題に取り組み、外国語学習者の語彙習得を改善するシステムの開発を提案する。

語彙力は外国語での意思疎通にとって不可欠なスキルであり、語彙習得は外国語をマスターするための最も重要な課題の一つである。本研究は、コード・スイッチング現象を用いたe-ラーニングシステムという新しい語彙習得方法の提案及びその有効性の確認を目的としている。本提案手法によって開発された実験システムでは、一つの言語のある単語が、もう一つの言語に属する文法構造を持つ文書の中で用いられる。このことは、学生に新しい単語について深く考える機会をもたらし、それゆえ第二言語の語彙を増やすことに役立つ。この技術は、今日まであまり関心を集めてこなかったが、近年、心理学者や言語学者、さらにコンピュータサイエンス研究者のような様々な分野の数多くの研究者の関心を集めている。

本学位論文では、言語習得のためのコード・スイッチングを利用した英語教育システム(Co-MIXシステムと呼ぶ)を提案する。このようなシステムは、言語チューター会話システムと考えることができる。本学位論文においては、その背景・基礎となる諸理論を説明し、それと共にこのプロジェクトをさらに展開するために、これまでの研究成果をどこまで、またどのように利用すべきなのかについて論じる。提案システムの各部について説明を行い、第二言語の語彙習得を促進するために、本手法がどのように適用されるのかについて明らかにする。さらに本提案システム全体がどのように学習者の興味を深めているのか、またどのようにシステムの利用に対する関心を高めているのかに関して、様々な手法を提案する。

著者の知る限り、e-ラーニングにおけるコード・スイッチングに基づく語彙習得のための基本的な方法論も教育用アプリケーションもこれまで存在しない。したがって、コード・ス

スイッチングを第二言語の語彙習得においていかに活用するかに関する研究が、本学位論文による貢献の中心となる。さらに、コード・スイッチングが第二言語の語彙習得に及ぼす影響は、コンピュータサイエンスの分野では未だに研究されていない。本研究の重要性は、コード・スイッチングに基づく語彙習得が、その手法の有効性を論証する証拠とともにe-ラーニングシステムという文脈において初めて報告されたという点にある。本提案システムのパフォーマンスを評価するために行われた評価実験では、コード・スイッチングを用いたCo-MIXシステムが、本学位論文で提案する手法を持たないベースラインよりも一般的に優れた成果を取めたことを示している。さらに、日本語の語彙を習得したい英語話者を対象としたもう一つの別の実験システムの初期評価も行われ、この実験結果もまた英語以外の言語に本手法が適用し得るということを実証している。

Table of Contents

Abstract	ii
List of Tables	viii
List of Figures	ix

1 Introduction

1.1 Research background.....	1
1.2 Purpose and novelty of research.....	3
1.3 Thesis structure.....	5

2 Conversational Agents

2.1 Introduction.....	7
2.2 Chatterbots: past and present.....	8
2.3 Task and non-task oriented classification.....	10
2.4 Chatterbots with occupation.....	13
2.5 Research idea.....	16
2.5.1 Baseline chatterbots.....	16
2.5.2 System components.....	17
2.6 Conclusions.....	22

3 Second Language Vocabulary Acquisition

3.1 Introduction.....	25
3.2 High frequency words.....	26
3.3 Code-switching.....	27
3.4 Incidental learning and extensive reading.....	31
3.5 Teaching vocabulary.....	33

4 Emotions and Learning

4.1 Introduction.....	35
4.2 General information.....	36
4.2.1 Background.....	36
4.2.2 The role of emotions in learning.....	37
4.3 The idea of an English tutoring system.....	38
4.3.1 Aim.....	38
4.3.2 Code-switching method.....	39

4.4	Preliminary experiment.....	39
4.4.1	Purpose.....	39
4.4.2	Method.....	40
4.4.3	Results.....	42
4.4.4	Discussion.....	44
4.5	Conclusions	46
5	The Co-MIX Project	
5.1	Introduction.....	47
5.2	Co-MIX Japanese.....	48
5.2.1	System introduction.....	48
5.2.2	Systems used.....	49
5.2.3	Study material generator.....	50
5.2.4	Language quiz generator.....	52
5.3	Evaluation experiments.....	56
5.3.1	Participants.....	56
5.3.2	Procedure.....	56
5.3.3	Experiment method.....	57
5.3.4	Experiment results.....	59
5.3.5	Participants' attitudes on the SD scale.....	59
5.3.6	Factor analysis.....	61
5.3.7	User rankings.....	64
5.3.8	Analysis of user mistakes.....	64
5.4	Discussion.....	66
5.5	Co-MIX English.....	67
5.5.1	System introduction.....	67
5.5.2	Study material generator.....	69
5.5.3	Language quiz generator.....	71
5.6	Evaluation experiments.....	73
5.6.1	Participants.....	73
5.6.2	Semantic Differential scale.....	74
5.6.3	Experiment results.....	75
6	Conclusions and Future Works	
6.1	Overall conclusions.....	77
6.2	Future works.....	78
	References.....	80
	Appendixes.....	90
	Research achievements.....	102
	Acknowledgements.....	105

List of Tables

Table 1. The percentage of correct answers for each question in both phases of the experiment.....	43
Table 2. The results of preliminary experiment compared in both test phases and presented by question number and percentage of correct answers.....	43
Table 3. Factor analysis results for proposed system. The highest significance in bold font.....	61
Table 4. Factor analysis results for baseline system. The highest significance in bold font.....	62
Table 5. Average number of mistakes for proposed and baseline systems. Lower number in bold.....	64

List of Figures

Figure 1. The overview of code-switching-based system with modules.....	18
Figure 2. Gamification-based point system to encourage users' interaction.....	21
Figure 3. Comparative user results of target vocabulary comprehension for all participants in both tests.....	44
Figure 4. Process of study material generator (Co-MIX Japanese).....	48
Figure 5. Flow process of the language quiz generator (Co-MIX Japanese).....	50
Figure 6. Example of the output file of study material generator (Co-MIX Japanese).....	51
Figure 7. Screenshot of system interface (Co-MIX Japanese).....	53
Figure 8. Process of third phase of the test (Co-MIX Japanese).....	54
Figure 9. Result file of the language quiz generator (Co-MIX Japanese).....	55
Figure 10. Semantic differential scale for proposed and baseline systems (Co-MIX Japanese).....	60
Figure 11. Process of study material generator (Co-MIX English).....	69
Figure 12. Process of language quiz generator (Co-MIX English).....	70
Figure 13. Screenshot of system interface (Co-MIX English).....	72
Figure 14. Output of Language Quiz Generator (Co-MIX English).....	73
Figure 15. Semantic differential scale for proposed and baseline systems (Co-MIX English).....	75

Chapter 1

Introduction

This thesis summarizes the results of the research in second language vocabulary acquisition. This chapter serves to introduce the thesis. The contents of the chapter are as follows: section 1.1. introduces the background of the research; section 1.2 describes the contribution made by the research. Finally, section 1.3. describes the structure of this thesis.

1.1 Research background

With the increasing demand from a global society, learning foreign languages has become a necessary skill in the modern world. According to Crystal (Crystal, 2012), English is most widely taught as a second language. It also has the the highest number of second language speakers and currently is the third most spoken language in the world¹. In Japan, English ability is increasingly valued as a global communication skill². It is useful in business situation, required to get into university and helpful in order to communicate with an increasing number of foreign visitors³. The Japanese society has become more aware of the benefits for knowing a second language. In a survey concerning the motivation for learning English among the Japanese (Honna, 2008), 79.3% of respondents said they would like to be able to use English. Furthermore, 83.1% of respondents stated that they would like to use English

¹ Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2015. *Ethnologue: Languages of the World, Eighteenth edition*. Dallas, Texas: SIL International. Online version. Retrieved from: <http://www.ethnologue.com> (accessed 2016/1/3).

² *The Concept of Global Human Resource Development Focusing on the East Asian Region*. Retrieved from: <http://www.mext.go.jp/english/highered/1303540.htm> (accessed 2015/3/14).

³ *No. of foreign visitors to Japan in 2014 highest ever*. Retrieved from: <http://www.japantoday.com/category/national/view/no-of-foreign-visitors-to-japan-in-2014-highest-ever> (accessed 2014/2/4).

when traveling, 79.0% of respondents would like to communicate with foreigners in English, and 58.6% of survey participants would like to be able to read books in English.

Teaching English has become mandatory in elementary schools for fifth- and sixth- graders since the 2011 school year. However, the educational establishments are not ready to face this growing challenge, which lead to many problems⁴. In spite of emphasis placed on the necessity of English communication from the Ministry of Education, Culture, Sports, Science & Technology in Japan, the expectations of English teachers towards students have been very low. In a survey presented by Honna (2008), primary school teachers in Japan were asked about “the minimum English ability students should be equipped before they graduate from the compulsory education and join the nation’s workforce”. According to the results, approximately half of survey respondents concurred that it is enough for students to “be able to exchange greetings and engage in simple interactions”.

Recent technological innovations bring the chance to enhance existing learning methods. Vocabulary is considered to be one of the most important parts of second language education (Ling, 2010). Over the years, many techniques have emerged to facilitate second language vocabulary acquisition, but none of them has proved to be universally effective. The outdated approach to second language vocabulary acquisition can be noticed among Japanese students. They need to memorize a great number of English vocabulary, usually without a chance to elaborate on what these words actually mean. The students mostly use traditional memorization techniques, like flashcards or rote learning. While there is no doubt that these methods may be useful for simple, test-oriented memorization, the retention of

⁴ *Are schools ready for English?* Retrieved from: <http://www.japantimes.co.jp/news/2011/02/26/national/are-schools-ready-for-english/> (accessed 2013/5/27).

vocabulary studied through simple memorization is rather poor for the long-term development of vocabulary. Studying new words in second language is considered to be an on-going process that takes time and effort (Nation, 2001). Without enough practice, the students tend to forget new words or do not know how to properly use them in a sentence. As a result, many Japanese students lose motivation to study English and usually learn no more than is needed to pass the examination requirements⁵.

New possibilities for second language vocabulary acquisition can be found in code-switching phenomenon in which a word from one language is used in a sentence whose structure belongs to another language. It brings a chance for students to think deeper about new vocabulary and has a potential for a longer-term expansion of second language (L2) vocabulary. In the past, this method has been paid less attention to, but recently it has become a point of interest of researchers in different fields, like psychologists (Macaro, 2001), linguists (Coady & Huckin, 1997) and computer scientists (Labutov & Lipson, 2014). There is also an increasing interest in code-switching between Japanese and English languages (Namba, 2012; Winch, 2015). The phenomenon of code-switching is described and compared to other popular techniques described in Chapter 3.

This thesis approaches the problem of vocabulary acquisition among second language learners. The code-switching-based vocabulary acquisition method for e-learning systems is proposed, and implemented as the Co-MIX Japanese system. Co-MIX is an abbreviation of the term *code mixing*, which is used as an interchangeable term for code-switching phenomenon (see subsection 3.3). Co-MIX Japanese is a teaching system using code-switching for English vocabulary acquisition and

⁵ *What's wrong with the way English is taught in Japan*. Retrieved from: <http://www.japantimes.co.jp/opinion/2009/02/05/commentary/whats-wrong-with-the-way-english-is-taught-in-japan/> (accessed 2012/7/3).

designed for Japanese users. However, to confirm the expandability of the proposed system, the preliminary experiments were conducted with a prototype system for English speaking users, who want to study Japanese vocabulary. These experiments gave some evidence that the presented code-switching-based method can also be adapted in other languages.

1.2 Purpose and novelty of research

This research presents work on implementation of code-switching method, an innovative learning approach that can give students a chance to expand their lexical knowledge and utilize the advantages of the code-switching phenomenon, in an e-learning application. To the author's best knowledge there exists no robust methodology or teaching application for code-switching based vocabulary acquisition in e-learning. In addition, the code-switching effect on second language vocabulary acquisition has not been studied in the field of computer science.

The novelty of this research comes from the fact that a code-switching based vocabulary acquisition method is presented for the first time in the context of an e-learning system, with evidence to support the effectiveness of my proposed method. The design of Co-MIX Japanese system that utilize this method is described, and the results of experiments to evaluate the usefulness of presented learning application in learning second language vocabulary are discussed.

Another novelty of this contribution is found in combining the phenomenon of code-switching with several other existing methods of teaching vocabulary (see Chapter 3), and implementing this method as an e-learning software to facilitate second language vocabulary acquisition. An e-learning application should be equipped with a function to encourage users to actively participate in a learning task or increase their motivation to study. Therefore, the results of the previous study

regarding the method of increasing user engagement (Mazur et al., 2011) are utilized by introducing various language quizzes to strengthen users' motivation. In this research users' attitudes towards the proposed system are measured to prove that it is well accepted by the users and properly designed for applying code-switching method in an e-learning system. Moreover, with research results presented in Chapter 5 we provide evidence showing that the users find presented research relevant and useful.

Another issue discussed in this work is the issue of creating valuable study materials for second language teaching. The presented research tackles the problem and follows the general idea of an exercise generator presented by Ginsburg (2012). A new method for assessing the second language vocabulary proficiency level of the user with high frequency word lists is presented. The assessment results are used by proposed system to choose the learning exercises appropriate to the user's language ability. The idea is described wider in Chapter 5.

The main research question posed in this research is as follows: "Since there seems to be an agreement on the positive effects of code-switching in learning, can this phenomenon be successfully employed in e-learning to expand students' second language vocabulary?". In order to address the research question two sub-questions are considered: The first one states: "Can the code-switching method enable students to understand words without given definitions?". The other one focuses on answering how the users respond to presented method and is formulated as: "What are the user attitudes towards the proposed system?".

This thesis also addresses the problem of encouraging users' interaction with a system and motivating them to use it for a long period of time. The problem of maintaining users interest with computer systems has become one of the most important challenges not only in the field of e-learning but in a variety of other fields,

e.g, the research in the field of human-computer interaction presented by Dybala et al. (Dybala, Ptaszynski, Maciejewski, Takahashi, Rzepka, & Araki, 2010).

1.3 Thesis structure

This thesis consists of 5 chapters. Chapter 1 serves to introduce the thesis. Chapter 2 reviews the field of conversational agents in greater detail, with a focus on teaching aspects of chatterbots in context of presented research. The chapter reviews the state of the art in this field and presents the wider vision of a research towards creating a conversational agent which utilizes code-switching phenomenon. Chapter 3 describes theoretical background of the research. It includes the underlying theories and relevant work that contributed to my research. Chapter 4 presents the foundation of my research in creating an e-learning system that utilizes code-switching method for second language vocabulary acquisition. It includes the preliminary experiment which also explores the role of emotions in studying. Chapter 5 describes implementations of the proposed code-switching-based method in two e-learning applications, designed to teach second language vocabulary. The method, the architecture of both systems and the effectiveness of the proposed method is described and verified through evaluation experiments. The experiment results and their implications for this research are also discussed. Finally, Chapter 6 concludes the doctoral dissertation, discusses future works and further perspectives of using the proposed code-switching method.

Chapter 2

Conversational agents

2.1 Introduction

In 1950 Alan Turing started his paper with famous question: “Can machines think?” (Turing, 1950). After this publication, the possibility for machines to think started to be explored by a newly founded scientific field of Artificial Intelligence (AI). A number of technologies that tried to attempt answering Turing’s question have emerged in Computer Science over the following years. Dialogue systems are an example of technology that made interaction between human and machine possible. Weizenbaum’s ELIZA (Weizenbaum, 1966) became the inspiration for linguistic researchers all over the world in creating a computer program that can understand and respond to the natural language.

Nowadays, conversational systems integrate computational linguistics techniques with the widely spread communication over the Internet to interpret and respond to statements made by users in natural language, as it is pointed in work by Lester et al (Lester, Bratning, Mott, 2004). Brennan describes chatterbots as “artificial constructs that are designed to converse with human beings using natural language as input and output” (Brennan, 2006). The typical usage involves receiving the user utterance in a given human language and providing the possibly reasonable or intelligent response to the given sentence. Then, the sequence is repeated as long as user keeps the conversation going. Since their early days, chatterbots have become more sophisticated along with maturing technology and able to respond to user utterance in both text and synthesized voice speech. Moreover, they evolved from

simple text environment into modern visual interfaces that incorporate human-like avatars or virtual presence through augmented reality interfaces (Matthias, Andre, 2005).

2.2 Chatterbots: past and present

Since their origin chatterbots were designed to perform certain functions or, as it could be said, perform certain roles or occupations. ELIZA, simple chatterbot that still influences the research in that field, used simple pattern matching algorithm to find a corresponding response to a given pattern. Despite of its simplicity, this program also performed a role, being a Rogerian psychotherapist able to answer users' questions without any traces of its own personality. ELIZA was later criticized for its lack of an internal world model that could influence and track conversation. In other research, a chatterbot named PARRY developed by Colby in 1975 simulated paranoid behavior (Colby, 1975). By using different tricks like admitting ignorance, changing the level of the conversation, rigidly and continuing previous topics it successfully fooled its human judges into believing that he is one of the patients showing typical paranoid behavior. That sufficiently explained its unusual responses that would be attributed to illness and make its appearance real.

Some progress in chatterbot technology came in 1995 when Wallace introduced Artificial Linguistic Internet Computer Entity (ALICE). This chatterbot, often described as modern ELIZA, is an open source project that became the prototype of many current chatterbots. When ELIZA held no memory of the conversation and its knowledge base was embedded right into the code, ALICE introduced Artificial Intelligence Markup Language (AIML) used to store the knowledge-based data. This solution responded to the existing necessity of a bigger knowledge base. The chatterbot created by Wallace has won Loebner's annual

competitions that awards prizes to the most human-like chatterbots, which is the realization of Turing Test⁶.

Exploiting natural language techniques to build conversational systems has been used for a broad range of applications. Nowadays, dialogue systems are deployed on commercial websites to respond to customers' inquiries about provided services. They also answer questions about financial services or verbally demonstrate the company portfolios. Many companies, such as VirtuOz², use conversational systems for customer support. They also have a potential to assist the patients as virtual physicians/doctors (Lokman, Zain, 2009), serve as bully and harassment advisors (Latham, Crockett, Bandar, 2010) or act as storytellers (Chamberlain, 1984). Some chatterbots are created purely for entertainment purposes, deployed in video games to inform players about in-game events (Hingston, 2009) or as companions assisting players in a virtual environment (Camilleri, Montebello, 2008). Finally, educational conversational systems assist students by providing guidance as they learn, i.e., system presented by Jia (Jia, 2009).

In recent times, chatterbots research has been drawn towards embodied conversational systems where body language and gestures are almost of the same importance as the natural language dialogue (Matthias, Andre, 2005). This new approach is motivated by necessity of making conversational systems appear more human like, thus become more engaging with the user. However, when a significant proportion of chatterbots research is concentrating on evaluating such interfaces, evaluating the quality of actual conversation seems to be a research challenge. Dybala et al. (Dybala, Ptaszynski, Rzepka, & Araki, 2009) stated that in case of task oriented systems it is much easier to achieve user satisfaction because of the very existence of task that requires an interaction between human and system to be completed. In case

⁶ <http://www.loebner.net/Prizef/loebner-prize.html>

of non-task oriented dialogue, the interaction depends more on the content of the conversation.

2.3. Task and non-task oriented classification

Chatterbots are often classified into two conventional groups. The first one often referred to as non-task oriented or free talking, considers the ability of program to chat freely on any given topic, without any obligations, mostly for entertainment purposes. In opposition to non-task oriented systems are task-oriented, specialized systems which put some limitations to system functionalities and usually aim to achieve specific goals. Most previously built conversational systems can engage in either task oriented dialogues to better understand human utterances or non-task oriented dialogues to allow users to enjoy the conversation.

