THE EFFECTS OF COMPUTER-ASSISTED TEXT PRODUCTION IN EARLY JAPANESE LEARNING
パソコン使用による日本語学習者の漢字タイピングエラーの分析

Michael Dixon
Indiana University Bloomington

Introduction

The purpose of this study is to examine how students of Japanese as a foreign language (JFL) utilize computer-based character input in their development as writers in the foreign script. Although there are various differences between the processes involved in producing handwritten and typed script, this study seeks to determine how students use this technology in their development as JFL writers, and to determine what the potential benefits or downfalls of a more computer-based approach to text production could be for learners of Japanese.

Due to the unique nature of the Japanese writing system, students in JFL classes typically spend a greater amount of time dealing with issues of the new orthography than one might find in the curricula of other non-character-based languages. This time and focus on the orthography in teaching JFL presents a particular challenge to educators because they are then required to tend to issues related to the unfamiliar orthography in addition to the normal tasks of teaching grammar, vocabulary, listening and speaking. The struggle to attend to each issue is not just seen in the instructors but also in students of Japanese, as there tends to be a high rate of attrition at the early stages of language learning. The introduction of kanji presents further potential difficulty for learners who come from non-character-based languages because they are required to memorize and reproduce kanji, invoking the need for new strategies to aid in visual memory and kinetic recall of the characters in their written forms.

There has been a significant amount of debate among JFL educators as to when orthography should be introduced, but research in this area has focused primarily on when to introduce the native system of writing (Hatasa, 2002), and has not delved into different ways of having students interact with text, such as a computer-based writing approach. Because this particular aspect of the Japanese pedagogical process has not been experimentally reviewed, this study seeks to quantitatively examine the difference between the hand-written and typed domains for JFL learners. This study seeks to determine the differences between these two types of writing production by examining the text produced by a groups of students who will produce text output by the traditional method of writing by hand as well as being allowed to complete certain class assignments with the aid of a computer. This group was given supplementary instruction on how to use a computer to type in Japanese, and this instruction was put into place by requiring students to use a computer to type essay assignments throughout the semester.

In addition to examining student affect regarding the task of writing by hand vs. writing by computer, the written output of students will also be tested through error analysis to determine if the cognitive processes in learning and using kanji are similar or different from those of native Japanese speakers (phonological-type errors, as observed by Horodeck, 1987) and to determine what types of errors occur in handwritten vs. typed

production, as to determine what challenges students face in writing by hand and with a computer.

With regard to the L2 production of text in *romaji* and any of the Japanese scripts, it is important to consider the sequence of processing that takes place in between thought and production. Consider the following flow of processing in an L2 learner of Japanese writing in the hiragana syllabary by hand: The student has the word in mind, parses the words into its respective mora, and then recalls the visual form and kinetic association necessary to write the syllable in the foreign script. If there is kanji involved, the student must recall the correct kanji and its corresponding compounds, recall the visual and kinetic association to the character, and then be able to write the character with the correct stroke order. In the case of writing in *romaji*, this process is much more native for JFL students of alphabetic backgrounds: the word in mind is written with the internalized script practices of the L2 learner. If this is being done by typing, the writer is presented with a set of kanji candidates that correspond with the *romaji* input and choose from the set of homophones. The JFL student is able to bypass the arduous step of having to memorize how to write these characters in a script that is very different, kinetically and visually complicated, and therefore unfamiliar to the learner.

One potential benefit of the use of a word processor is that it takes advantage of the fact that orthographic complexity does not necessarily result in difficulty in processing kanji. In fact, the opposite may be true to a certain extent in terms of how it applies to the English writing system, and some studies have shown that characters that are more orthographically complex are in fact more easily remembered than orthographically simple characters (Steinberg & Yamada, 1978-1979; Kess & Miyamoto, 1999). One of the most important factors in kanji recognition is the frequency in which the kanji appears in everyday reading and therefore contributes to visual familiarity. High frequency kanji are therefore more easily recognized because they are more visually familiar to the reader. Even in the case of kanji that share similar frequencies, Kawai (1966) found that more complex characters are in many cases easier to read and easier to remember than less complex characters.