To the author's best knowledge, even though there have been some attempts to provide such conversational systems, there are not many existing ones that would successfully employ both functions. A system presented by Nakano et al. (Nakano, Hoshino, Takeuchi, Hasegawa, Torii, Nakadai, Kato, & Tsujino, 2006) can dynamically change dialogue strategy based on speech recognition results. This system features different control modules called experts and each of it is dedicated to perform different kind of task. Non-task oriented functionality is incorporated into this model as one of the modules, the chat expert, which employs standard chatterbot technology (ALICE). This integrated system for both kinds of dialogues can adaptively change the strategies and more accurately respond to human utterances. However, the authors mentioned that it is important to improve chat expert and operate on a larger base of vocabulary – currently limited due to the problems with voice recognition accuracy.

Crocket et al. (Crockett, O'Shea, Bandar, 2011) presented an overview of methodology for constructing goal oriented conversational systems (GO-CA) with a mixed-initiative strategy (from time to time either human or chatterbot may take control of the conversation). By goal-oriented they understand a deep strategic purpose of conversation and directing it to achieve a certain goal, e.g. task oriented dialogues for services and non-task oriented dialogues for chat and entertainment. Authors express the concern that traditional chatterbots prolong the conversation with humans using pointless chat.

Therefore, GO-CA attempts to get the conversation back on track using its rule base and knowledge obtained during the conversation. The authors presented two different chatterbots with different goals. The first one called Adam served as university student debt advisor and was focused on providing advice to students. The main requirement of the system was to cope with upset students and deal with abusive language. During the usability evaluation with 200 undergraduate students the subjects were asked to provide their opinion about Adam. 75.0% of students were satisfied by the advice and 47.0% stated that they would use this system instead of going to a real advisor. The majority of student comments were positive, e.g. "(...) I felt like I was seeing a real advisor".

Among the areas where Adam could be improved, subject mentioned better understanding of slang and mobile text talk. The current version of Adam has been run at the university and system continues to learn how to relate to life as a student. As Nagano et al. have noted, the latter was used as an advisor on "bullying and harassment in the workplace" (Nakano, Hoshino, Takeuchi, Hasegawa, Torii, Nakadai, Kato, & Tsujino, 2006). A complex policy was automated using a conversational system for advice to be available 24 hours a day. The chatterbot allows using natural language to discuss a person's individual cases and offer valuable

advice. During the evaluation process 94.0% of responders indicated that they found the advice without difficulty and they didn't feel like consulting the human advisor.

I found these results to be consistent with the research on chatterbots as language learning tools by Fryer et al. (Fryer, Carpenter, 2006). The researchers conducted an experiment with 211 students participating in a sessions with two popular chatterbots ALICE and Jabberwacky. According to the results, most of participants enjoyed using the chatterbots and generally felt more comfortable conversing with the chatterbots than a human language partner or teacher. Such positive foreign language communication experience could enhance students' interest in language learning and improve motivation. Seventy four percent of participants were asked to write about their experience, and they defined the communication with a system as "funny and entertaining".

These results bring some evidence that conversational systems offering both pedagogical functions and a chance of engaging, non-stressful conversation have potential to become artificial teachers. However, the authors did not provide conclusive data and there are still questions whether their results can be treated as reliable. Therefore, more conclusive results should be provided to confirm their claims.

Of course a mere chatterbot system that correctly answers the students' questions may not be enough to create the artificial representation of a teacher. Among the functionalities that should be considered are following: to provide conversational agent with a sense of humor, the ability to react to changing emotional states of students and to offer some representation of emotions from the side of conversational system in order to make it more human-like. By fulfilling this

condition this technology has a potential to offer much more than just mere engaging dialogue.

2.4. Chatterbots with occupation

From the beginning chatterbots performed different roles supplying counsel and service to others. Just like their human counterparts, conversational systems tend to be specialized in their fields and aim to bring the high standards of professional and intellectual excellence. By the term “occupation” I mean the activity to which one a certain individual devotes oneself that also requires special knowledge or training in a given field is described. However, performing a specified function does not mean that chatterbot should be strictly task-oriented. In order to develop a system that could meet users’ expectations the non-task oriented aspects of chatterbot should also be taken into consideration.

This chapter is not proposing a new type of chatterbot systems, but a different way they should be classified. After all, the ability to interact with others by asking a question, commenting and giving an answer is the foundation of human communication. According to Bogdanovych et al. (Bogdanovych, Simoff, Sierra, & Berger, 2005) the purpose of increasing amount of chatterbots in e-commerce is to provide consumers with answers to their questions from an entity that resembles human.

Chatterbots have been widely used in the human-computer interaction and in learning they can potentially serve not only as task-oriented tools, but also as companions that would build students’ self-confidence in foreign language communication. Non-task oriented functionality seems to be a great way to allow students a chance to practice their language skills after finishing the main assignment.

They would have a chance to review a newly acquired material and have an opportunity to practice it. Students could also do the self-analysis by seeing the transcript of their conversation and finding/correcting the problematic sentences. A more advanced system could also do it automatically by preparing the summaries of more frequently occurring mistakes and offer guidance how to correct them or dynamically change the teaching strategy to concentrate on filling the gaps in the knowledge of currently studied second language.

Also for the teacher it is a chance to track progress of the students and see what language they use and spot the most common mistakes. Finally, by turning text-based conversation into voice recognition input and speech generation output they could read and listen at the same time. Creating a chatterbot with task and non-task oriented functions has one more potential advantage – it could be a good language tool for beginners to deal with specific language tasks as well as casual conversation partner for higher level students. As teachers, chatterbots should appear as human as possible to become believable teachers.

So what actually makes the ideal teacher? One of the possible answers has been brought in a detailed report that sets out the ideal educator⁷. According to the simplified definitions provided by surveyed 12-13 years old pupils a good teacher should: listen to students, encourage them, like teaching his/her subject, take time to explain things, help students when they are stuck, allow them to have a chance to speak, not give up on them, care on their opinion, take time to explain thing, be friendly and supportive (“tells you how you are doing”). There are still many possible drawbacks in bringing a full-fledged chatterbot able to serve as language instructor. Coniam (Coniam, 2008) in an evaluation of language resources of a few popular

⁷ “What makes the ideal teacher”. BBC NEWS. (2000/6/16). Retrieved from http://news.bbc.co.uk/2/hi/uk_news/education/793594.stm (accessed 2012/05/30).

chatterbots (including Jabberwacky, ALICE and Jenny) for their potential in teaching English describes a few commonly occurring problems. According to this research, current chatterbots work best when they process one-clause sentences and the topic of conversation is common. Some of them can deal with misspellings but colloquial forms of English cause interaction problems, especially for younger learners accustomed to use this kind of language in everyday conversation. In general, if user inputs incorrectly spelt words, the chatterbot experiences query processing problems.

Therefore, a chatterbot that would suggest corrections to certain ungrammatical utterances is necessary for better communication with artificial language conversational agent. Teenagers often use colloquial and simplified forms of language in their everyday communication, so a successful teacher should be able to recognize such language and, if possible, suggest a correctly spelt form or more appropriate word. Second language students often use different audio-visual materials, i.e. movies, songs, games, to improve their language skills. However, they may encounter inappropriate expressions and a teaching conversational agents could try to make users aware of the problem of their appropriateness.

Teaching chatterbot should be able to respond this issue in an appropriate way. Previously, different conversational systems have been developed for second language learning, for example systems proposed by Coniam (Coniam, 2008) and Tatai et al. (Tatai, Csordas, Kiss, Szalo, Laufer, 2003). The one who has received the most attention is CSIEC (Computer Simulation in Educational Communication) is an Artificial Intelligence (AI) framework developed at Peking University (Jia, 2009). This system is worth attention because it has been developed over the years and has been successfully integrated into selected Chinese classrooms where it has been used as a language tool to assist teachers during English classes. Even though CSIEC provides learners with a chatting partner, the creators of the system admit that there

are still many unsolved problems regarding the ability of understanding and generating natural language or dealing with textual ambiguity.

2.5 Research idea

This research aims to create a system that would perform both task and non-task oriented functions to give students not only a chance to expand their language knowledge, but also practice with a virtual language partner who could support their endeavors.

2.5.1 Baseline chatterbots

To create the baseline of the project, a casual conversation system using modality and word association retrieved from the Web(Modalín), developed by Higuchi et al. (Higuchi, Rzepka, Araki, 2008), has been adapted. Modalín uses those words association to create propositions and automatically extracts sets of words related to conversation topic selected by user. It also adds the modality to generated utterance and verifies the semantic reliability. It extracts word associations in real time using Goo search engine snippets and does not rely on any off-line database. Next, it applies the extracted word associations into proposition templates and checks its naturalness on the Internet. If the proposition has low result frequency, it is considered unnatural and system simply omits it to generate a new one in the similar way.

For the purpose of this research, two different baseline chatterbots are considered. As a consequence of occurring problems with Modalín system, highly dependent on search engines to answer the user's utterance, it was decided to also incorporate the other available chatterbot in this research. For each turn of the

dialogue multiple searches have to be made and it leads to overuse of search engine. Therefore, it was decided to use the second conversational agent that, unlike Modalin, uses limited amount of queries. A conversational system called Maru-chan created by Takahashi (Takahashi, 2009) extracts keywords from the user's utterance (nouns and adjectives) and uses them to perform fewer queries in the Google search engine. Maru-chan does not only extract word associations from the Internet, but also n-grams, the strings of words containing sets of 3- and 4-grams for every word from the list. Next, these extracted n-grams and candidates are used in order to generate chatterbot response candidates. This system also checks the length of the candidate and automatically deletes the longest and the shortest candidate to deal with sentences that convey too little or too much information. The priority goes to the medium candidates that are scored according to manually set threshold (decided after a preliminary experiment) and the top one are selected as the agent's response.

Meanwhile, a new version of Modalin that overcomes the occurring problems is being made. The aim of the new implementation is to possibly integrate functionalities of both chatterbots to utilize their best features and enrich chatterbot system's conversational strategies.

2.5.2 System components

The architecture of presented system consists of a few independent modules that cooperate with each other to perform the main task (teaching English) as well as the other, supportive functions, e.g. spelling correction, language normalization and responding to users' emotional states in an appropriate way (Fig. 1). The system is built upon the baseline chatterbot (Modalin) with other modules performing a supportive role during the input utterance analysis and generation of output utterance. The functions of each module are briefly described above.

The next step of this research is to bring an implementation of a context learning method based on the phenomenon of code-switching into chatterbot in order to generate code-mixed phrases. This is an innovative way to use this occurrence with conversational agents and to our best knowledge it has not been yet done in the field of computer science. Code-switching is the transition between linguistic units (words or phrases) from one language into another, within one sentence, where original grammar of the native language is usually preserved unchanged (Sridhar, Kamal, 1980). This research aims to create the environment where students will be able to learn new language units by context in a natural way, just like children who learn language by connecting meaning of words with the people, things, and activities around them. Namely, during conversations with a chatterbot using the base language (Japanese) students will encounter words in second language.

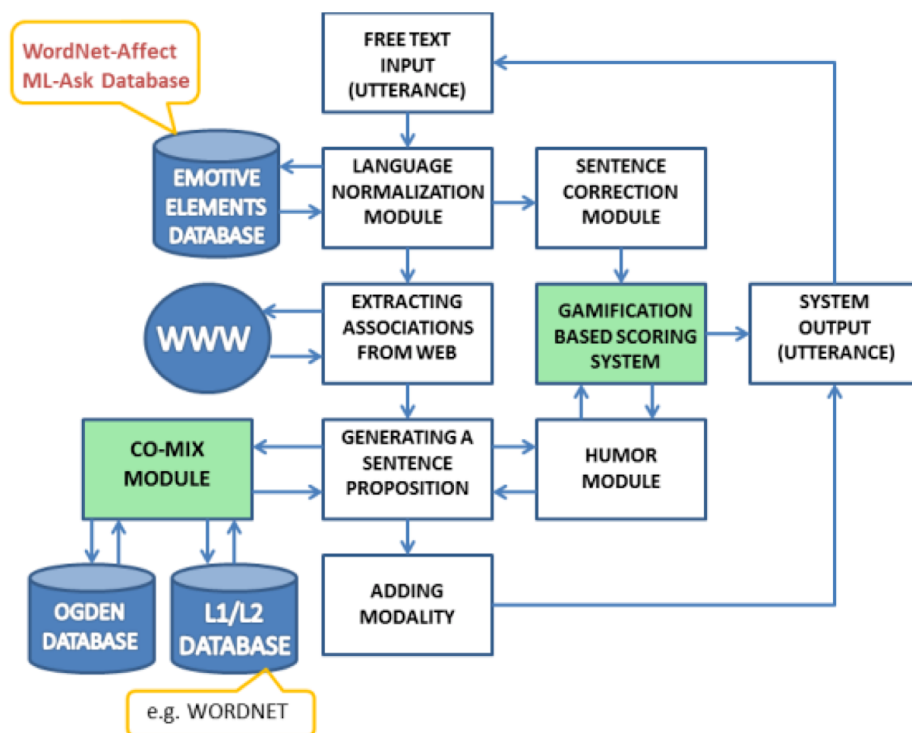


Figure 1. The overview of code-switching-based system with modules

By establishing a connection between the meaning of a certain difficult word in the first (L1) and second (L2) language, students may understand the meaning without other references, e.g. using traditional dictionaries. For example: “Let’s have a lunch together in the *shokudou*⁸ (cafeteria)” (English-Japanese code-mixed phrase).

The results of preliminary experiments described by Mazur et al. (Mazur, Rzepka, Araki, 2010) indicate that code-mixed sentences may be an effective way of expanding one’s L2 vocabulary. The recent experiment concluded with high percentage of correct answers in a given language understanding test (82.5% and 79.1% for each test phase) also brought an insight that code-switching method may be an effective way of presenting new vocabulary units to the students (Mazur, Rzepka, Araki, 2012). Code-switching module will let users engage in both task and non-task oriented conversation. In a free conversation mode this system will automatically generate Japanese sentences with English words and allow users to elaborate on their meaning while conversing with a system.

The non-task oriented functionality will allow users to select vocabulary from a given list (currently nouns) and system will use them to generate sentences using Conceptnet⁹, a free common sense knowledge base. Users will encounter both real commonsense sentences and fake sentences that do not convey the meaning of a studied word. In a sentence generation I also would like to use Wordnet¹⁰ to check the semantic distance between response candidates and eliminate the ones which are too close and may interfere with the context learning approach.

⁸ From here onward romanized transliterations of Japanese are italicised. English translations in parentheses accompany the transliterations when needed.

⁹ ConceptNet: <http://conceptnet5.media.mit.edu>

¹⁰ WordNet: <https://wordnet.princeton.edu>

Next, it is planned make use of an emotion recognition module that will recognize emotive sentences with ML-Ask emotive elements/emotive expressions analysis system for Japanese, developed by Ptaszynski et al. (Ptaszynski, Maciejewski, Dybala, Rzepka, & Araki, 2009). This system determines utterances' emotiveness and detects types of emotions. The two main features are determining general emotiveness of a sentence (emotive/non emotive) and specifying the types of emotions.

For example, the sentence “*Kyo wan nante kimochi ii hi nanda!* (Today is such a nice day!)!” is recognized as emotive, and contains emotive expression (“*kimochi ii* (nice)”) as well as emotive elements (“*nanda* (emphasis)”) and exclamation mark that make the utterance more emotive. In the second step, system checks the emotive values of the utterance and conduct the analysis of specific types of emotions. For example, the above example sentence contains an emotive expression “*kimochi ii* (nice)” which belongs to the group called “*yorokobi* (joy)”. This group is considered as positive, in opposition to negative emotion group, such as “*iya* (dislike)”. Emotion recognition system will help us deal with shifting attitude of users towards the system, as well as the problem of keeping user engaged and eager to continue the discussion with conversational system.

The other aspect taken into consideration is humor that can be used to enhance positive and reduce negative engagement. I intend on using pun generation system made by Dybala et al. (Dybala, Ptaszynski, Rzepka, & Araki, 2009). This system extracts a base word from a user utterance (mostly nouns) and transforms it using Japanese pun phonetic generation patterns to create candidate list. After checking the candidates in the Goo search engine and selecting the one with the highest hit rate it looks for a sentence with a chosen word and extracts its part to form a response or randomly selects a pun from its pun database. Dybala et al. (Dybala, Ptaszynski,

Rzepka, & Araki, 2009) also conducted a study on the role of engagement in non-task oriented conversation and the influence of humor as a measure to improve the engagement. The authors underline the fact that there is no comparative study concerning the role of engagement in conversation with task and non-task oriented systems. Finally, I would like to aid non-native speakers' reading comprehension of informal English with language normalization module based on a CEGS, a system for generating casual English short sentences from regular English input using a phonetic rule-based approach, developed by Clark et al. (Clark, & Araki, 2011).

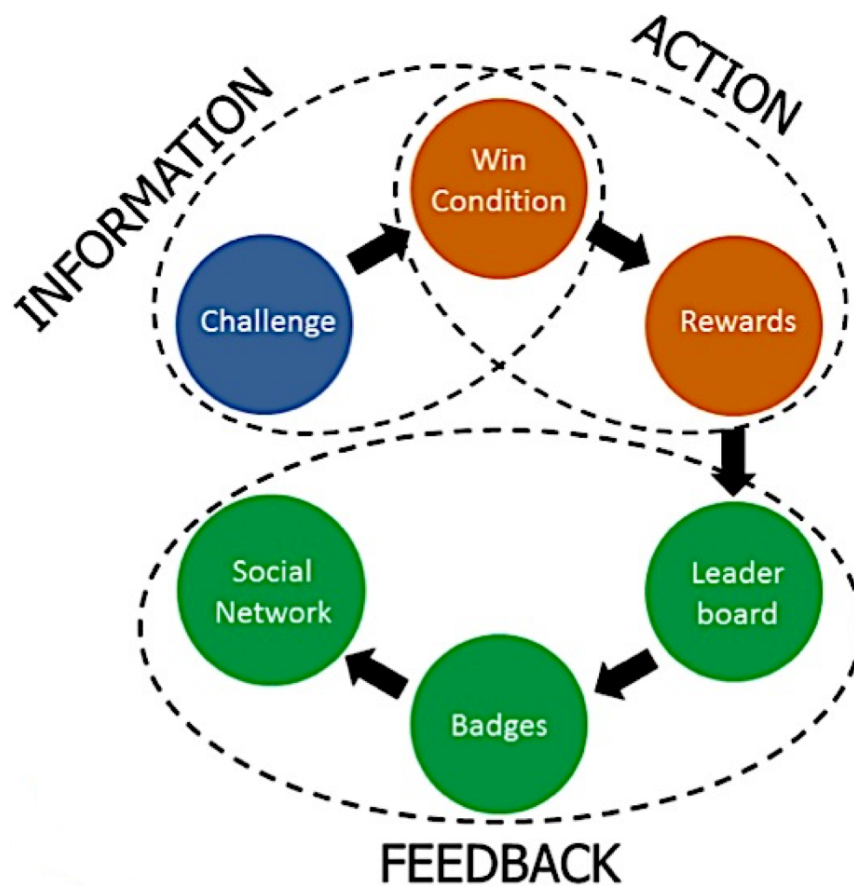


Figure 2. Gamification-based point system to encourage users' interaction

The other way I would like to deal with user willingness to enforce our system with a gamification module, presented in a work by Mazur et al. (Mazur, Rzepka,

Araki, 2011), dedicated to encourage user engagement and increase willingness to interact with a chatterbot. *Gamification* is an informal term for the use of elements of video game design in non-gaming systems. It is an innovative way of designing modern applications highly inspired by electronic entertainment. Some of existing works on dialogue systems, especially the one by Jia et al. (Jia, 2009) show that the quality of conversation with chatterbot can be measured by its duration and general willingness of the user to interact with a system. According to this research, without a specific goal of the conversation user engagement slowly decreases.

Since our initial target group consists of high school students, I decided to provide a scoring system inspired by gamification ideas, with a sole purpose to motivate users to learn foreign language and reward their efforts. Most of the schools use grades to motivate students and control their progress – high grades for excellent results, and low grades as a consequence of the lack of studying. However, an artificial teaching system should offer more than typical grading system. Using gamification ideas may bring the motivation and positive attitude of users towards the system. Therefore, such mechanism may encourage chatting with an agent and serve as a mean to keep user engaged and eager to study at longer intervals of time.

2.6 Conclusions

A successful chatterbot must offer more than just performing its designated tasks, because potential users are expecting more human-like interlocutors that will not only answer their questions but also allow them to have a free conversation. Therefore, designing a new chatterbot should take into consideration both aspects of conversational systems and create a bridge between them in order to fulfill the expectations of potential users.

Currently used scoring systems, like the one presented by Jia et al. (Jia, 2009), seem to be limited with their progress assessment methods and founded on simple grading rules, thereby allowing users to review their performance and scores only in a very basic way. Nevertheless, this work proves that such function is important to self-learning and evaluation. Therefore, I would like to introduce a more recent approach to this matter with a gamified point system that not only assess student progress, but also rewards them for their achievements and encourage doing more effort in studying. This point system is an iteration starting with a presenting a challenge for students (Fig. 2). Firstly, I explain rules of the game: present the challenge (e.g. acquiring and actively using newly learnt vocabulary units), specify winning conditions and rewards for reaching certain goals. Among the possible gratifications of user efforts the elements such as leaderboard and badges - commonly used game mechanics - will provide motivating competitiveness to users. Leaderboards show the top rank students and their best scores. Just like in video games, where the well-known urge to beat the best score and place one's own name among the top rank users has a positive motivation quality. Badges are commonly used in games and also effectively serve its purpose many popular social network applications, such as Foursquare¹¹.

Finally, the popularity of social media networks will give the opportunity to reward users with recognition of their achievement in the web community among their friends and relatives. A reward, such as dedicated badge, can be obtained automatically, after completing the necessary set of tasks.

This chapter presented the insight on the roles the conversational systems play to benefit the modern society. I underlined the fact that existing informal classification of chatterbots is not sufficient enough to express the growing expectations of potential

¹¹ Foursquare: <https://foursquare.com>

consumers. In a discussion about chatterbot occupation I emphasized the possible adaptation of conversational agents in learning. As an example, I described my work on an English tutoring system that presents a new approach to the subject by taking into consideration both aspects of task and non-task oriented conversation to face the student expectations towards a successful teaching application. I presented some examples showing that such system should perform both task and non-task oriented functions to become a useful language-learning tool and properly perform its designated role. I also discussed some existing problems in currently used chatterbots and presented some possible solutions to solve these issues.

The originality of presented approach comes from using different modules performing separate functions, e.g. a module that use code-switching-based context method to generate sentences used in conversation with a chatterbot; gamification-based scoring system to encourage user engagement; a humor module that generates simple language jokes to deal with changing user attitude towards system and motivate; a language normalization module to improve comprehension of informal English; finally a module for affect analysis to generate and analyze emotive sentences.

Chapter 3

Second Language Vocabulary Acquisition

3.1 Introduction

In the modern world, English has become a global language spoken in over sixty countries. English language proficiency is essential for cross-cultural communication and the necessity to know the language is constantly growing. This necessity leads to the problem of finding qualified teachers that can give students a chance to improve their language skills. Vocabulary is the most essential part of second language learning (Ling, 2010), and recently has become a topic of discussion for language researchers (Nation, 2001), school curriculum designers (Zhang & Anual, 2008) and others.

The problem of vocabulary acquisition is one of the most significant issues that has to be faced in order to master foreign languages. In a spoken language, about 1,800 words constitute about 80% of the spoken corpus (McCarthy, 2004). Because some frequent words are often repeated, it is thought that learners can understand a large proportion of foreign language conversation with a relatively small vocabulary (McCarten, 2007). The importance of vocabulary is explained by Wilkins (Wilkins, 1972), who states: “without grammar very little can be conveyed, without vocabulary nothing can be conveyed”. However, previous research also suggests that although second language learners need to know a large number of words as a long-term goal, this is not an essential short-term goal (Nation, 2001). The number of words known by native speakers and the number of words actually needed to use the language should both be considered.

A variety of approaches to vocabulary learning have been presented, but none have been proved to be the most effective way of studying language units. The way that new words are presented to students usually depends on teachers' individual beliefs and teaching strategies. New possibilities for vocabulary acquisition can be found in the code-switching phenomenon, in which a word from one language is used in a sentence in which the grammatical structure belongs to another language. Code-switching presents a chance for students to think about new foreign language words in a deeper manner and offers the potential for expansion of the second language vocabulary. Recently this technique has become a target of interest for researchers in different domains, such as psychologists (Macaro, 2001), linguists (Coady & Huckin, 1997) and computer scientists (Labutov & Lipson, 2014).

3.2 High frequency words

The presented research is aimed to create an environment in which students can build self-confidence in English communication. One of the ways to achieve this goal is to measure students' second language ability and adjust the difficulty of a given task to their personal level. Proposed system offers the opportunity for incidental learning using graded readers. These are texts that are graded according to high frequency word count and therefore can be used by learners at different levels of second language understanding. These words are very important because they account for a large proportion of the content in spoken and written texts and are linguistically adjusted to a given user second language level. For example, in the Oxford Bookworm series Level 1 covers 400 words and Level 6 encompasses a total of 2,500 cumulative words. Knowledge of about 4,000 words usually covers 86.0% of a given text (Nation, 2001).

Graded readers are used to support the extensive reading approach to second language learning and provide comprehensible learning materials with word difficulty levels corresponding to users' language ability. The high frequency words of the

language are important, and therefore both teachers and learners should spend considerable time on their acquisition. In this research, vocabulary acquisition is facilitated by repetition of target high frequency words through a series of tests (quizzes) and presenting comprehensible examples of their usage. If there is a delay between the presentation of a word form and its meaning, students have an opportunity to make an effort to guess or recall its meaning. Nation (Nation, 2001) states that guessing can only be successful if a good clue to the meaning (incidental learning) is provided. These quizzes offer the chance to experience both incidental vocabulary learning and direct teaching - two popular ways of learning high frequency words.

The problem of automatic generation of teaching materials has also been approached by Ginsburg (Ginsburg, 2012) who proposed a learning program to generate editable English lesson materials to speakers of Japanese. The author underlines that higher vocabulary skills are crucial in order to understand the words while using this application. This research also provides reading exercises with highlighted lexical units and their dictionary definitions. The author underlines that higher vocabulary skills are crucial in order to understand the words while using this application. He also provides reading exercises with highlighted lexical units and their dictionary definitions.

3.3 Code-switching

Code-switching (other names that commonly refer to this phenomenon are code mixing and code changing) is a phenomenon that exists in many multilingual societies where people use more than one language to communicate on a daily basis, such as Singapore, the Philippines, India, the USA, Spain and China (Velasquez, 1995). Code-switching has also become a target of interest of Japanese scientists. Namba (2012) used data collected from English and Japanese speaking siblings identified the structural processing which underlies code-switching and stated that in

bilingual interaction switching between two languages is a common phenomenon. Winch (2015) described the phenomenon of code-switching-based Lou Language (*Ruu-go*) that has become popular among Japanese teenagers between 2006 and 2007. The researcher summarizes the success of this language phenomenon by two factors, which is a bilingual experience among the Japanese and the influence of globalization.

Code-switching can be defined as “the alternation of two languages within a single discourse, sentence or constituent” (Jamshidi & Navehebrahim, 2013) or explained as a brief insertion of a word from one language into the other by two people engaged in a conversation. Code-switching can occur on the lexical level, when there is no existing translation for a word or sentence to L1, or on the semantic level, when an idea can be better explained in second language. There are also two possible locations where the switch can occur, i.e. intersentential (after the sentence) and intrasentential (within the sentence) (Zirker, 2007).

Examples of vocabulary code-switching in English and Japanese sentences are given below:

- “*Watashi ha **apple** wo tabetai desu.*” (‘I’d like to eat an APPLE.’) - English word within a Japanese sentence
- “*She bought it as **omiyage**.*” (‘She bought it as a SOUVENIR.’) - Japanese word within an English sentence

Code-switching often occurs as a stage of bilingual children development, when they naturally mix elements of different languages and move from one system to another without noticeable discrimination (King, 2006). This phenomenon reflects the transition between linguistic units (words) of one language into another, within one sentence, when original grammar of native language is usually preserved unchanged. Code-switching occurs when a speaker takes some components of one language to use it while speaking another. Sridhar explains it as a kind of transition of

units of one language (L1; mother language) to using those of another (L2; foreign language) within a single sentence (Sridhar et al., 1980). For example:

“Let’s have a lunch together in the *shokudou* (cafeteria)”
(English-Japanese code-mixed phrase; a real life example).

The point of interest of this research centers on using a code-switching effect to introduce new linguistic items. The units such as words, phrases or clauses can be interchanged and by understanding the context of the whole phrase a student would be able to understand a foreign word without knowing the direct translation. By inserting a single L2 words into the L1 sentence it is possible to guess the meaning by L2 - L1 word association. Some of existing works on teaching vocabulary through code-switching (Celik, 2004) show that this phenomenon can be effectively used during EFL (English as a Foreign Language) classes. Code-switching brings a chance for students to think deeply about new linguistic units and may bring the chance of L2 vocabulary expansion.

The majority of people using code-switching are bilinguals who are able to speak multiple languages. They can easily code-switch and use the method as a resource to find better ways to convey different meanings. Velasquez (1995) states that for bilinguals code-switching is a linguistic option, because they are proficient in both languages. However, there is an increasing amount of evidence that code-switching can also be used by less proficient people to fill linguistic gaps whenever the learner of a second language encounters the problem of insufficient vocabulary (Nishimura, 1995; Oxford & Crookall, 1990). These two aspects create a common distinction between code-switching as an asset for bilingual people with a high competence in both languages, and as a “reparation tool for insufficiency in the second language” (Hamers & Blanc, 2000). A similar notion comes from Song et al.

(Song & Andrews, 2009) who describe code-switching as “an attempt to keep the conversation flowing without having to pause or abandon the message”.

In the last two decades of research, there seems to be an agreement on the positive effects of code-switching in learning and that it could be applied in second language learning to facilitate the learning outcome (Celik, 2003; Jamshidi & Navehebrahim, 2013; Song & Andrews, 2009; Lin, 2013). There is a common agreement that when code-switching is planned ahead of use, it can greatly contribute to more efficient understanding of a given topic, and be used as a part of second language learning. These effects of code-switching have been studied by Cook (2001) who also found that the use of code-switching in L2 studying does not have a negative effect on learning a foreign language, and moreover, that vocabulary learning can be facilitated by mixing two languages.

Another interesting finding is demonstrated in the work presented by Lin (2013), who discovered that code-switching seems to increase the amount of cognitive processing by students. According to Lin, a larger cognitive effort is necessary to process new vocabulary, which means that students will learn new vocabulary more thoroughly. Although Lin’s research did not show conclusive proof for higher effectiveness of code-switching for learning new vocabulary, it indicated that code-switching does not affect the vocabulary acquisition in a negative way.

In this research I focus on intrasentential code-switching, where switching takes place inside a sentence. An idea introduced by Celik (2003), who presented the assumption that in the human mind second language vocabulary is organized in the same way as synonymic expressions are organized in the first language (L1), is also incorporated. Celik’s research discusses that a link potentially exists between the first and the second language by means of associations and semantics. This link would allow interlocutors to consciously switch elements between two languages. Celik also

presents some benefits of code-switching technique for more successful learning of new vocabulary in a storytelling context.

Some attempts have been made to implement bilingual teaching based on code-switching in an online e-learning system, like the idea of an e-learning system for code-switching bilingual teaching by Shi et al. (Shi, Liu & Chen, 2013). However, this Google Translate-based method and the limited focus on improving students' international employment competence is significantly different from the approach presented in this thesis. The idea presented by Shi et al. also lacks experimental data and does not seem to contribute significantly to the field of e-learning systems and code-switching based methods. Abdali (2010) notes that code-switching does not negatively affect the acquisition of new vocabulary except for minor spelling problems, and that even a simple text can be sufficient to present the new vocabulary to students. The author mentions that some problems with proper spelling of acquired words may occur, but the possibility of wider application of code-switching as a teaching device was suggested.

3.4 Incidental learning and extensive reading

A study by Nation (2001) shows that a considerable amount of second language word acquisition happens through incidental learning. The author also stresses the fact that the long-term vocabulary growth is greater via incidental learning. The outcome of this research is the incidental learning theory, inspired by the way children learn vocabulary in their first language. The theory states that the vast majority of vocabulary is learned gradually through multiple exposures to a word over time. However, a typical student who tries to learn a second language vocabulary will encounter the widespread tradition of studying through rote memorization and making word lists. This vocabulary learning strategy comes from the previous century and offers only a minimal amount of context, and therefore is widely criticized and not generally recommended by teachers. For example, an article by Oxford et al. (Oxford

& Crookall, 1990) expresses concern that students might not be able to use new word pairs in any communicative way without further assistance.

Another idea of how to improve vocabulary memorization that contributed to this research originated in a study by Nagy et al. (Nagy & Herman, 1985). This study claims that extensive reading should be promoted, because it can lead to greater vocabulary growth. Further research by Nagy et al. (Nagy, Anderson & Herman, 1987) also showed that even one or two exposures to a target word are able to contribute to a students' initial vocabulary knowledge. Krashen believes that language learners can learn new vocabulary and spelling more efficiently by receiving comprehensible input while they read, and postulated this concept in the Input Hypothesis (Krashen, 1985).

This notion was also confirmed by Day et al. (Day, Omura & Hiramatsu, 1991), who carried out a similar study with Japanese high school and undergraduate university students. The experiment subjects read an adapted version of an English story and then took a vocabulary test on some of the words from the story. The subjects who read the story learned significantly more words than the control group who only took the test. This outcome provides some insights into the way foreign language vocabulary is studied and proves that extensive reading has a highly contributive role in the process. Horst et al. (Horst, Cobb & Meara, 1998), in a study using a graded reader with two kinds of vocabulary test, found that an average of five new words were learned. This high acquisition came from the effect of the length of the text and the use of a simplified reader where the unknown words do not appear too densely.

According to the study by Labutov et al. (Labutov & Lipson, 2014), humans tend to be good at inferring meaning from context and the authors also state that single words can be understood from the context of the surrounding text in the paragraph. Further proof for the potential effectiveness of code-switching for

language acquisition comes from the recent findings of Borovsky et al. (Borovsky, Elman & Kutas, 2012). This research contributes evidence that even a single exposure to a novel word in a constrained context results in the integration of the word to the reader's personal semantic base. It is contrary to a prior belief that a given word must be repeated with a sufficiently high frequency, and shows that words can be remembered incidentally while reading. Coady et al. (Coady & Huckin, 1997) have stated that many adult L2 learners are acquiring new words through reading to improve language proficiency. It has been emphasized that repeated exposure to new vocabulary is needed for a progressive learning process, however previous studies in both first and second language acquisition by Horst et al. (Horst, Cobb & Meara, 1998) did not specify how much exposure is necessary for effective vocabulary acquisition.

3.5 Teaching vocabulary

There are many existing methods of teaching second language vocabulary. One of these is the “diglot method” presented by Christensen et al. (Christensen, Merrill & Yanchar, 2007). In this approach foreign language units are introduced gradually by using a first language storybook. This method requires a human translator who manually prepares the book. The reader is exposed to different words and phrases expressed in the foreign language, but the shortcoming of this method is that student has to look up new words in a dictionary to understand the contents of the book.

Teaching vocabulary mostly involves making students remember new language units by presenting words many times before students can successfully memorize them. Repetition is one of the popular techniques, along with actively recalling a word to be learned. It is also important to organize words in meaningful ways to make it easier to learn, as is often presented in textbooks, as thematic sets benefit the memory.

The other popular way to teach vocabulary is personalization, e.g., using the vocabulary meaningfully to form statements about the student's everyday life, as presented in the work by Schmitt (2000). It is common to encourage students to make their own vocabulary notebooks which consist mostly of written translations of single words in a list, but this method seems to be very limited without actual sentences and dialogs that represent the meaning of words (McCarten, 2007).

Coady et al. (Coady & Huckling, 1997) claim that acquiring a substantial vocabulary in the second language is necessary to attain competencies in other language skills, such as listening, speaking or writing. This research follows Krashen's belief that "language learners acquire vocabulary and spelling more efficiently by receiving comprehensive input while reading" (Krashen, 1985). This general idea is followed with the graded readers approach. This assumption is in strong opposition to the conventional rote memorization technique of lexical studying. Therefore, it can be asserted that new vocabulary units should not be presented only with their first language equivalents and without the presence of verbal second language context.

According to Gibbons's research (Gibbons, 1987), nouns are the most often code-switched word class. As many single lexical item switches are nouns, the presented research focused on utilizing this findings in order to make the best use of the code-switching phenomenon. Further evidence of the importance of nouns is presented in the studies by Lanza (1997) and Vihman (1998), who also proposed that nouns are more commonly used for the embedded language, i.e. a given base language with the inserted elements of the second language, because translations to the second language are often provided for nouns. A research paper by Yen (2009) also contributes evidence that English often acts as the embedded language with the base Mandarin and Cantonese languages in Hong Kong, and that nouns are the most frequently mixed part of speech.

Chapter 4

Emotions and Learning

4.1 Introduction

In recent years English has become a dominant language in over 60 countries and as an international language it is represented in every continent. Its role as a tool for cross-cultural communication cannot be neglected. With a growing necessity to study English comes problem of finding qualified teachers that would give student a chance to practice language skills and improve. However, it is not only difficult but also expensive to get such assistance. Therefore, creating an application that would build students' self-confidence in English communication is one of main goals of the research presented in this thesis.

This chapter presents the introductory work that created the foundation of the research presented in the chapter that follows. In this chapter we present the first step towards this goal - the research on a system that would give students an opportunity to expand their knowledge of second language vocabulary and to practice second language vocabulary is presented. The main purpose of this introductory research is to check how emotive language with example sentences that appeal to emotions will affect the learning process of test subjects and their performance while studying with code-switching method.

Two issues are considered: the first one is to observe how sentences with emotions will help dealing with shifting attitude of users towards the system; the

second one is to check whether humor that can be used to enhance positive and reduce negative engagement of users.

4.2. General information

4.2.1 Background

Human languages can be divided into two parts: one is its grammar, and the other is vocabulary. The dictionary of words is constantly expanding and growing larger every year. The problem of vocabulary acquisition is probably the biggest drawback that has to be faced in order to master the foreign language. In spoken language about 1800 words make up about 80% of the spoken corpus (McCarthy, 2004). Because some frequent words are often repeated it is said that learners can understand a large quantity of foreign language conversation with a relatively small vocabulary (McCarten, 2007). According to Wilkins (Wilkins, 1972) “without grammar very little can be conveyed, without vocabulary nothing can be conveyed”.

During English conversational classes teaching vocabulary is a significant part of course. There have been many existing method of teaching second language vocabulary, e.g. the “diglot method” invented by Burling (Christensen et al., 2007). In this approach foreign language ingredients are introduced gradually by using a native language storybook. This method requires a human translator that manually prepares the book and reader is exposed to more and more words and phrases expressed in foreign language. The shortcoming of this method is that student has to look up for new words in a dictionary to understand the contents of a book.

Teaching vocabulary is mostly about making students remember new language units by presenting words many times before students can successfully memorize them. Repetition is one of the popular techniques, along with actively recalling a word

to be learned. It is also important to organize words in meaningful ways to make it easier to learn just like it is often presented in textbooks, as thematic sets to benefit the memory. The other popular way to teach vocabulary is personalization, e.g. using the vocabulary meaningfully to form statements about student's everyday life, as presented in the work by Schmitt (Schmitt, 2000). It is common to encourage students to make their own vocabulary notebooks which consist mostly of written translations of single words in list, but this method seems to be very limited without true sentences and dialogs that represent the meaning of words (McCarten, 2007).