Phonological and Semantic Aspects of Word Processing

Although there is a significant amount of research which suggests that Japanese (like all other developed writing systems [DeFrancis, 1989]) are essentially phonologically-accessed writing systems, their systems of typing are of interest to reading researchers because although the input method heavily relies on the phonetic encoding of the words to be typed (in *romaji*, which is a traditionally less-familiar orthography for Japanese writers), the character selection process then calls the semantic value of these characters into play once again when the writer chooses the correct character or combination of characters from among its homophones. This heightened importance on the differentiation of common phonetic characters could be seen as placing a newer, stronger (or at least different) emphasis on the semantic value of the characters (an emphasis on the comparative semantics of *all* of the related characters instead of just a kinetic reaction to the word unit as a whole), because the writer is required to know what the correct compound *is* as well as what it *phonetically could be but is not*.

It could be argued, however, that for non-native learners of Japanese electronic text input could have negative benefits in terms of the depth of kanji knowledge. According to the Orthographic Depth Hypothesis, the L1 orthography affects L2 word recognition processes. While Japanese is considered to be a deep orthography, speakers of languages with roman alphabets would be considered to come from L1s with more shallow orthography (Koda 1998. 1990). It would follow that native speakers of languages with Romanized alphabets would approach the Japanese writing system with certain biases acquired from their L1 writing system. It could also be argued that the depth of Japanese orthography is in fact greatly lessened by its transfer to electronic text input, as it is essentially reduced to a phonetic coding with an added measure of semantic confirmation. The difference in orthographic depth between the written and typed domains in Japanese writing have yet to be formally explored, and it is the intention of this study to present findings which can be applied to the Orthographic Depth Hypothesis and how it applies to typed Japanese.

Implications of Computer-Mediated Text Input on Kanji Use, Cognition

The development of the word processor leads to interesting new views of text in Japanese, because it both emphasizes the role of *romaji* as a tool for input in many electronic devices, but it also cements the wider use of kanji because of its easier access (Gottlieb, 1995). The nature of the word processing task shifts the writing task from production (which has been proven to be mostly kinetic in the case of writing by hand [Watanabe, 1991]) to one of recognition. Despite the documented difficulty that Japanese readers have in detecting phonologically acceptable errors in kanji form (Horodeck, 1987; Matsunaga, 1994), Ishii (2000) states that if the kanji is phonetically entered without errors, that NS writers are very likely to correctly access any kanji they can recognize, and are able to do so at a much greater rate than they would be able to if they were producing the characters by hand. This facilitated access to the kanji corpus has been cited as creating a change in the number of kanji being used as well, because the ability to recall the kanji from memory is no longer a requirement for its use (Tanaka, cited in Gottlieb, 2000; Taylor and Taylor, 1995). In terms of production, the focus has shifted from an emphasis on the visual (or as some would argue, kinetic) makeup of the character to the ability to recognize and utilize the correct kanji in context (Takase, 2001). Although the production of kanji is the primary focus of this study, it is worth noting that this is not the only part of the Japanese language that is subject to constraint-based input on a computer. The production of okurigana is also subject to the effects of a constraint-based system similar to the ways that kanji is, facilitating through typing what is an arbitrary and ambiguous task by hand.

Implications for JFL Pedagogy

The prevalence of computer-mediated text input in Japanese is certainly not likely to lessen in the future, emphasizing the importance of drawing meaning from the changes that this shift in the notion of literacy brings about. For JFL learning, the shift in the skills associated with computer-mediated text input would indicate potential for benefit for JFL learners from alphabet-based languages in the following ways:

- The ability to use a keyboard they are already familiar with.
- The ability to enter the text through a regular phonetic system.
- The ability to produce text without having to memorize the orthography of the characters they are intending to write.
- The ability to have greater exposure to the written system as a whole.
- The ability for more incidental vocabulary and kanji learning to take place as the result of increased exposure to authentic text.
- The ability to have increased exposure to the semantics of kanji through the selection process.
- The ability to spend more time on other areas of language development without spending too much time on writing practice.
- The ability to have a greater sense of the writing system as a whole before embarking on the writing task.