This research is following Celik's assumption (Celik, 2004) that learners' second language vocabulary and other linguistic components are organized in the same way the synonymic expressions are organized in the first language. Celik's research brings some evidence that there might be a link between two languages by associations and semantics. In other words, it can be asserted that it is possible to consciously switch elements between two different languages.

4.2.2 The role of emotions in learning

In this research, different existing classifications of emotions have been considered, and Ekman's (Ekman, 1999) list of six basic emotions have been selected. It contains such emotions, like: happiness, surprise, sadness, disgust, anger and fear. This classification has been used to create sentences in the preliminary experiment described later. Craig et al. (Craig et al., 2004) provides evidence for a link between learning and some affective states like confusion and boredom and points to an existing connection between emotions and learning. It has been proved that emotions modify thinking and problem solving. Some research suggested that emotional stimuli tend to be better remembered than unemotional stimuli (Bower, 1992).

There are also psychological evidences for the existence of emotional selectivity in human memory. For example, items associated with emotional reactions tend to be learned better, depending how strong was the reaction. There seems to be a memorizing advantage for affective material (Parrot et al., 2000).

4.3 The idea of an English tutoring system

4.3.1 Aim

The aim of the project is to combine a few independent modules that cooperate with each other to perform its main teaching function, but also to deal with related problems, such as encouraging users engagement, aiding in reading comprehension of informal English and enhance positive engagement with humor. Therefore, an implementation of an emotion recognition module that will recognize emotive sentences with ML-Ask, a system for affect analysis, developed by Ptaszynski et al. (Ptaszynski et al., 2009) is considered.

The main assumption states that emotion recognition system will help to deal with shifting attitude of users towards the system, as well as with the problem of keeping user engaged and eager to continue the discussion with the conversational system. The other aspect taken into consideration is humor that can be used to enhance positive and reduce negative engagement. This project is intended to utilize pun generation system presented by Dybala et al. (Dybala et al., 2009). Finally, aiding non-native speakers' reading comprehension of informal English with language normalization module by Clark et al. (Clark et al., 2011) is considered.

4.3.2 Code-switching method

This research continues to explore code-switching method, as an effective way of acquiring second language vocabulary. Previous experiments brought some evidence that code-switching method can be an effective way of acquiring second vocabulary. In the previous experiments presented by Mazur et al. (Mazur et al., 2010) most of the participant successfully remembered and recreated test sentences in mixed language and were able to successfully understand the meaning of new words in the second language from the context of given sentences. This general idea is further analyzed in the preliminary experiment presented in this chapter of the dissertation. The preliminary experiment (presented below) was conveyed to examine how emotive language with sentences that appeal to emotions will affect the learning process of test subjects and their performance.

4.4 Preliminary experiment

The experiment was conducted with 15 Japanese students (age 10-13) from local elementary and junior high schools in Sapporo, Japan. Test subjects were selected as a target age group that represented our future system users. The subjects were mostly males (73.3%). In this section, I will describe the purpose of the experiment, method, present the results and discuss our findings.

4.4.1 Purpose

The purpose of the experiment was to teach students new English vocabulary using presented code-switching method. The idea was to test if the subjects are able to understand the meaning of a whole sentence by associating English word with its

Japanese counterpart. Test subjects had to guess the meaning of new words by context of the whole sentence, without using dictionary. They were not told the purpose of the experiment, just simply asked to follow teacher's instructions. For purpose of this experiment the research interest focused on replacing nouns. Gibbons's research (Gibbons, 1987) confirmed that nouns make up the most often code-switched word class, so in the first step it was decided to follow this general trend.

4.4.2 Method

The experiment consisted of two parts. First, a learning phase occurred, followed by a testing phase. The contents of each part of the experiment are described below:

1) Learning phase

Four pattern sentences in Japanese with English word inside were presented to test subjects. Each experiment introduced two English words. In the first round (sentences without emotions), the selected words were "culture" and "sightseeing". In the second round (sentences with emotions) two more, "relationship" and "announcement" were used to form sentences. Vocabulary lists from Sunshine 3 textbook (Shimaoka et al., 2011) have been used to choose vocabulary for the experiment. Sunshine is the official English course book used at Junior High School in Japan, and it was selected to assure that test subjects were not familiar with this words before the test. Sentences were presented on tablet (iPad 2nd generation) and subjects were asked to read them from the screen.

The learning phase took approximately 1 minute for each subject (about 20 seconds for each question). The knowledge of selected words was double-checked by

asking our subjects after the test whether they knew presented vocabulary. 87% of subjects never heard any of presented words before the experiment, and 13% mentioned they may have heard it before, but they were not sure of the meaning.

The participants were asked to memorize Japanese sentences with English nouns replacing Japanese counterparts. All the pattern sentences in the learning phase were used in a correct context that corresponded to the presented vocabulary.

2) Testing phase

Two separate tests were performed to check the effect of emotions on problem solving: one with plain sentences without emotional load and the other one with sentences that convey emotions to see to what extent they can modify participants thinking and influence the test results. The order of questions was randomized. Each set consisted of 8 sentences. Half of the sentences were semantically correct and the rest were incorrect, i.e., they used the English noun in the wrong context.

The subjects were given two minutes to read sentences and decide if they are correct or not (“O” for yes and “X” for no). After the second turn of the experiment, the students’ understanding of presented vocabulary was checked to assure if they properly learned the new vocabulary.

Example sentences:

a) Sentences with no emotional load:

- *Kaigai ni sunde iru indojin ni totte wa, boukoku no **culture** wo ikasu no ga juuyou desu.*

(Translation: For Indians who live overseas, keeping their home culture (alive) is important.)

- *Kyoto **sightseeing** wa aki ga ichiban yoi to omoimasu.*

(Translation: As for sightseeing in Kyoto, I think that autumn is the best.)

b) Sentences that convey emotions:

- *Gakkou no tomodachi ga tenkou ni naru to iu **announcement** ga kite, ima wa kanashii desu.*

(Translation: There was an announcement that my friend will transfer schools, so I'm sad.)

- *Shiawase to okane wa amari **relationship** ga nai no de, kyuuryou no hikui hito de mo ureshiku naremasu.*

(Translation: There's not much relationship between money and happiness, so even people with low income can be happy.)

c) Wrong sentences with a semantic mistake.

- *Resutoran de **announcement** wo chuumon shitara, oishikunakatta no de kanashikatta desu.*

(Translation: I ordered announcement in the restaurant, but it was not tasty, so I was sad)

- *Kare no **sightseeing** ga warukatta no de, jugyou wo yasumimashita.*

(Example: Because his sightseeing was bad, he was absent in the class.)

4.4.3 Results

The results of the preliminary experiment in comprehending the test sentences using discussed code-switching-based context learning method are presented below. The percentage of results in both tests is compared in Table 1. Test number 1 refers to sentences with emotive expressions, and Test number 2 indicates non-emotive sentences. The maximum score the participant could receive was 120 (15 points for each correct answer).

Table 1. The percentage of correct answers for each question in both phases of the experiment

Test number	Number of correct answers	Percentage of correct answers
1	99 out of 120	82.5%
2	95 out of 120	79.1%

The percentage of correct answers for each question is presented in Table 2. Highest and lowest scores are indicated in bold.

Table 2. The results of preliminary experiment compared in both test phases and presented by question number and percentage of correct answers

Test number	Question number (1-8)							
	1	2	3	4	5	6	7	8
1	60	80	80	93.3	80	93.3	86.6	86.6
2	93.3	80	66.6	93.3	40	93.3	86.6	80

The comparative user results of target vocabulary comprehension in both tests are presented in Figure 3. Horizontal axis shows the question number (from 1 to 8) and vertical axis represents the number of participants (out of 15). Blue color represents Test 1 questions (sentences with emotive expressions) and red represents Test 2 questions (sentences with non-emotive expressions).

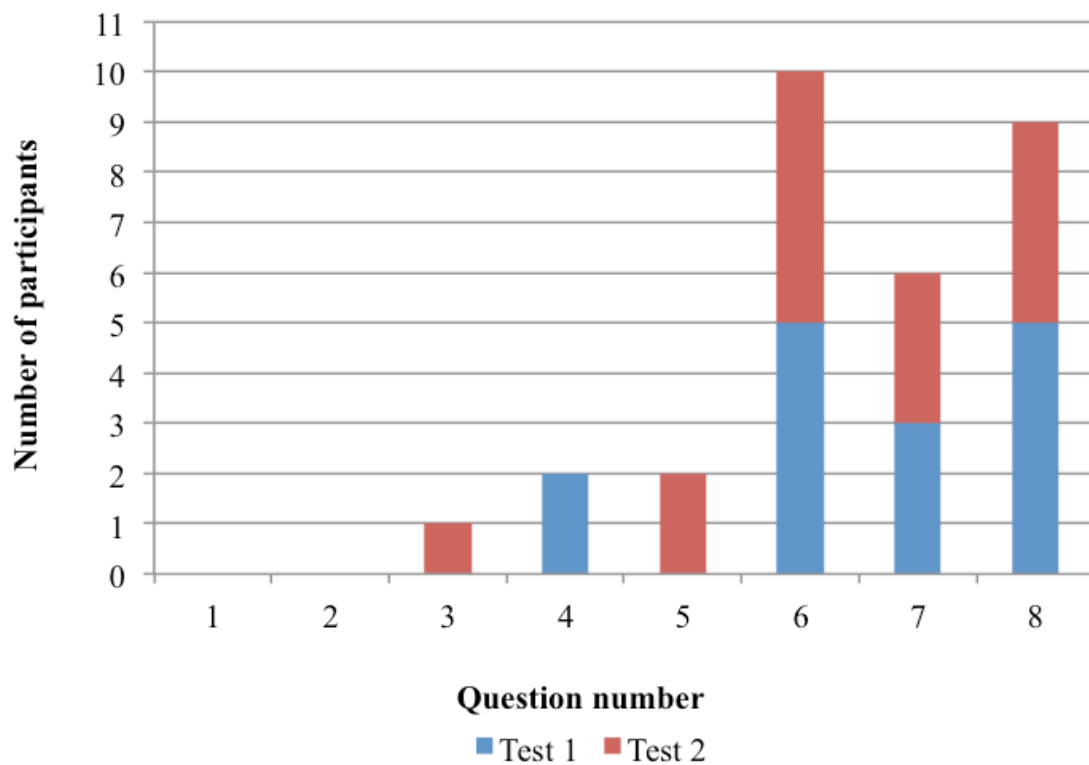


Figure 3. Comparative user results of target vocabulary comprehension in both tests

4.4.4 Discussion

After analyzing the experiment results the following things have been noticed:

a) the percentage of correct answers in both phases were relatively high which proves that code-switching method may be the effective way of presenting new vocabulary units

b) since 87% of subjects were not familiar with presented vocabulary before the test, we can assume that they successfully linked two languages by associations and semantics, thereby understood the foreign word without the direct translation

c) by getting a higher percentage of correct answers for each question in the second phase and a larger variety of scores compared to the first phase, it can be interpreted that sentences with emotions had an impact on participants' responses

Although it was expected that emotional sentences will be easier to memorize and results will be higher, there was a 3.4% difference between the number of correct answers in both phases, in favor of non-emotional sentences. An assumption has been made that it may be due the fact that sentences in second phase were significantly longer and thus perhaps more difficult to process and rationalize.

After further analyzing the participants results, it was noticed that one of the presented sentences in the second phase caused a particular comprehension difficulties, with only 40% of correct answers. As a result, this question negatively influenced the final score. It is also worth noticing that the percentage of correct answers for three questions in second phase was higher than in the first phase, as it has been shown in results.

In an attempt for further investigation of results it was decided to check how excluding of the question with the lowest score will influence the overall percentage of correct answers. To keep the balance between both tests the corresponding question from the first phase were excluded. As an outcome, it was noticed that the second phase had 1.8% higher score than the first phase (77.6% correct answers in first test and 79.4% for second test).

In spite of the gap between the scores received by the participants in two phases, the ambiguity of scores in the second phase shows that this matter should be a

subject of further investigation. This results give some insights about the influence of emotions on the process of human reasoning and the motivation to further investigate this matter.

4.5 Conclusions

In this chapter, the idea of an English tutoring system aimed to bring new approach in second language vocabulary teaching was described. With the results of the preliminary experiment in understanding the code mixed sentences, it has been shown that the proposed method might be an effective way of expanding students' second language vocabulary. The high percentage of correct answers, making successful connections between meanings of words in two languages and the outcome of results brought the motivation for further study of this matter.

In further research, another experiment on a wider scale is planned for more conclusive results. An idea of testing the proposed method on sentences with emotions expressed by emoticons has also been formed. By doing so, it will be possible to shorten the test sentences and avoid occurring problem described in results section. The results also confirmed that presented code-switching method is effective for second language vocabulary acquisition.

Therefore, in the next step of this research code-switching module will be implemented in the second language vocabulary teaching system to observe how code-switching would function in an e-learning application, what are users attitudes towards presented method and how users benefit from using such an e-learning application.

Chapter 5

The Co-MIX Project

5.1 Introduction

The purpose of the Co-MIX Project is to bring applications that utilize code-switching method and create systems able to teach second language vocabulary to students. This chapter describes the work on Co-MIX project and presents the major contribution to fulfill research goals - Co-MIX Japanese, a teaching system using code-switching for English vocabulary acquisition.

In section 5.5 of this chapter the preliminary evaluation of another system, aimed at English-speaking users who want to study Japanese vocabulary, is also presented. The purpose of the other research was to find more evidence that the code-switching-based method used in Co-MIX Japanese system can be adapted in other languages.

A few assumptions have been formed while creating Co-MIX Japanese system. They are based on a preliminary sample of 10 subjects and prior research. The assumptions include the following: 1. The use of code-switching phenomenon in computer systems will expand students second language vocabulary; 2. It is expected that the user attitudes towards the proposed system to be generally positive; 3. The average number of mistakes is expected to be higher for the baseline system; 4. It is assumed that the users will benefit from code-switching method and the scores in language tests while using the proposed system will be higher; 5. Users will be able to successfully obtain the meaning of a number of L2 words from context, without referring to a dictionary.

5.2 Co-MIX Japanese

5.2.1 System introduction

Co-MIX Japanese is an English vocabulary teaching system that uses code-switching for vocabulary acquisition. The purpose of the system is to introduce a new vocabulary learning method for e-learning systems using intrasentential code-switching, incidental learning and the benefits that are generated with learning vocabulary in context. The following is an example of a Japanese sentence with intrasentential code-switching:

“Migi ni magaru to, konbini ga arimasu. LEFT ni magaru to kouen no geeto ga miemasu”(If you turn right, there is a convenience store. If you turn LEFT, you will see a gate to the park”).

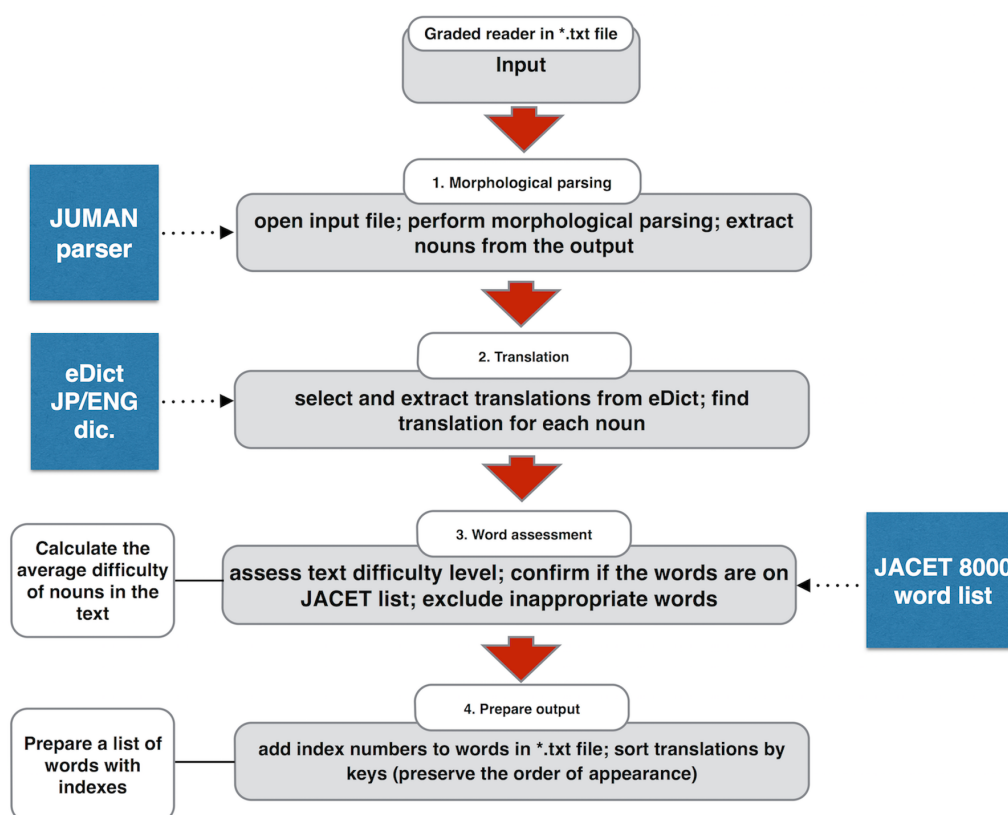


Figure 4. Process of study material generator (Co-MIX Japanese)

The modular architecture allows easy extensions of the system with additional functionalities. Co-MIX Japanese system currently has two modules. The first one is responsible for the automatic generation of study materials based on given graded readers in L1 (Figure 4). Its functions are described below. The other module of the Co-MIX Japanese system, the Language Quiz Generator, checks the users' understanding of L2 vocabulary through a series of language tests. These quizzes use different strategies to prompt students to match second language vocabulary with their counterparts in the L1 (Figure 5). The system also responds to users' mistakes by providing them with example sentences that show the correct usage of a given word.

In the three available test phases we use different language quiz strategies, e.g. spelling exercises to reinforce students' understanding, or presentation of phonetically similar words to make students' aware of pronunciation similarities between English words. Since each user possesses a different level of English vocabulary knowledge, in the first step the Co-MIX Japanese system evaluates user language competency and presents a study article appropriate to the user's English vocabulary knowledge level. Graded readers are used to introduce new vocabulary units and are followed by the vocabulary understanding test phase (Figure 4). The current version of the system is provided with a selection of readings (graded articles) in Japanese for the proposed system, and their translations into English used by the baseline system.

5.2.2 Systems used

In this research, comparative evaluation experiments were conducted with two systems. The baseline system is a variation of our proposed system without the code-switching module. It is equipped with some commonly-used vocabulary learning techniques described above, such as repetition and the diglot method; however, it lacks the advantages of the intrasentential code-switching method presented in the proposed system. Another difference comes from the fact that instead of using

readings in a mixed language with intrasentential code-switching, the baseline system presents reading exercises entirely in English using the same set of articles manually translated and proofread by a native speaker.

5.2.3 Study material generator

The main function of this application is to automatically create editable learning materials from a given text file (selected graded readers). The study material generator is fed with text files with graded readers in English and Japanese and automatically generates an editable output file with tagged vocabulary items (nouns), their translations in English for the proposed system, and their positions on the JACET 8000 vocabulary list (Uemura & Ishikawa, 2004). JACET 8000 covers approximately 85% of the words used in the graded readers and provides the average difficulty of a text, which allows the system to categorize the text to an appropriate difficulty level.

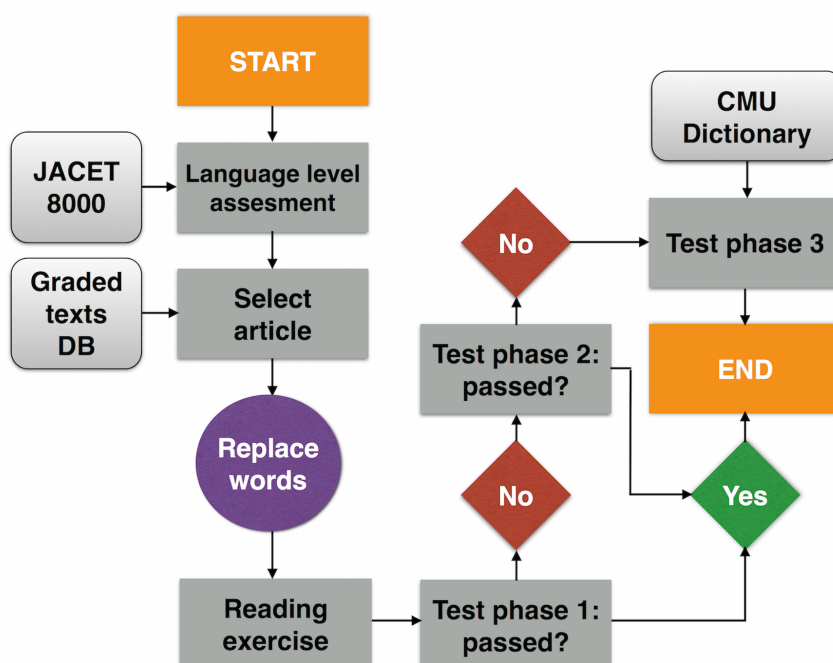


Figure 5. Flow process chart of the language quiz generator (Co-MIX Japanese)

Usually, preparing suitable readings for students presents a challenge for teachers. The study material generator can be used to manually create additional learning materials from given input texts. Selected materials must have a high percentage of vocabulary appropriate to the specific group of learners in order to help students to properly infer word meanings. JACET 8000 is a list of basic words designed specifically for English learners in Japan. It is suitable for our system's needs because words are presented in different levels of difficulty in accordance with their frequency and educational significance. The list is considered to be the most reliable and usable English high frequency word list for graded readers in Japan, mostly for high text cover rate including the vocabulary in both easy reading materials and contemporary English articles (Uemura & Ishikawa, 2004).

時間のドロボー

テレビは時間ドロボーである。寝れて家に帰ってきて、もうなんにもする(1)気がしないときなど、ついテレビをつけると、なんだかにぎやかに騒いでいたり、その日のスポーツの結果を繰り返し映したりしているから、ぼーっと見てしまう。

見ている番組が終わると今度はもの足りなくなって、チャンネル切り替えのコントローラーを探し出し、パッパッパッと、じつに便利に(2)画面を変えていく。

あんまり見たいと思う番組はなくても、ピーナツを食べ始めたのと同じで中毒に近くなっている。無いもの足りないから、(2)画面を消さないのでも。そうやって30分(3)単位で時間を潰され続けて、結局(7)局深夜の(4)映画を見終わったら、(5)明け方の4時に近かった、という経験がある。それですっかり(6)生活のリズムが狂ってしまう。

目に悪い、体に悪い、アタマに悪い、などと勝手なことを言うけれど、テレビ(7)局としては、どんな(8)手段を使っても、視聴者を(2)画面に惹きつけておくのが(9)商売であるから、それに引っかかるこちらが悪いのである。

休日、人に会って、(10)昨日のテレビの話をしてみると、みんな見ていたとみえて、じつによく話が通じる。みんな時間ドロボーにあっていたのである。その時間に眠を使って(11)読書をする、(12)会話を交わすというのではなく、みんな私と同じようにぼんやりして、光り輝く(2)画面を眺めてアホになっていたのである。

逆にテレビを見ていなければ、人と話が通じない。(13)漫画を読んでも、ギャグが解らない、クロスワードパズルも解けないということになる。静かに(14)秋の月でも眺めようと、テレビを消し、(15)電灯を消して窓を開け放つと、葉家のテレビの大音声とバカ笑いがとび込んでくる。

1. Indexed vocabulary

2. Editable vocabulary list

3. Average difficulty of article

(1)気 | spirit | 1
 (2)画面 | terminal screen | 6
 (3)単位 | unit | 2
 (4)映画 | movie | 1
 (5)明け方 | dawn | 4
 (6)生活 | living | 6
 (7)局 | bureau | 7
 (8)手段 | means | 6
 (9)商売 | trade | 1
 (10)昨日 | yesterday | 1
 (11)読書 | reading | 6
 (12)会話 | conversation | 1
 (13)漫画 | cartoon | 11
 (14)秋 | autumn | 3
 (15)電灯 | electric light | 1

3.93333333333

Figure 6. Example of the output file of study material generator (Co-MIX Japanese)

The process of the study material generator is as follows. The first program searches the input file for nouns and checks if their English translations are available in the dictionary. In the next step, the program analyzes the word rankings (difficulty level) in a given reading and an average difficulty level is provided for the whole article. Then, the system generates an output text with selected words preceded by an index number and supplemented with a list of all chosen words, their proposed translations and their difficulty ranking. In this step, human teacher is allowed to review the result and eliminate possible mistakes that may be caused by limitations of the JACET 8000 list word coverage. This is a necessary step before handing over the materials to the language quiz generator (explained in the subsection below) and presenting them to users. A few search rules had to be established.

This study focused on words that are nouns, have an available translation in the dictionary, have a translation that consists of no more than three words, are on our frequency list and are not on the list of the first 500 most common words (stop words). This list includes some frequent words that students are most likely to know. Therefore, the last condition was introduced to exclude these common words, e.g: “the”, “is”, “at”, “which”, “on”, etc.

5.2.4 Language quiz generator

The language quiz generator utilizes reading materials prepared by the study material generator. Firstly, it checks users’ vocabulary level and provides them with reading exercises corresponding to their language level. It also generates quizzes to check the understanding of given vocabulary units after completing the selected reading exercise.

The program works in the following manner: the user is presented with a set of 20 polar (yes/no) questions to determine the initial level of language proficiency. Next, the user is assigned to one of the five levels of English proficiency according to

his/her final test score. The JLPT (Japanese Language Proficiency Test) scale, which corresponds with the JACET 8000 vocabulary list, has been used. Based on this selection, an appropriate text is selected and presented to the user in Japanese with randomly selected nouns replaced with English words. After finishing the reading exercise, users confirm their readiness to start the test phase. Next, users are presented with a set of a maximum of 20 test questions. The number of test questions depends on the user's language level, the difficulty of the selected reading and number of nouns extracted from the input file.

CO-MIX Japanese ・ 英単語学習システムへようこそ！

Correct! [Q 2/20] Please type the English word for 通訳 (i____e__r):
interpreter
Incorrect. It should be: 'interpreter' Example sentence: 叔父は僕に通訳をやらせた。 My uncle made me serve as interpreter. [Q 3/20] Please type the English word for 文化 (c____e):
culture
Correct! [Q 4/20] Please type the English word for 都市 (c__y):
city
Correct! [Q 5/20] Please type the English word for 欠勤 (a____e):
absance
Incorrect. It should be: 'absence' Example sentence: この間の金曜日の欠勤を弁解できるからね。 Can you account for your absence last Friday? [Q 6/20] Please type the English word for 生活 (l_v__g):
livi

Figure 7. Screenshot of system interface (Co-MIX Japanese)

The main purpose of these quizzes is to bring new vocabulary into productive use and enhance its memorization. The objective of each question is to type the word in English in response to the given vocabulary unit in Japanese. In the first phase of questions, random letters (not more than three and depending on word length) are presented to the user. If the user provides the system with correct answers, the test ends and automatic grading is provided. In the case of error, the system searches for an example sentence with the proper use of a given word and presents it to the user.

The system uses sentences from the Tanaka corpus¹², a large public-domain collection of parallel Japanese and English sentence pairs.

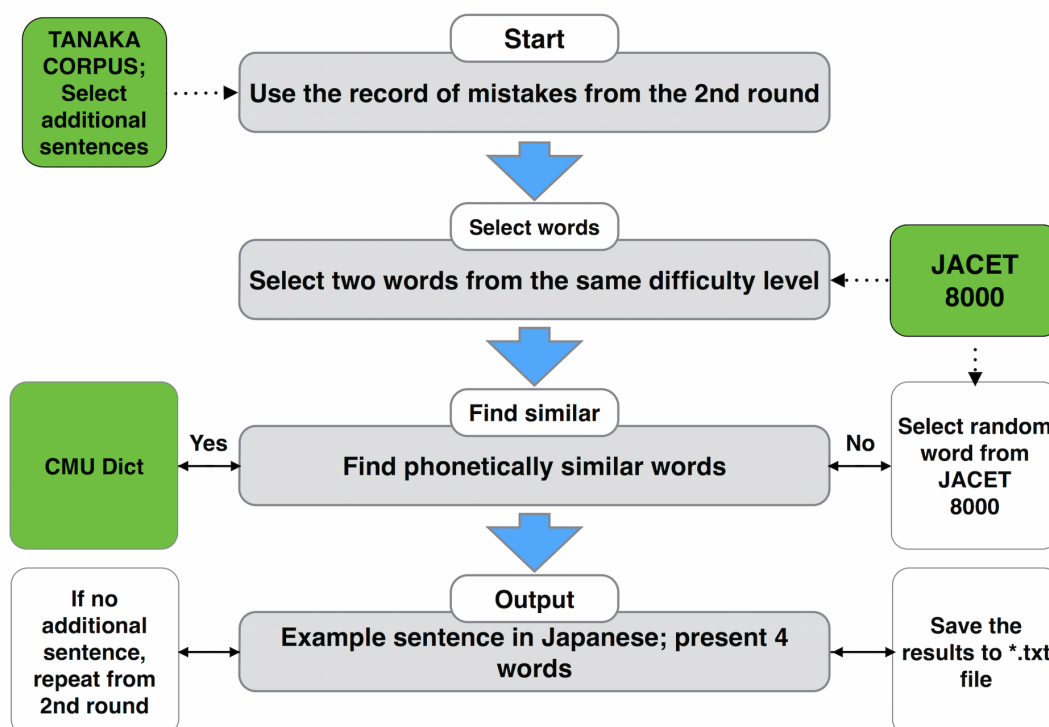


Figure 8. Process of third phase of the test (Co-MIX Japanese)

Next, mistaken words are repeated in the second phase of questions, where the difficulty level is slightly higher due to lack of support from random letters and the necessity to write the target vocabulary in the sentence. Similarly to the second phase, the system repeats user mistakes in the third and final phase (Figure 8), where four words are presented and user has to input the correct one.

However, one of the randomly selected words out of three is presented with phonetically similar variation, to make students aware of pronunciation similarities between English words and to indicate some vocabulary that could be easily mistaken, e.g. *write: reit, rite, wright; tale: tail; nose: noes, noos*. The CMU

¹² Tanaka Corpus: <http://dict.longdo.com/about/hintcontents/tanakacorporus.html>

Pronouncing Dictionary¹³, a machine-readable pronunciation dictionary, is used to select words with similar readings. To extract target vocabulary items from texts, the JUMAN Morphological Analyzer for Japanese¹⁴ is used. Finally, the Edict dictionary¹⁵ (Japanese-English) is used to extract English translations of word units.

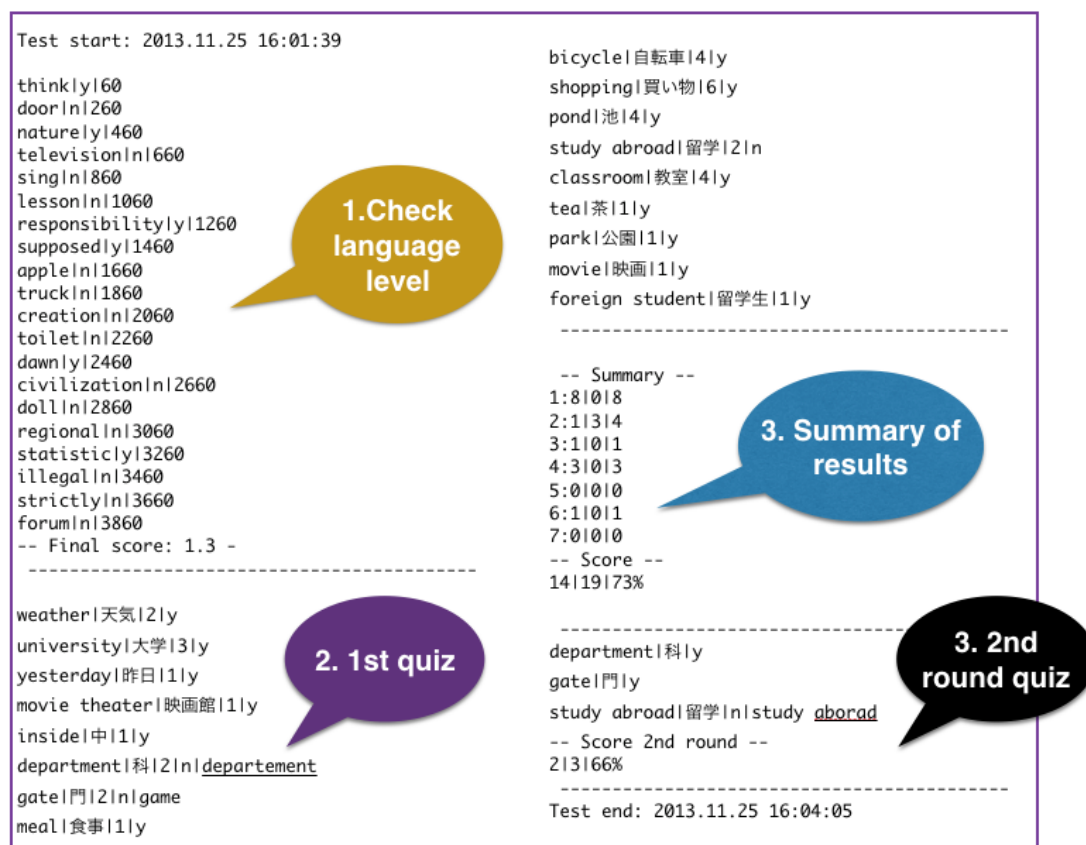


Figure 9. Result file of the language quiz generator (Co-MIX Japanese)

¹³ The CMU Pronouncing Dictionary: <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>

¹⁴ Juman: <http://nlp.ist.i.kyoto-u.ac.jp/index.php?cmd=read&page=JUMAN&alias%5B%5D=日本語形態素解析システムJUMAN>

¹⁵ Edict Dictionary: <http://www.edrdg.org/jmdict/edict.html>

5.3 Evaluation experiments

5.3.1 Participants

Forty test subjects (university students and adults) participated in two evaluation experiments and interacted with the proposed and baseline systems. The participants were recruited through flyers posted at the local universities and community centers. Twenty-four participants were university students (undergraduate and graduate) and 16 were adult English learners (aged between 24 and 43) who volunteered to participate in the survey (see Appendix C). The average age of the participants was 28.3 years. Among the respondents 21 were male and 19 were female.

The participants were selected using the following criteria: due to the fact that presented method requires basic English skills, such as reading ability and basic comprehension of English grammar, the target group were people who had some background in English education. Therefore, experiments were performed with Japanese university students and adults who had completed mandatory English education, thus represented comparable language ability. Test samples were only accepted if the number of years the participants studied English was similar. Thirty-eight respondents started English education in junior high school, and only two had been studying English from elementary school. Among the adults, eight people (50.0%) had been studying English for fewer than two years after graduating, three people for two to five years, and five people for more than five years.

5.3.2 Procedure

The procedure of the experiment was as follows:

1. Respondents were asked to interact with two presented systems. Sessions were conducted in a computer laboratory with 4 identical personal computers. Up to

4 respondents were tested simultaneously. Respondents were seated separately. They were informed about the basic research objectives. Although the delivery of instructions was fully automated, the experimenter remained present to answer possible questions and to prevent interaction among respondents. User results were automatically saved in a result file to be later used in data analysis.

2. Respondents were asked to fill out a paper survey after finishing the learning tasks with both systems (see Appendix C). All participants were informed about the purpose of the study and the confidentiality of the results. Respondents firstly filled out basic information: age, gender, and English education background.
3. Respondents rated both systems by marking 14 word pairs on a 7-point scale. The same test subjects participated in both experiments, but they were assigned to either start the experiment with the proposed or baseline system; i.e. half of the participants first interacted with the baseline system to assure the reliability of the results.
4. Respondents ranked both systems based on their overall performance.

5.3.3 Experiment method

In order to investigate users' attitudes towards the proposed and baseline systems, the semantic differential (SD) scale created by Osgood et al. (Osgood, Suci, Percy & Tannenbaum 1957) was used. Since it has been introduced, the SD Scale has been used in a variety of studies and number of contexts to measure people's attitude towards stimulus words and concepts. Heise (1965) describes SD scale as simple, effective mean for obtaining data on people's reaction towards a given object and provides many evidence that support the validity of the SD as a technique for attitude

measurement. In recent research, Urabe et al. (Urabe, Rzepka & Araki, 2014) used the SD Scale to investigate users' perception of an emoticon recommendation system; Ardito et al. (Ardito, Costabile, De Angeli & Lanzilotti, 2006) proposed semantic differential method to analyze users' satisfaction and usability in evaluation of e-learning systems.

The SD scale utilizes polar word pairs such as "good - bad", "positive - negative" and "valuable - worthless" to measure the meaning of a given concept. The position marked 4 is labeled "neutral," the 5th position is labeled "slightly," the 6th position "quite," and the seventh position "extremely. " Ratings on SD scale are often correlated and the average score for each pair of bipolar words indicates the correlation between the word and the object being tested. According to Heise (1970), three major dimensions of word meanings usually account for most of co-variation in ratings. Heise (1970) gives an example of numerous studies that used the SD scale and discusses the three major dimensions of words meanings: Evaluation, Potency and Activity (EPA). They were isolated through the factor analysis by analyzing hundreds of word pairs related to these three dimensions.

In this experiment 14 bipolar words (shown in Figure 10) were used with a seven-point scale to derive subjects' attitudes towards the given concepts and measure their opinions on a metrically controlled scale. Additionally, the polar word pairs were drawn from three general dimensions of word meanings: Evaluation (e.g., valuable – not valuable), Potency (e.g., attractive – unattractive), and Activity (e.g., fast – slow). The bipolar words were determined to represent all major dimensions of word meanings in order to evaluate the participants' attitude towards a computer system. In selection of appropriate word pairs the author referred to semantic differential profiles suggested by Heise (1965), and works by Urabe et al. (Urabe, Rzepka, Araki, 2014), Osgood et al. (Osgood, Suci, Percy & Tannenbaum 1957) and Newby (2010).

Randomization of word pairs was also performed by presenting concepts in a random order.

5.3.4 Experiment results

The language level of participants was determined by their number of correct answers in the initial vocabulary test described above (numbers close to zero represent an elementary level of word understanding; numbers close to five represent an upper-intermediate/advanced level of English vocabulary knowledge). The average vocabulary level was 3.61 for the proposed system (eight people reached the maximum score of five, and the lowest score was 0.7). As for the baseline system, the average score was 3.18 (three people finished the vocabulary test with the maximum score, and the lowest result was 0.4).

In the test results for the proposed system, six people reached 100.0% correct answers in the first round (15.0% of the users) for the proposed system, followed by 19 out of 34 in the second round (55.0%) and 12 out of 14 in the third round (85.0%). The average answering time for the proposed system was 11.6 minutes. As for the baseline system, 10 people reached 100.0% correct answers in the first round (25.0%), followed by 11 out of 30 (36.0% of respondents) in the second round and 16 out of 19 (84.0%) reaching the maximum score in the third round. The average answering time for the baseline system was 8.8 minutes.