As this review of literature has shown, the complexity of the Japanese writing system is manifest in a variety of seemingly paradoxical phenomena, such as the higher recognizability of graphically complex characters, the reliance on kinetic memory over visual memory in producing characters by hand, the relative ignorance of the semantic meaning of kanji during fluent reading, and the reliance on phonological access when the characters themselves have relatively little graphical-to-phonological correspondence. The seemingly "deep" orthography of kanji has been shown to be treated like a shallow orthography in fluent reading, and the Japanese system of computer-mediated text input requires decoding of the phonological manifestation of a lexical item into a script in which a single keystroke represents a sub-morphemic component. In addition to these phenomena, the asymmetry in reading and writing ability (at least in the hand-written domain) would indicate that the Japanese system of computer-mediated text input could help to level this asymmetry through its increased emphasis on the reading of characters while writing, and that this system may in fact actually increase awareness of the semantic nature of the characters through the selection of the correct homophone during the kanji selection process. Based on the aforementioned evidence, it would be useful to investigate the effects of these potential benefits on JFL learners, as this research could be of benefit to the field of literacy in Japanese as well as the field of JFL pedagogy. It is the intent of the next section to show how these effects could be effectively assessed and implemented in the classroom.

Rationale

Computer-mediated text input has become the predominant form of text production in Japanese (Gottlieb, 2000:26), and the fast-paced emergence and evolution of this system have not afforded very many careful examinations of how this change has affected notions of literacy in the Japanese language, let alone to look at how this change in literacy might affect non-native speakers (NNS) of Japanese who might employ this system of text. The shift in technological literacy to the skills associated with computer-mediated text input hint at potential for benefit for JFL learners from alphabet-based languages in the ways described previously.

Although there have been studies that have looked at the cognitive processing of kanji as it is read by native speakers of Japanese (Horodeck, 1987; Matsunaga, 1994; Law and Caramazza, 1995; Hatta and Kawakami, 1996; Mori, 1998), there are not very many studies that have either looked at how this cognition affects the reading NNS of Japanese, or have looked at how this affects the writing of either NS or NNS speakers of Japanese. Horodeck's (1987) study remains one of the few to examine the phenomenon of writing errors and accuracy, but his study utilized native speakers of Japanese and focused specifically on using hand-written writing errors to determine that the reading of kanji is an essentially phonological construct in the minds of Japanese readers. Because of this, his methodology is primarily focused on the categorization of errors based on his research question, although for the purposes of this inquiry, this classification can also be used to compare the errors in kanji output in the handwritten vs. typed domains in Japanese.

Emergent Themes in the Literature

In looking at how computer-mediated text input might affect the processing and cognition of the kanji writing system, there are several themes which emerged as being particularly pertinent to the study of the computer-mediated text input phenomenon and that provide clues as to how this phenomenon can be scientifically analyzed.

The first and most commonly cited theme in the literature on kanji processing is that kanji is primarily phonologically accessed by native speakers in fluent reading (Tzeng, Lin, Hung, & Lee, 1995; Horodeck, 1987; Matsunanga, 1994). This finding provides a valuable connection to computer-mediated text production in Japanese because this method relies on phonologically coding the words to be written into *romaji* which are then converted into kanji through the computer's corpus of kanji that match a given phonological "spelling." The phonological bias in reading may be seen as a contributor as to why the Japanese computer-mediated text input system is also set up to require phonological input, although it could be said that phonological input simply happens to represent the most concise and accessible method currently available for getting text from the mind of the writer through to its appearance on the page (Gottlieb, 2000). Regardless of the impetus for this development, the first step in kanji production on a computer is to decode the intended phonology into its *romaji* equivalent.

The second theme in the literature on kanji processing and cognition that is pertinent to this research is found in the data stating that native writers are less prone to make phonological errors (e.g., errors that are phonologically acceptable but otherwise incorrect) than semantic or purely orthographic errors in writing (Tzeng, Lin, Hung, & Lee, 1995; Horodeck, 1987; Matsunanga, 1994). This finding, in addition to being an important clue as to how we know that kanji is primarily phonologically accessed by native speakers in fluent reading, also provides a very important connection to the computer-based text input system because this system could be seen as enhancing the probability of the occurrence of these errors by requiring the user to choose between a set of phonologically similar kanji with different meanings, in a sense creating a "trap" for "lazy" NNS *and* NS writers who might rely on the kanji conversion process to make up for their lack of knowledge as to what the right kanji should look like for a given phonology.

The third theme in the literature dealing with the implications of a new literacy that interacts frequently with computer-mediated text input is that the constraint-based corpus-access processes of computer-mediated text input in Japanese facilitates the correct production of the intended text, in that it allows users to produce kanji that they might not be able to otherwise write by hand (Ishii, 2000, Gottlieb, 2000). The implications of this range from pro to con, since on the one hand, computer-mediated text input could be seen as an empowering tool for JFL learners, but on the other hand, the reliance on computers could be seen as contributing to attrition in the hand-written skills of NS Japanese. Regardless of the stance on this issue, what is missing from the literature is quantitative data on how much of a difference there is between written and typed ability, so it is the goal of this research to find out what the differences in hand-written vs. typed text production are both in terms of accuracy and the variety of kanji that are used.