5.3.5 Participants' attitudes on the SD Scale

The calculated average of system users' attitudes towards the proposed and baseline systems were collected using the SD Scale. In Figure 10, numbers closer to one represent the part of the spectrum assigned to the words on the left, whereas numbers close to seven are better characterized by the words on the right. Two different colors were used to present systems: black for the proposed system and grey for the baseline. Pair of bipolar words are translated and presented in English, but the

original word pairs in Japanese used in the survey are presented in Appendix B for additional reference.

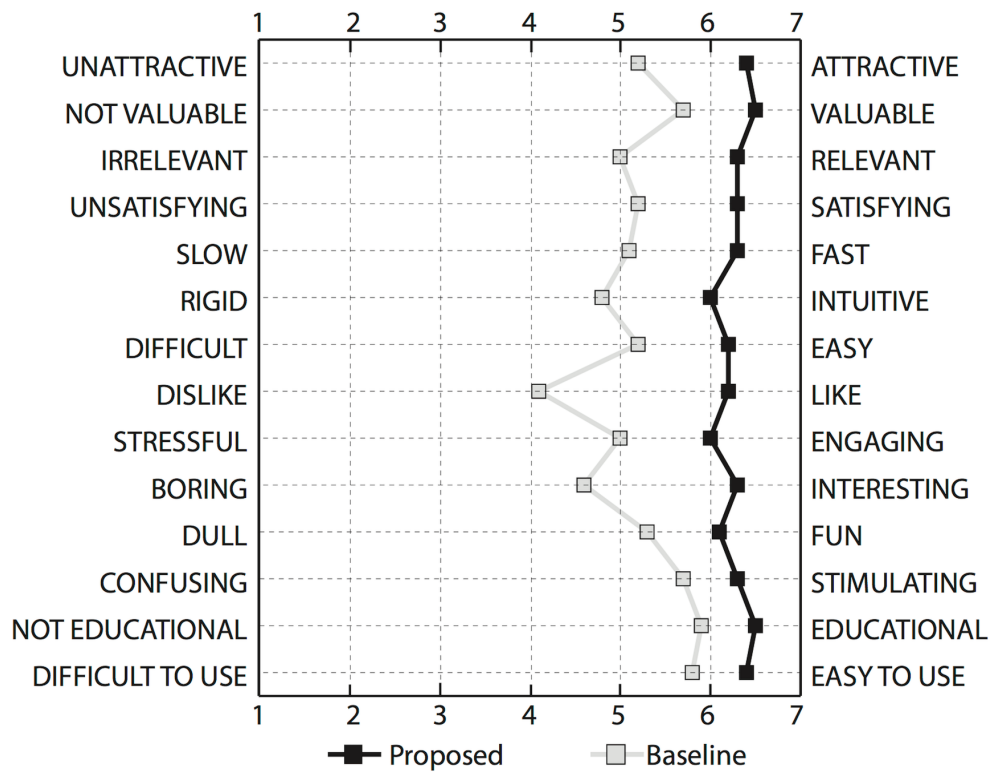


Figure 10. Semantic differential scale for proposed and baseline systems (Co-MIX Japanese)

After analyzing the results shown in Figure 10, the conclusion was reached that the average results of the proposed system were distinctly higher than the baseline. The overall average of the proposed system was 6.2 points, which was superior to the baseline system (5.1 points). The general trend line for the proposed system stayed within 6/7 at the higher end of the scale, whereas items in the baseline system such as “like/dislike” and “interesting/boring” diverted from the given trend line. On the other hand, three items (“stimulating/confusing”, “educational/not educational”, “easy to use/difficult to use”) had similar results in both systems, which indicates general user satisfaction with both systems.

5.3.6 Factor analysis

Two separate sets of statistical tests were carried out on the baseline (System B) and proposed (System A) systems using the result data acquired from the semantic differential analysis (see Figure 10). The reason for this form of analysis was that different questionnaires were given to respondents to evaluate each system independently.

Table 3. Factor analysis results for the proposed system. The highest significance is in bold font.

Proposed (System A)	Factors	
	1	2
Value	0.82	
Attractiveness	0.80	
Stimulating	0.80	0.67
Satisfaction	0.80	0.66
Fun	0.77	0.70
Relevance	0.77	0.60
User Interest	0.69	0.51
Learning Speed	0.64	0.57
Easy to use	0.57	
Engaging	0.64	0.84
Likeability	0.54	0.68
Educational	0.57	0.65
Intuitive		0.61
Difficulty		0.58
Eigenvalue	6.71	1.10
Explained Variance	47.98	7.92
Total	55.90	
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Items loading <0.50 were deleted for no statistical significance.		

Since putting both sets of data together did not generate reliable comparative results, performing further analysis was decided to evaluate each set of data separately and then compare their characteristic features. Data from the system evaluations were submitted to Factor Analysis (on SPSS 21). KMO & Bartlett's Test were run on the data for the correlation test and the sampling size to check if the sampling size was fit for factor analysis.

Table 4. Factor analysis results for baseline system. The highest significance in bold font.

Baseline (System B)	Factor	
	1	2
Stimulating	0.88	0.72
Engaging	0.83	0.50
Likeable	0.81	0.76
Easy to use	0.77	0.60
Intuitive	0.74	0.50
Educational	0.68	0.57
Learning Speed	0.63	
Difficulty	0.57	
Attractiveness	0.55	0.85
Fun	0.80	0.85
Value		0.82
User Interest	0.69	0.80
Relevance	0.67	0.80
Satisfaction	0.60	0.75
Eigenvalue	7.76	1.26
Explained Variance	55.43	9.03
Total	64.46	
Extraction Method: Principal Axis Factoring. Rotation Method: Promax with Kaiser Normalization. Items loading <0.50 were deleted for no statistical significance.		

The test resulted with 0.91 sampling adequacy for the proposed system and 0.93 for the baseline system, which means that the samples were adequate and therefore fit for factor analysis and the p-value (significance level) for both tests was 0.01.

A second set of analysis, Principal Axis Factoring, was run on the data with the two extracted factors to validate the results of the preliminary statistical test. Items loading <0.50 in whichever factor were deleted for having no statistical significance. For the proposed system, two factors were extracted explaining 55.9% of the data variances. As for the baseline system, 64.4% of the data variances were explained with the extracted two factors.

In the baseline system (see Table 4) the items “Stimulating”, “Engaging”, “Likable” and “Fun” were extracted with high rates in Factor 1, named Likability. In Factor 2 five items, “Attractiveness”, “Fun”, “Value”, “User Interest” and “Relevance” named Educational Value were rated highly.

The items for Educational Value were mostly rated highly in Factor 1 for the proposed system, whereas in the baseline system they were extracted in Factor 2, e.g. the item “Valuable” was rated 0.82 in Factor 1 for system A and 0.30 in Factor 1 for system B. It can be interpreted that the item “Valuable” was selected more significantly for the proposed system.

For the proposed system (see Table 3), the items “Value”, “Attractiveness”, “Stimulating” and “Satisfying” were rated high in Factor 1, named Educational Value. This means that a significant number of the respondents gave high rates to the latter semantic values. “Fun” and “Engaging” were rated high in Factor 2, and named Enjoyment.

5.3.7 User rankings

In the last step, the respondents were asked to rank the two systems based on performance, and state which of the two systems they prefer. The results of the survey indicate that the majority of participants selected the proposed system as performing better, and expressed their eagerness to use it in the future. Based on this result, it can be asserted that the proposed system achieved a significant improvement over the baseline system in terms of system performance and user preferences. Below, the results of the ranking are presented:

- **Which system do you think is superior?** 35 out of 40 (87.5%) chose the proposed system
- **Which system would you like to use more after the experiment?** 35 out of 40 (87.5%) chose the proposed system
- **Which system do you like more?** 35 out of 40 (87.5%) chose the proposed system

5.3.8 Analysis of user mistakes

The average number of mistakes the participants made was higher for the baseline system in all three rounds of the vocabulary knowledge evaluation test (Table 5).

Table 5. Average number of mistakes for proposed and baseline systems. Lower number in bold.

SYSTEM	1st round mistakes	2nd round mistakes	3rd round mistakes
Proposed	2.62	0.62	0.05
Baseline	3.75	1.10	0.07

The wide gap between the number of mistakes in the first round could support an assumption that users were able to memorize new vocabulary better in the first round. Most of the mistakes the users made can be analyzed by categorizing them in three groups.

The first group, which was labeled as “**transfer mistakes**”, is related to the phonetic matching of loan words and lexical borrowings in the Japanese language. It corresponds to typical phonological problems many Japanese students encounter while learning English. Some examples that fit into this category are mistakes like: “*conbersashon*” or “*conberstation*” (conversation), “*pamito*” (permit), “*pasonal*” (personal) and “*abusensu*” (absence), “*curassu*” (classroom), “*libing*” (living), “*buroku*” (block) and “*elejible*”(eligible). The words in parenthesis are user input words and the correct answers are shown in brackets.

The second group is referred to as “**vocabulary mistakes**”, and it includes some obvious examples of erroneous target L2 word associations, i.e., students associated the target English word with some other word they may previously have heard. For example: “*creation*” (criterion), “*convocation*” (conversation), “*action*” (activity), “*thick*” (trick), “*mood*” (mind), “*conference*” (conversation), “*scene*” (screen), “*mails*” (means), “*root*”(rest), “*becoming*” (beginning) and “*preparation*” (presentation).

The third group was distinguished as “**spelling mistakes**”, which contains typical spelling mistakes that happen to many learners of second language: simple performance errors that include random guesses, slips, typos, etc. For example: “*burau*” (bureau), “*univercity*” (university), “*wintin*” (within), “*bycycle*” (bicycle), “*interduction*” (introduction), “*studying abroud*” (studying abroad), “*hourely pay*” (hourly pay), “*expearence*” (experience), “*libing*” (living), “*biginning*” (beginning) and “*screan*” (screen).

5.4 Discussion

Experiment results seem to concur with the findings by Chen (1998), who showed that second language learners commonly lack sufficient word knowledge to support proper guessing of the word meaning from a given reading context. Therefore, in the case of the baseline system, the chance of comprehending complex graded readers presented entirely in the second language and effectively acquiring new vocabulary was significantly lower. There is a visible gap between the second round of test between two systems, and higher number of correct answers in the case of the proposed system may indicate higher effectiveness of code-switching method, not present in the baseline system. This confirms the previous assumption that the users benefit from code-switching method and its presence will have a positive impact on their test scores.

Another insight comes from the fact that the time users spent with the baseline system was shorter. It also indicates the advantage of code-switching based method used in the proposed system, with higher number of correct answers and lower number of users' mistakes. Analysis of user mistakes confirms another assumption, and demonstrates that the number of spelling mistakes was higher for the baseline system. Smaller number of mistakes indicate that the proposed system helped students to better memorize the spelling of new words in a foreign language. Some opinions contributed by the experiment participants (see Appendix A) also indicate that the proposed system was more encouraging to interact with, while the baseline system seemed to discourage some participants to sufficiently elaborate on a given task, and thus was used for a shorter duration than the proposed system.

The results of the experiments answer the first and second research questions and clearly show that the code-switching phenomenon applied to an e-learning system can expand students' vocabulary and enable them to understand many words without

given definitions. These findings are consistent with the previous research, in which most of the participants successfully remembered and recreated a series of test sentences in a mixed language, thus proving code-switching based method to be an effective way of acquiring new vocabulary (Mazur et al., 2012).

Judging from both the results and analyses, it can be asserted that user attitudes to the proposed system were positive and in some categories it significantly outperformed the baseline system, which answers the third research question. Additional rankings, based on the systems' performances and users' preferences, also indicated the proposed system as more favorable. Another insight in the proposed system's educational value and general user satisfaction was provided by the factor analysis results. It demonstrated that the users had more positive impressions towards the proposed system, especially in terms of its educational value and users' preferences.

Although the number of answers ranking the proposed system as preferable was the same for each of the questions, there were some differences in the answers. Only three out of 40 respondents indicated the baseline system as their first choice (7.5%) and two participants chose a different system in each category, which may indicate that they saw the proposed system as superior, but liked the other one more (13.0%). There are many possible explanations, one of them emphasizing the baseline system as more similar to the methods of studying previously encountered by some users. For these users the baseline system seems to be more challenging. For further explanation, some relevant examples of user comments after using both systems are presented in Appendix A.

5.5 Co-MIX English

5.5.1 System introduction

This subsection presents the preliminary implementation of a new system that follows the previous work on Co-MIX Japanese system to confirm that the proposed code-switching-based method can also be adapted in other languages. This work also utilizes my proposed vocabulary teaching method for e-learning systems that harnesses the benefits of code-switching, incidental learning and advantages of learning vocabulary in context. This method has been successfully implemented in Co-MIX Japanese for Japanese learners of English. This section presents another implementation of the proposed code-switching-based method. This work in progress is called the Co-MIX English system. Opposite to the previous research aimed at Japanese native speakers, this new system is aimed at English-speaking foreigners who want to study Japanese language vocabulary.

Similar to the previous system, the modular architecture allows easy extensions of the Co-MIX English system. The current implementation consists of two modules. The first one is responsible for the automatic generation of study materials based on given graded readers in the first language (Fig. 11). The other module, the Language Quiz Generator, checks the users' understanding of second language vocabulary through a series of language tests. The system also responds to users' mistakes by providing them with example sentences that show the correct usage of a given word.

In the first step the Co-MIX English system evaluates user language competency and presents study article appropriate to the user's vocabulary knowledge level. The text is used to introduce new vocabulary units and the proposed system is provided with a selection of graded articles in Japanese. The baseline system is a variation of the proposed system for experimental purposes, without the code-switching module. It is equipped with some commonly used vocabulary learning techniques, however, it lacks the code-switching method presented in the proposed

system. Another difference comes from the fact that instead of using readers in a mixed language with code-switching, the baseline system presents reading exercises entirely in L2 using the set of articles manually translated and proofread by a native speaker.

5.5.2 Study material generator

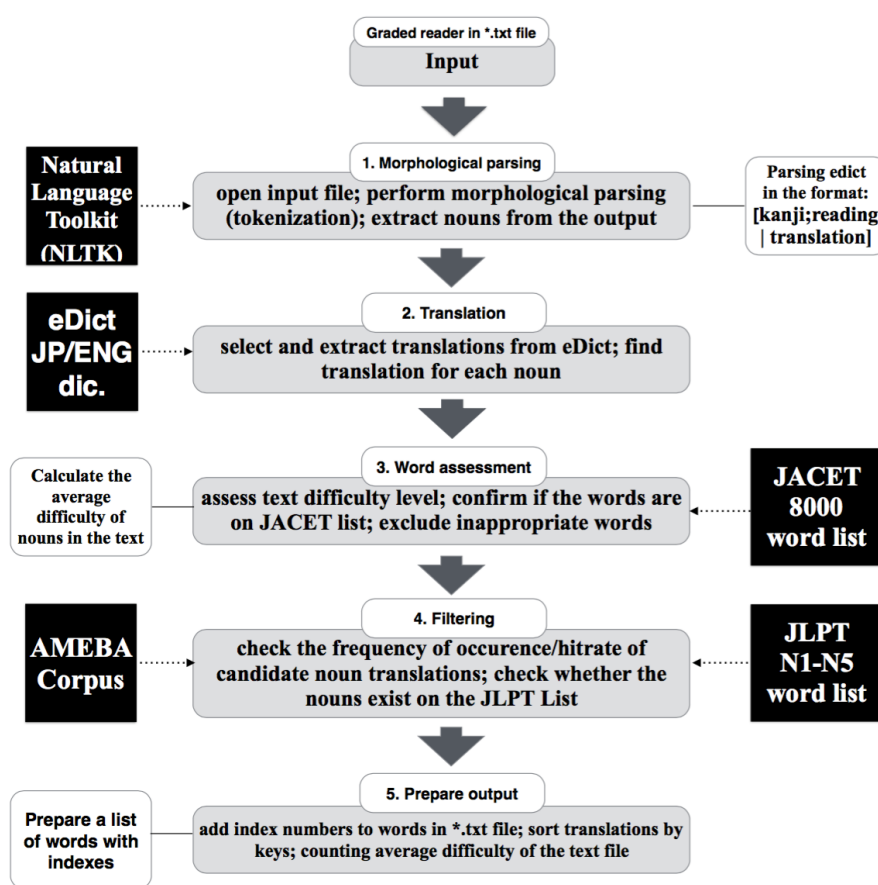


Figure 11. Process of study material generator (Co-MIX English)

The main function of this application is to automatically create editable learning materials from a given text file. The study material generator is fed with text files with graded readers in a given language and automatically generates an editable output file with tagged vocabulary items (nouns), their translations in second language

for the proposed system, and their positions on the JACET 8000 vocabulary list (Uemura & Ishikawa, 2004). In case of the proposed system we use the JLPT vocabulary list¹⁶. Vocabulary lists are used to assess users' language ability and their prior knowledge. The system is using standardized lists of over 8,000 JLPT¹⁷ words required for the exam across all 5 levels. Both lists provide the average difficulty of a text, which allows the system to categorize the text to an appropriate proficiency level.

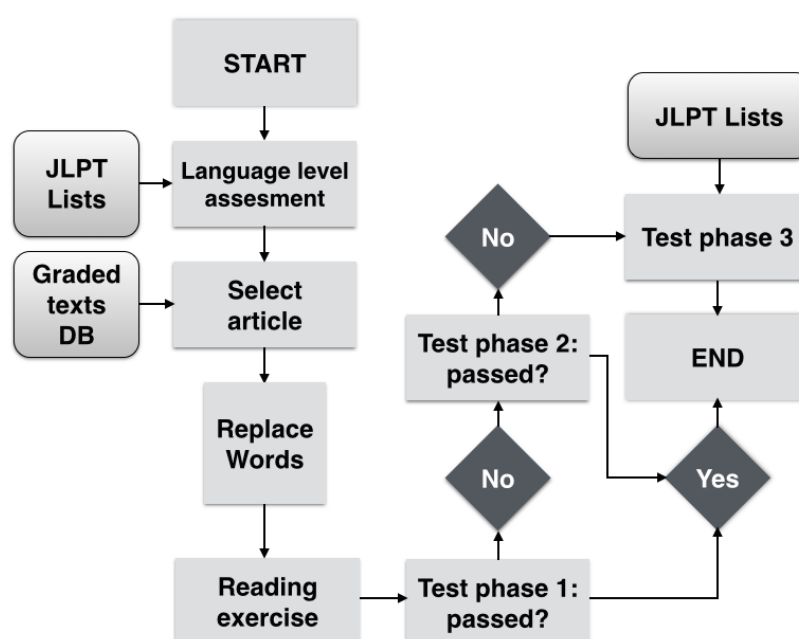


Figure 12. Process of language quiz generator (Co-MIX English)

Usually, preparing suitable reading materials for students presents a challenge for teachers. The study material generator can be used to manually create additional learning materials from given input texts. The process of the study material generator is as follows. First program searches the input file for a selected part of speech (e.g.

¹⁶ The Japanese Language Proficiency Test, or JLPT, is a standardized test to evaluate and certify the Japanese language proficiency of non-native speakers. "The JLPT is offered in five levels (N1, N2, N3, N4, N5). In order to measure Japanese-language proficiency as thoroughly as possible, test items are designed for each level." - Japanese Language Proficiency Test, accessed 1 October 2014, < <https://www.jlpt.jp/e/about/points.html> >.

¹⁷ This selection is based on a database of about 18,000 vocabulary words that have appeared on the test with high frequency over the past 20 years.

noun) and checks if their English translations are available in the dictionary. Then the program analyzes the word rankings in a given reading and an average difficulty level is provided for the whole article. In the next step, the system generates an output text with selected words preceded by an index number and supplemented with a list of all chosen words, their proposed translations and their difficulty ranking. In this step, the human teacher is allowed to review the result and eliminate possible mistakes that may be caused by limitations of the available list word coverage. This is a necessary step before handing over the materials to the language quiz generator and presenting them to users. A few search rules had to be established. This study focused on nouns, which have an available translation in the dictionary, have a translation that consists of no more than three words, are on our frequency list and are not on the list of the first 500 most common words that students are most likely to know.

5.5.3 Language quiz generator

The language quiz generator utilizes reading materials prepared by the study material generator (Fig. 12). In the first step, it checks users' vocabulary level and provides them with reading exercises corresponding to their language level. It also generates quizzes to check the understanding of given vocabulary units after completing the selected reading exercise. The program works in the following manner: the user is presented with a set of 20 polar (yes/no) questions to determine the initial level of language proficiency.

Next, the user is assigned to one of the five levels of language proficiency according to his/her final test score. The JLPT Scale was used, because it corresponds with the JACET 8000 vocabulary list (Japan Association of College English Teachers List of 8000 Basic Words) described by Uemura & Ishikawa (2004). Based on this selection, an appropriate text is selected and presented to the user in English with

randomly selected nouns replaced with Japanese words. After finishing the reading exercise, users are presented with a set of a maximum of 20 test questions.

CO-MIX English ・ 日本語単語学習システムへようこそ！

Please fill in the missing word:

[Q 1/20] Please type the Japanese word for friend (t____a_hi):

tomodachi

Correct!

[Q 2/20] Please type the Japanese word for park (k_ue_):

公園

Correct!

[Q 3/20] Please type the Japanese word for center (c____shi_):

ちゅうしん

Correct!

[Q 4/20] Please type the Japanese word for place (tok_r_):

ところ

Correct!

[Q 5/20] Please type the Japanese word for shopping (k____no):

買い物

Correct!

[Q 6/20] Please type the Japanese word for stationery (bu____ugu):

文房具

Figure 13. Screenshot of system interface (Co-MIX English)

The number of test questions depends on the user's language level, the difficulty of the selected reading, the number of available example sentences, and number of nouns extracted from the input file. The main purpose of these quizzes is to bring new vocabulary into productive use and enhance its memorization. The objective of each question is to type the word in second language. In the first phase of questions, random letters (not more than three and depending on word length) are presented to the user.

If the user provides the system with correct answers, the test ends and automatic grading is provided. In the case of error, the system searches for an example sentence with the proper use of a given word and presents it to the user. The sentences from the Tanaka corpus¹⁸, a large public-domain collection of parallel

¹⁸ Tanaka Corpus of parallel Japanese-English sentences: http://www.edrdg.org/wiki/index.php/Tanaka_Corpus

Japanese and English sentence pairs, were used. Next, mistaken words are repeated in the second phase of questions, where the difficulty level is slightly higher due to lack of hints and the necessity to write the target vocabulary in the sentence. Similarly to the second phase, the system repeats user mistakes in the third and final phase, where four words are presented and user has to select the correct answer.

[Test 1] Please fill in the missing word:

[Question 1/20] Please type the Japanese word for "tea" (o _ h _):

> ocha

Correct!

[Question 2/20] Please type the Japanese word for "center" (c _ _ _ _ i _):

> 中心

Correct!

[Question 3/20] Please type the Japanese word for "place" (t _ _ _ r _):

> ところ

Correct!

[Question 4/20] Please type the Japanese word for "exchange" (k _ _ _ a n):

> kokan

Incorrect. It should be: '交換 (こうかん)'.

Example sentence:

"このセーターを大きいのと交換していただけますか。"

"Will you exchange this sweater for a larger one?"

[Question 5/20] Please type the Japanese word for "food" (t _ _ _ m _ n _):

> たべ

Figure 14. Output of Language Quiz Generator (Co-MIX English)

5.6 Evaluation

5.6.1 Participants

Sixteen test subjects (university students and adults) participated in a preliminary evaluation experiment and interacted with the proposed system (Co-MIX English). Among the respondents 12 were male and four were female. Ten participants were in their twenties and six participants were over 30 years old. In this

experiment the target group were people who had some previous Japanese education, due to the fact that presented method requires basic Japanese skills, such as reading ability (hiragana, katakana and some basic Japanese kanji characters).

Therefore, the experiments were performed with university students and adults with comparable language ability (the number of years both groups had studied foreign language was similar). 16 samples were also selected from baseline system experiment without code-switching to compare user attitudes towards baseline and proposed systems. Sixteen samples were selected to correspond with age, gender and education level of the proposed system users'.

All participants were informed about the purpose of the study and the confidentiality of the results. Ten participants were adults (eight company employees, one self-employed and one housewife) and six were university students (undergraduate and graduate). Ten respondents (62.5%) have been learning Japanese from 2 to 5 years, three have been learning from 1 to 3 years, two have been studying less than 6 months, and one have been studying more than 5 years. Seven respondents previously took the JLPT Exam (4 of them took Level 4, 2 of them level 5 and 1 of them Level 3). Thirteen participants also expressed their interest in taking JLPT exam in the future, two participants did not plan to take an exam, and one was not sure. Seven people were interested in taking Level 3, two participants wanted to start with Level 5, two participants were aiming at Level 2, one participant wanted to try Level 4 and one aimed at the highest Level 1.

5.6.2 Semantic Differential Scale

In this experiment, 14 bipolar words and a seven-point scale to derive subjects' attitudes towards the given concepts and measure their opinions on a

metrically controlled scale were used. The bipolar words were determined based on the most typical pairs of opposite expressions used to evaluate the peoples' attitude towards a computer system. The SD scale was selected as a conceptual model of assessing the value of e-learning system because it fits in the framework to measure overall users perceived satisfaction with an e-learning system and learners behavior/usage of the system proposed by Levi (2006).

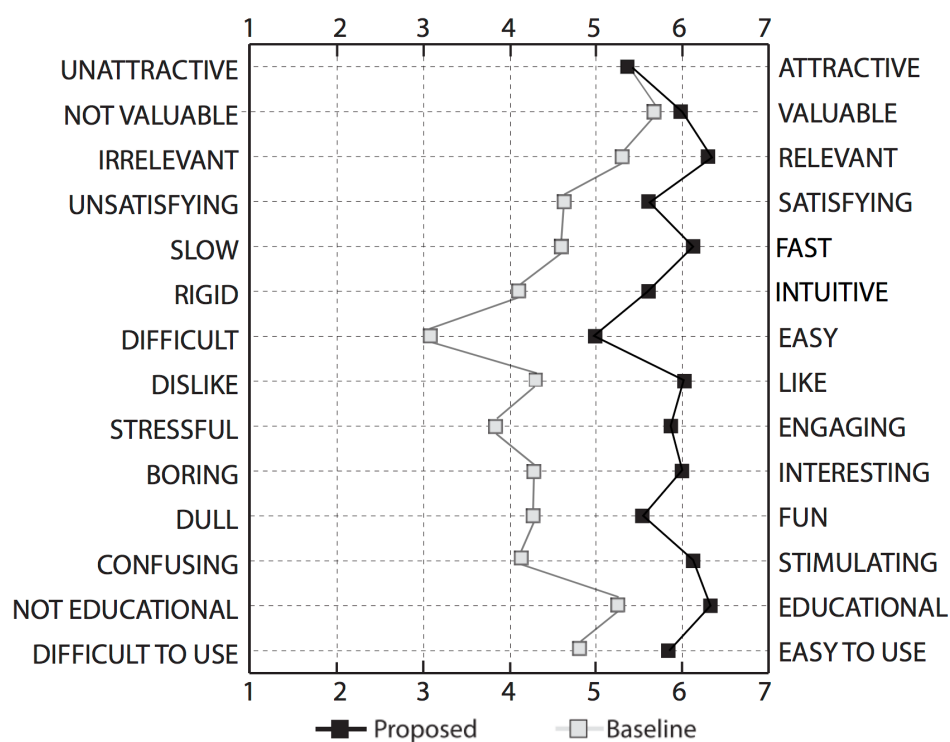


Figure 15. Semantic differential scale for proposed and baseline systems (Co-MIX English)

The participants of the experiment were asked to fill out a paper survey after they finished the learning tasks with a given system (see Appendix D). In order to investigate users' attitudes towards the system, the semantic differential (SD) scale created by Osgood (1957) was used. The SD scale utilizes polar words pairs such as “good - bad”, “positive - negative” and “valuable - worthless” to measure the meaning

of a given concept. The position marked 4 is labeled "neutral," the 5th position is labeled "slightly," the 6th position "quite," and the seventh position "extremely".

5.6.3 Experiment results

The language level of each participant was determined by their number of correct answers in the initial vocabulary test. The average vocabulary level was 2.09 which corresponds to the Level 4, according to JLPT scale. In the test results only one person reached 100.0% correct answers in the first round, followed by 5 out of 15 in the second round (33.3%) and 7 out of 10 in the third round (70.0%). The average answering time for the proposed system was 11.8 minutes. In case of the baseline system, 2 people out of 16 reached 100.0% correct answers in the first round (12.5%), followed by 4 out of 14 (28.5% of respondents) in the second round and 10 out of 10 (100.0%) reaching the maximum score in the third round. The average answering time for the baseline system was 8.8 minutes.

The calculated average of system users' attitudes towards the Co-MIX Japanese system using the SD Scale were collected. In Fig. 15., numbers closer to one represent the part of the spectrum assigned to the words on the left, whereas numbers close to seven are better characterized by the words on the right. After analyzing the results, it can be asserted that the average results of the proposed system were distinctly higher than the baseline. The overall average of the proposed system was 5.8 points, which was superior to the baseline system (4.5 points). The general trend line for the proposed system stayed within 5/6 at the higher end of the scale, whereas items in the baseline system such as "like/dislike" and "interesting/boring" diverted from the given trend line. On the other hand, three items ("stimulating/confusing", "educational/not educational", "easy to use/difficult to use") had similar results in both systems, which indicates general user satisfaction with both systems.

Chapter 6

Conclusions and Future Works

6.1 Overall conclusions

In chapter 5 of this thesis, Co-MIX Japanese, a system for English vocabulary acquisition in context, which uses a code-switching based approach for Japanese learners, was presented. The presented results demonstrated evidence that the proposed method is useful for students' second language vocabulary acquisition. Experiments with the proposed and baseline systems and the analysis of users results also showed that the proposed method is an effective way of expanding students' knowledge of a given second language vocabulary.

The high percentage of correct answers, making successful connections between meanings of words in two languages and positive comments from the participants (see Appendix A) suggest that a system for learning English vocabulary in context using a code-switching based approach has the potential to provide Japanese learners with engaging learning activities with an educational value. The participants improved their vocabulary scores after using the application, which demonstrates the pedagogical benefits of the presented code-switching-based method. The users of the proposed system were also able to memorize new second language vocabulary better than those who used the baseline system. Higher user satisfaction with the proposed system also provides motivation to improve and refine it in a number of different ways in the future. Judging from the results, it can be asserted that user attitudes towards the proposed systems were positive and in most categories they significantly outperformed the baseline system. Additional rankings, based on the systems'

performances and users' preferences, also indicated proposed systems as more favorable.

Chapter 5 of the thesis also presented the most current work on the Co-MIX English system. This is another implementation of an experimental method for second language vocabulary acquisition in context using a code-switching based approach for studying Japanese vocabulary. Further experiments confirmed the results of a previous study on a system for English vocabulary acquisition based on code-switching. The high percentage of correct answers, making successful connections between meanings of words in two languages and positive comments from the participants suggest that system for learning English vocabulary is a successful continuation of our previous work on Co-MIX Japanese system. The participants improved their vocabulary scores after using the application, which demonstrates the pedagogical benefits of the presented code-switching-based method. The experiment also demonstrated some evidence that the proposed method is useful and effective way of expanding students' second language vocabulary.

6.2 Future works

There are various ways in which our system could be further improved. In the previous research, a foundation for further research on artificial tutors was created. Both presented Co-MIX systems are a step towards the general idea of creating a conversational system that would act as a second language tutor and not only transmit knowledge to students, but also offer a variety of other functions including a representation of emotions to make it more human-like and react to the changing emotional states of the users.

As for the future improvements, the Co-MIX project will aim to incorporate additional functionalities and advance towards the initial idea to provide a chatterbot

able to process user queries in mixed languages, as presented in Chapter 2. Such a Co-MIX system enabled with an artificial tutor could teach users second language vocabulary using the presented code-switching method in a human-like conversation.

It is also planned to take into consideration suggestions provided by the experiment participants in their comments (see Appendix A) and extend the system to support other languages. The first step in this direction was already taken by introducing Co-MIX English. The preliminary experiment provides some evidence that presented code-switching-based method can be easily implemented in other languages with a proper morphological analyzer, dedicated study materials (graded readers) and high frequency word lists.

The next step will be to compare the previous research on the Co-MIX Japanese system with the Co-MIX English system. Since my latest work on the system dedicated to aid English-speaking users in studying Japanese vocabulary and its initial results gave more implications on possible adaptation of this method into other languages, performing more experiments to compare both systems performance is planned. With the constant improvement of proposed method, I believe this approach to second language vocabulary acquisition in E-learning systems may bring even better results in the future.

References

Abdali, M. (2010). Teaching Vocabulary through Code-Mixing. *Peyke Noor Journal*, 8(1), 138-145.

Ardito, C., Costabile, M. F., De Angeli, A., & Lanzilotti, R. (2006). Systematic evaluation of e-learning systems: an experimental validation. In Morch, A., Morgan. K, Bratteteig, T., Ghosh, G., Svanaes, D. (eds.), *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (pp. 195-202). New York, NY: ACM.

Bogdanovych, A., Simoff, S. J., Sierra, C., & Berger, H. (2005). Implicit training of virtual shopping assistants in 3D electronic institutions. In: *Proceedings of the IADIS International e-Commerce* (pp.50–57). Lisbon: IADIS Press.

Borovsky, A., Elman, J., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meaning. *Language, Learning & Development*, 8(3), 278-302.

Bower, G. H. (1992). How might emotions affect learning. *The handbook of emotion and memory: Research and theory*, 3, 31.

Brennan, K. (2006). The managed teacher: Emotional labour, education, and technology. *Educational Insights*, 10(2), 55-65.

Camilleri, V., & Montebello, M. (2008). SLAVE–Second Life Assistant in a Virtual Learning Environment. *RELIVE08–Researching Learning in Virtual Environments*.

Celik, M. (2003). Teaching vocabulary through code-mixing. *ELT Journal*, 57(4), 361-369.

Chamberlain, W. *The Policeman's Beard is Half Constructed*. (1984). Warner Books.

Chen, H. J. (1998). A preliminary investigation on Taiwanese EFL learners' vocabulary size. In *Proceedings of the fifteenth conference on English teaching and learning in the Republic of China* (pp. 193-209). Taipei: Crane Publishing Co.

Christensen, E., Merrill, P. & Yanchar, S. (2007). Second language vocabulary acquisition using a diglot reader or a computer-based drill and practice program. *Computer Assisted Language Learning*, 20(1), 67-77.

Clark, E., & Araki, K. (2011). Text normalization in social media: progress, problems and applications for a pre-processing system of casual English. *Procedia-Social and Behavioral Sciences*, 27, 2-11.

Coady, J., & Huckin, T. (1997). *Second language vocabulary acquisition: A rationale for pedagogy*. Cambridge: Cambridge University Press.

Colby, K. *Artificial Paranoia: A Computer Simulation of Paranoid Process*. Pergamon Press, New York, 1975.

Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL*, 20(01), 98-116.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 57(3), 402-423.

- Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media*, 29(3), 241-250.
- Crockett, K., James, O. S., & Bandar, Z. (2011). Goal orientated conversational agents: Applications to benefit society. In *Agent and Multi-Agent Systems: Technologies and Applications* (pp. 16-25). Springer Berlin Heidelberg.
- Crystal, D. (2012). *English as a global language* (2nd edition). Cambridge: Cambridge University Press.
- Day, R., Omura, C. & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, 7(2), 541-551.
- Dybala, P., Ptaszynski, M., Rzepka, R., & Araki, K. (2009). Activating Humans with Humor--A Dialogue System That Users Want to Interact with. *IEICE TRANSACTIONS on Information and Systems*, 92(12), 2394-2401.
- Dybala, P., Ptaszynski, M., Maciejewski, J., Takahashi, M., Rzepka, R., & Araki, K. (2010). Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, 2(1), 31-48.
- Ekman, P. (1999). Facial expressions. *Handbook of cognition and emotion*, 16, 301-320.
- Fryer, L., & Carpenter, R. (2006). Bots as language learning tools. *Language, Learning & Technology*, 10(3), 8.

Gibbons, J. (1987). *Code-mixing and code choice: A Hong Kong case study*. Clevedon: Multilingual Matters.

Ginsburg, J. (2012). Automatic generation of English lesson materials for native speakers of Japanese, In *Proceedings of the 2012 Joint International Conference on Human-Centered Computer Environments* (pp. 14-18).

Hamers, J. F., & Blanc, M. H. (2000). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.

Heise, D. R. (1965). Semantic differential profiles for 1,000 most frequent English words. *Psychological Monographs: General and Applied*, 79(8), 1.

Heise, D. R. (1970). The semantic differential and attitude research. In Summers, G.F. (Ed.), *Attitude measurement* (pp. 235-253). Chicago: Rand McNally.

Higuchi, S., Rzepka, R., & Araki, K. (2008, October). A casual conversation system using modality and word associations retrieved from the web. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 382-390). Association for Computational Linguistics.

Hingston, P. (2009). A turing test for computer game bots. *Computational Intelligence and AI in Games, IEEE Transactions on*, 1(3), 169-186.

Honna, N. (2008). *English as a Multicultural Language in Asian Contexts: Issues and Ideas*, 121-122. Tokyo: Kuroshio Publishers.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, 11(2), 207-233.

Jia, J. (2009). CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning. *Knowledge-Based Systems*, 22(4), 249-255.

Jamshidi, A., & Navehebrahim, M. (2013). Learners Use of Code Switching in the English as a Foreign Language Classroom. *Australian Journal of Basic and Applied Sciences*, 7(1), 186-190.

King, K.A. Child language acquisition. (2006). In: Fasold, Connor-Linton. *An Introduction to Language and Linguistics*, Cambridge University Press, pp. 205-224.

Krashen, S.D. (1985). *The Input Hypothesis: Issues and Implications*. New York: Longman.

Labutov, I., & Lipson, H. (2014). Generating Code-switched Text for Lexical Learning. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Volume 1: Long Papers* (pp. 562–571). Baltimore: Association for Computational Linguistics.

Lanza, E. (1997). Language contact in bilingual two-year-olds and code-switching: Language encounters of a different kind?. *International Journal of Bilingualism*, 1, 135-162.

Latham, A., Crockett, K. A., & Bandar, Z. (2010). A Conversational Expert System Supporting Bullying and Harassment Policies. In *ICAART (1)* (pp. 163-168).

Lester, J., Branting, K., & Mott, B. (2004). Conversational agents. *The Practical Handbook of Internet Computing*.

Lin, A. (2013). Classroom code-switching: Three decades of research. *Applied Linguistics Review*, 4(1), 195-218.

Ling, Z. (2010). Effectiveness of L2 Vocabulary Learning and LT-WM. *Journal of Xinjiang University (Philosophy, Humanities & Social Sciences)*, 6, 135-139.

Lokman, A. S., Zain, J. M., Komputer, F. S., & Perisian, K. (2009). Designing a chatbot for diabetic patients. In *International Conference on Software Engineering & Computer Systems, ICSECS*.

Macaro, E. (2001). Analysing student teachers' codeswitching in foreign language classrooms: Theories and decision making. *The Modern Language Journal*, 85(4), pp. 531-548.

Matthias, R. E. H. M., & André, E. (2005). From Chatterbots to Natural Interaction: Face to Face Communication with Embodied Conversational Agents. *IEICE transactions on information and systems*, 88(11), 2445-2452.

Mazur, M., Rzepka, R., & Araki, K. (2010). Co-MIX Project: Towards Artificial Tutors Using Code Mixing as Foreign Language Teaching Method. In *Proceedings of IWMST (2010)*, pp.196-201.

Mazur, M., Rzepka, R., & Araki, K. (2011). Proposal for a conversational English tutoring system that encourages user engagement. In *Proceedings of the 19th International Conference on Computers in Education* (pp. 10-12).

Mazur, M., Rzepka, R., & Araki, K. (2012). Mixing Words and Emotions: New Natural Methods for Artificial Language Tutors (人工知能学会全国大会 (第 26 回) 文化, 科学技術と未来). 人工知能学会全国大会論文集, 26, 1-5.

McCarthy, M. J. (2004). *Touchstone: From Corpus to Course Book*. Cambridge: Cambridge University Press.

McCarten, J. (2007). *Teaching Vocabulary. Lessons from the Corpus. Lessons for the Classroom*. Cambridge: Cambridge University Press.

Nagy, W., & Herman, P. (1985). Incidental vs instructional approaches to increasing reading vocabulary. *Educational Perspectives*, 23, 16-21.

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, 24(2), 237-270.

Nakano, M., Hoshino, A., Takeuchi, J., Hasegawa, Y., Torii, T., Nakadai, K., & Tsujino, H. (2006, December). A robot that can engage in both task-oriented and non-task-oriented dialogues. In *Humanoid Robots, 2006 6th IEEE-RAS International Conference on* (pp. 404-411). IEEE.

Namba, K. (2012). *English-Japanese Code-Switching and Formulaic Language: A structural Approach to Bilingual Children's Interaction*. LAP Lambert Academic Publishing

Nation, I. S. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Newby, P. (2010). *Research methods for education*. New Jersey: Pearson Education.

Nishimura, M. (1995). A functional analysis of Japanese/English code-switching. *Journal of pragmatics*, 23(2), 157-181.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: Univer. of Illinois Press.

Oxford, R., & Crookall, D. (1990). Vocabulary learning: A critical analysis of techniques. *TESL Canada Journal*, 7(2), 9-30.

Parrott, W. G., & Spackman, M. P. (2000). Emotion and memory. *Handbook of emotions*, 2, 476-490.

Ptaszynski, M., Dybala, P., Shi, W., Rzepka, R., & Araki, K. (2009). A system for affect analysis of utterances in Japanese supported with web mining. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 21(2), 194-213.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Shi, G. Y., Liu, S. & Chen, P. (2013). Design and Implementation of E-Learning System for Code-Switching Bilingual Teaching. *Applied Mechanics and Materials*, 241, 2886-2890.

Shimaoka, T., Aoki, S., Matsuhata, K., & Wada, M. (Eds.). (2011). *Sunshine English Course 3*. Tokyo: Kairyudo.

Song, Y., & Andrews, S. (2009). *The L1 in L2 Learning: Teachers' Beliefs and Practices*. Muenchen: Lincom Europa.

Sridhar, S. N., & Sridhar, K. K. (1980). The syntax and psycholinguistics of bilingual code mixing. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 34(4), 407.

Stewart, I. A., & File, P. (2007). Let's chat: A conversational dialogue system for second language practice. *Computer Assisted Language Learning*, 20(2), 97-116.

Takahashi, M. (2009). Utterance Generation Method Using Web Search Results and Word n-Grams. Bachelor Dissertation. *Hokkaido University, Japan*.

Tatai, G., Csordás, A., Kiss, Á., Szaló, A., & Laufer, L. (2003, January). Happy chatbot, happy user. In *Intelligent Virtual Agents* (pp. 5-12). Springer Berlin Heidelberg.

Turing, A. Computing Machinery and Intelligence. *Mind LIX* (236): 433-460 (1950).

Uemura, T., & Ishikawa, S. (2004). JACET 8000 and Asia TEFL Vocabulary Initiative, *The Journal of Asia TEFL*, 1(1), 333-347.

Urabe, Y., Rzepka, R., & Araki, K. (2014). Emoticon Recommendation System to Richen Your Online Communication. *International Journal of Multimedia Data Engineering and Management (IJMDEM)*, 5(1), 14-33.

Velasquez, M. D. G. (1995). Sometimes Spanish, Sometimes English: Language use among rural New Mexican Chicanas. In Hall, K., Bucholtz, M. (Eds.), *Gender Articulated: Language at the Socially Constructed Self* (pp. 421-446). New York: Routledge.

Vihman, M. (1998). A developmental perspective on codeswitching: Conversations between a pair of bilingual siblings. *International Journal of Bilingualism*, 2, 45-84.

Weizenbaum, J. ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the ACM* 9 (1): 36–45 (1966).

Winch, J. (2015). *Code-Switching: A new language in Japan?* Germany: Scholars' Press

Wilkins, D. A. (1972). *Linguistics in language teaching*. London: Edward Arnold.

Yen, M. J. (2009). *Code-mixing among Hong Kong Trilingual Teenagers*. Unpublished research paper: The Chinese University of Hong Kong.

Zirker, K. A. H. (2007). *Intrasentential vs. Intersentential Code Switching in Early and Late Bilinguals*. Unpublished master thesis, Birmingham Young University, Birmingham.

Zhang, L. J., & Anual, S. B. (2008). The Role of Vocabulary in Reading Comprehension: The Case of Secondary School Students Learning English in Singapore. *RELC Journal*, 39(1), 51-76.

Appendix A. Selection of participants’ comments on performance of both systems

Examples showing participants’ comments after using the proposed (System A) and baseline (System B) systems:

“Mixing Japanese and English is easy to understand and interesting, so I enjoyed using system A”

“I like both but I feel that System A helps me to gradually learn without too much material at once.

“System B is for someone more confident in his/her language ability”

“I would like this system to also have a listening exercise, which would make it even better.”

“I had a great time learning with System A. I learned and could remember many new words! I’d like to use it more!”

“Mixing Japanese and English is easy to understand and interesting, so I enjoyed using System A.”

“I like both but I feel that System A helps me to gradually learn without too much material at once. “

“Presenting new vocabulary inside Japanese readings was interesting and allowed me to guess their meanings quite well. On the contrary, the texts in System B were

entirely in English and I often felt lost and even though the exercise was enjoyable, I don't feel like I have learned many new words."

"I had a great time learning with System A. I learned and could remember many new words! I'd like to use it more!"

"System A is better for beginners, and B for people with advanced level of English. A is better for high school students, while B would be better for university level students."

"It was unique to see English mixed with Japanese language in that specific way. It also made learning new vocabulary easier for me. I think that both systems are interesting and they work pretty well!"

"I liked the idea of repetition to learn the words which were previously mistaken."

"System A is better for studying vocabulary and System B for studying grammar."

"I think it is a very simple and easy way to learn new words. I always learn collocations from sentences. Learning from context helps much more than just making a list of words to remember. I can know the right way to use word in a sentence and I don't like just memorizing words from lists."

"I liked both systems, but A seemed more effective."

"It is hard for me to learn vocabulary, because I don't know how to do it well. But, while using System A I could see English vocabulary inside Japanese sentences and understand its meaning. I think it is very good!"

"If study texts are entirely in English, probably many people will stop reading in the middle. I think that placing new vocabulary inside Japanese texts is a good way to increase the knowledge of words. I just wish there were even more words available."

The test was fun, but I would like to see even more exercises, because the mastery of English spelling is a very difficult problem to many Japanese. I also think that schoolgirls would appreciate a cuter interface with some nice looking characters!”

"I think B is good for reading short sentences and extending writing proficiency in English. In the case of System A, I was able to read it easily and it was a nice and unusual feeling that I can learn English while reading something in Japanese (I think many Japanese often have problems with reading quickly in their own language). Both systems are interesting and I like that they can check mistakes and show the proper usage on the fly. As a learning system, both are enjoyable to use.”

Appendix B. Fourteen Japanese bipolar word pairs used in our survey

“魅力がある—魅力がない

価値ある—価値のない

適切な—不適切な

満足だ—不満足だ

早く学べる—学ぶに時間がかかる

直間的だ—直間的ではない

簡単—難しい、好き—嫌い

安心—不安、興味深い—つまらない

面白い—面白くない

やる気がわく—やる気がわかない

教育的である—教育的ではない

使いやすい—使いづらい”

Appendix C. Japanese language questionnaire

英語教育システム比較実験

北海道大学大学院

情報科学研究科博士3年

マズル・ミハウ

本日は、実験に参加していただき誠にありがとうございます。本実験で得られる結果は研究以外の目的で一切使用することはないのでご了承ください。

本実験では、2つの英会話教育システムを使用いただき、どのシステムが最も使いやすく優れているのかを評価していただきます。英語教育システム「Co-MIX」とは、革新的な方法を使用して、ユーザーの新しい英語の語彙を教えるシステムです。

これからこの紙の指示に従って実験を進めてください。

質問がございましたら近くに待機しておりますので随時質問ください

それでは、次のページに移って質問に答えてください。

「英語教育システムA」ご利用ください。終わりましたら、評価用紙に移って質問に答えてください。以下に14個の対語（つまらない⇔楽しい、等）があります。先ほど使用したシステムに対しあなたの印象を-3～+3内で印（○・×・✓等）をつけて評価してください。

	-3	-2	-1	0	+1	+2	+3	
魅力がない	_____	_____	_____	_____	_____	_____	_____	魅力がある
価値のない	_____	_____	_____	_____	_____	_____	_____	価値ある
不適切な	_____	_____	_____	_____	_____	_____	_____	適切な
不満足だ	_____	_____	_____	_____	_____	_____	_____	満足だ
学ぶに時間がかかる	_____	_____	_____	_____	_____	_____	_____	早く学べる
直感的ではない	_____	_____	_____	_____	_____	_____	_____	直感的
難しい	_____	_____	_____	_____	_____	_____	_____	簡単
嫌い	_____	_____	_____	_____	_____	_____	_____	好き
不安	_____	_____	_____	_____	_____	_____	_____	安心
つまらない	_____	_____	_____	_____	_____	_____	_____	興味深い
面白くない	_____	_____	_____	_____	_____	_____	_____	面白い
やる気がわからない	_____	_____	_____	_____	_____	_____	_____	やる気がわく
教育的ではない	_____	_____	_____	_____	_____	_____	_____	教育的である
使いづらい	_____	_____	_____	_____	_____	_____	_____	使いやすい

評価し終わったら次のページに進んでください。

「英語教育システムB」ご利用ください。終わりましたら、評価用紙に移って質問に答えてください。以下に14個の対語（つまらない⇔楽しい、等）があります。先ほど使用したシステムに対しあなたの印象を -3～+3 内で印（○・×・✓等）をつけて評価してください。

	-3	-2	-1	0	+1	+2	+3	
魅力がない								魅力がある
価値のない								価値ある
不適切な								適切な
不満足だ								満足だ
学ぶに時間がかかる								早く学べる
直感的ではない								直感的
難しい								簡単
嫌い								好き
不安								安心
つまらない								興味深い
面白くない								面白い
やる気がわかない								やる気がわく
教育的ではない								教育的である
使いづらい								使いやすい

【アンケート】

1.あなたの性別に丸をつけてください。【 男 ・ 女 】

2.あなたの年齢を記入してください。【 歳】

3.あなたの所属(中学校・高校・大学・その他)を記入してください。

4.あなたはいつから英語を勉強していますか。【 】

(小学校・中学校・高校・大学・その他)

5.【卒業生の場合】卒業してからどのくらい英語を勉強していますか。

(2年間以下・2－5年間・5年間以上)

最後に次の質問にお答えください。

先ほど使用していただいた2つのシステムに対し、次の条件で順位をつけてください。

(A ・ B) で回答してください。

・ どのシステムが優れているか

1 位 2 位

・ 実在するとしたらどのシステムを使いたいか

1 位 2 位

・ どのシステムが好きか

1 位 2 位

・ 意見スペース

以上で実験は終了です。

ご協力いただき誠にありがとうございました。

Appendix D. English language questionnaire

Japanese Vocabulary Learning System Experiment Questionnaire

Michal Mazur
Ph.D. Candidate, 3rd Year
Graduate School Graduate School of Information Science and Technology
Sapporo, Japan

Thank you for participating in this experiment. Please note that the results obtained in this experiment will not be used for any purposes other than research.

In this experiment, you will test a system which is using innovative ways to teach Japanese vocabulary. I would like to ask you to interact with the application and afterwards evaluate your user experience with "CO-MIX English" system.

Your opinions count and this questionnaire will help us to understand the user's response to our research.

Please complete the experiment in accordance with the instructions in this paper and following the instructions given by an application. If you have any questions, please feel free to ask.

Now, please turn the page.

Please answer the following questions:

1. Please select your gender:

- a) male
- b) female

2. Your age:

- a) under 10
- b) between 10-19
- c) between 20-29
- d) over 30

3. Your affiliation:

- a) junior high school
- b) high school
- c) university
- d) other; please specify:
(e.g. graduate, employed, etc.)

4. How long have you been studying Japanese?

- a) less than 3 months
- b) 3 - 6 months
- c) 6 months to 1 year
- d) 1 - 2 years
- e) 2 - 5 years
- f) 5 - 10 years
- f) other; please specify:
(e.g. brought up in bi/multilingual environment, etc.)

5. Did you take the JLPT (Japanese Language Proficiency Test) before?

- a) yes
- b) no

6. [if you answered “yes” in the previous question]: Which level of JLPT did you take?

- a) N5
- b) N4
- c) N3
- d) N2
- e) N1

7. Are you planning to take the JLPT (Japanese Language Proficiency Test) Exam?

- a) yes
- b) no
- c) not sure

8. [if you answered “yes” in the previous question] Which level of JLPT exam would you like to take?

- a) N5
- b) N4
- c) N3
- d) N2
- e) N1

9. The space below (feel free to use the back as well) is for you. Tell us anything else you would like us to know about your impressions after trying the system:

Please complete the list of 14 opposite word pairs below with your impressions of the system. Please use one of the following symbols ○ · ✕ · ✓ to indicate the impression of a concept on a scale from -3(weakest) to +3(strongest).

	-3	-2	-1	0	+1	+2	+3	
unattractive	_____ : _____ : _____ : _____ : _____ : _____ : _____						attractive	
not valuable	_____ : _____ : _____ : _____ : _____ : _____ : _____						valuable	
irrelevant	_____ : _____ : _____ : _____ : _____ : _____ : _____						relevant	
unsatisfying	_____ : _____ : _____ : _____ : _____ : _____ : _____						satisfying	
slow	_____ : _____ : _____ : _____ : _____ : _____ : _____						fast	
rigid	_____ : _____ : _____ : _____ : _____ : _____ : _____						intuitive	
difficult	_____ : _____ : _____ : _____ : _____ : _____ : _____						easy	
dislike	_____ : _____ : _____ : _____ : _____ : _____ : _____						like	
stressful	_____ : _____ : _____ : _____ : _____ : _____ : _____						engaging	
boring	_____ : _____ : _____ : _____ : _____ : _____ : _____						interesting	
dull	_____ : _____ : _____ : _____ : _____ : _____ : _____						fun	
confusing	_____ : _____ : _____ : _____ : _____ : _____ : _____						stimulating	
not educational	_____ : _____ : _____ : _____ : _____ : _____ : _____						educational	
difficult to use	_____ : _____ : _____ : _____ : _____ : _____ : _____						easy to use	

Research achievements

1. Academic Journals

- 1) Michal Mazur, Krzysztof Karolczak, Rafal Rzepka and Kenji Araki. A System for English Vocabulary Acquisition Based on Code-Switching. In: *International Journal of Distance Education Technologies (IJDET)*, (in-press).
- 2) Michal Ptaszynski, Michal Mazur, Pawel Dybala, Rafal Rzepka, Kenji Araki and Yoshio Momouchi. (2013). Towards Computational Fronesis: Verifying Contextual Appropriateness of Emotions. In *International Journal of Distance Education Technologies (IJDET), Special Issue on Emotional Intelligence for Online Learning*, Vol. 11, No. 2, pp. 16-47.

2. International Conferences

- 1) Michal Mazur, Rafal Rzepka and Kenji Araki. (2010). Co-MIX Project: Towards Artificial Tutors Using Code Mixing as Foreign Language Teaching Method. In *Proceedings of the International Workshop on Modern Science and Technology (IWMST 2010)*, pp.196-201.
- 2) Michal Mazur, Rafal Rzepka and Kenji Araki. (2011). Proposal for a Conversational English Tutoring System that Encourages User Engagement. In *Proceedings of the 19th International Conference on Computers in Education*. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education (ICCE2011), pp.10-12.

- 3) Michał Mazur, Rafał Rzepka and Kenji Araki. (2012). Chatterbots with Occupation - Between Non Task and Task Oriented Conversational Agents. In *Proceedings of Linguistic and Cognitive Approaches to Dialogue Agents (LaCATODA 2012)*, 61, pp. 61-66.
- 4) Michał Mazur, Rafał Rzepka, and Kenji Araki. (2015). Applying Code-Switching Method for E-Learning in English and Japanese Vocabulary Acquisition Systems. In *Proceedings of 7th Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pp. 487-491.

3. Domestic Conferences

- 1) Michał Mazur, Kohichi Sayama. (2009). Towards Artificial Tutors – Robots, as Teachers and Partners for Children. In: *Language Acquisition and Understanding Research Group (LAU) Technical Reports*, Winter 2009, pp. 11-18.
- 2) Michał Mazur, Rafał Rzepka & Kenji Araki. (2011). The Idea of English Tutoring System Based on Code Mixing and Emotion Recognition. In *Proceedings of 「第4回若手研究者支援のための産学協同GCOE国内シンポジウム」*, pp.159-160.
- 3) Michał Mazur, Rafał Rzepka & Kenji Araki. (2012). The Idea of a Dialogue System That Encourages User Engagement. In *Proceedings of the 5th International Symposium on Global COE Program of Center for Next-Generation Information Technology Based on Knowledge Discovery and Knowledge Federation (GCOE-NGIT 2012)*, Hokkaido University, Sapporo, pp. 277-278.

- 4) Michal Mazur, Rafal Rzepka and Kenji Araki. (2012). Mixing Words and Emotions-New Natural Methods for Artificial Language Tutors. In *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*.
- 5) Michal Mazur, Rafal Rzepka and Kenji Araki. (2012). Chatterbots with Occupation-Between Non Task and Task Oriented Conversational Agents. *Proceedings of Linguistic and Cognitive Approaches to Dialogue Agents (LaCATODA 2012)*, 61, pp. 61-66.
- 6) Rafal Rzepka, Michal Mazur, Austin Clapp and Kenji Araki. Global Brain That Makes You Think Twice. To appear in *Proceedings of AAAI Spring Symposium*, Stanford, March 2016.

Acknowledgements

I would like to sincerely thank my supervisor, Prof. Kenji Araki, from the Graduate School of Information Science and Technology, Hokkaido University, for the invaluable guidance in writing this thesis and all assistance he has provided during my doctoral course.

I am also extremely grateful to Professor Tsuyoshi Yamamoto, Professor Miki Haseyama and Professor Yuji Sakamoto who acted as examiners during the defense of this thesis, and wish to extend this thanks to all the professors in the Media Networks Division of the Graduate School of Information Science and Technology.

I would also like to sincerely thank Assistant Professor Dr. Rafal Rzepka, from the Graduate School of Information Science and Technology, Hokkaido University, for the countless hours of assistance and fruitful discussion over the course of performing the work described in this thesis.

Furthermore, I would like to sincerely thank everyone at the Language Media Laboratory, Graduate School of Information Science and Technology, Hokkaido University for their invaluable support and assistance.

I would also like to thank all my fellow students, my family and other people who helped me and supported in various ways.