The fourth theme in the literature is that the cognitive processes of native writers are very different than those of NNS (Naucler, 1983:596), and the same applies to beginning readers and skilled adults (Tzeng, 1980). Because of this, it is important to determine if there are actual parallels that can be drawn between beginning readers and skilled adults and NNS and NS in terms of their approaches to writing, or more simply, do people learning Japanese as a foreign language approach the task of learning kanji in the same way that Japanese children do? Studies such as (Mori, 1998) have attempted to determine if strategies and biases in the L1 are transferred to L2 reading in Japanese, and although this study argues that speakers of English speakers look for phonetic representation in the Japanese writing system, it ultimately only stands to show that visual familiarity with the forms of kanji helps in kanji identification. Research on child learning of Japanese kanji has shown that there is a bias towards the centrality of meaningfulness in text (Steinberg & Yamada, 1978-1979) which suggests that children look for meaning first in text before pronunciation, and similar research by Rozin, Poritsky, and Sotsky (1971) showed that English speaking children with reading disabilities also benefited from learning Chinese characters as logographs, also indicating a bias towards the centrality of meaningfulness in text at early reading. These findings for early reading are of course different from the findings in adult reading, as studies have shown that phonological representation is the primary means of access in the fluent reading of a variety of languages (Tzeng, Lin, Hung, & Lee, 1995; Horodeck, 1987; Matsunanga, 1994). There is, however, no definitive study which describes the cognition of kanji in L2 early reading, therefore the body of research in this field would benefit from a scientific examination of this phenomenon.

Based on the themes cited previously, there are certain questions that emerge when comparing the research that has been done on reading and writing in Japanese and comparing those findings to the possible implications of the use of computer-mediated text input. In using computer-mediated text input, there are a variety of differences between this method of text production and the traditional method of writing by hand. Based on the previously mentioned theoretical premises, this study will attempt to determine how the differences between hand-written and typed production manifest themselves, and how these differences affect JFL learners at the early stages of language learning. In looking at these differences, one might benefit from a general working model of how these processes differ in the written and typed domains.

It is perhaps useful to look at the processes of the kanji production act in the written domain to roughly look like this:

- Word/Concept Formation (deciding what needs to be written)
- Kanji Lexicon (deciding how that should be written in kanji)
- Motor Act (producing the kanji from kinetic memory as a chunk)

In typing, the order might look something more like this:

- Word/Concept formation (deciding what needs to be written)
- Phoneticization (decoding the intended words into *romaji*)
- Motor act (producing the kana via *romaji* [as components] on the keyboard)
- Kanji lexicon (choosing the right kanji from its homophones)

The major differences between these two models are that after the word/concept formation, phonetic decoding takes places in the typed domain, followed by the motor act of typing, *after* which the kanji lexicon is accessed. This difference would naturally lead to changes in the way that the writing is processed and produced.

Based on the aforementioned conditions that effect the processing of kanji during the reading and writing process, in order to find out more about how the processing of hand-written text production differs from the processing involved in typed text production, these two conditions need to be manipulated in order to reveal if performance and student affect differ in these two domains.

Hypotheses:
1. NNS of Japanese will be able to accurately produce above their handwritten ability by means of computer-mediated text input
2. Phonological bias for NNS will be found in computer-mediated text input as it has been found in the hand-written domain (errors will be predominantly of the right-sound/wrong-form/wrong-meaning type)
3. As the semantic nature of kanji is called into question in the kanji selection process during computer-mediated text input, computer users will tend to use meaning-based strategies for dealing with unfamiliar characters when using the word processor, resulting in fewer semantic errors (errors of the right-sound/wrong-form/wrong-meaning type) NNS
4. The use of the English keyboard *romaji* will provide affective advantages for English-speaking JFL students

Research questions:
1. What are the differences in the types of errors that JFL writers make based on their hand-written and typed output?
   a. Are these errors primarily phonological, semantic, or orthographic?

2.   Do JFL learners find more time to focus on other aspects of language learning when the task of writing is facilitated through computer-mediated text production?

3.   How much time is spent typing vs. writing by hand for a given section of text?

4.   Do JFL writers use more kanji when they access it through computer-mediated text input?

5.   Do NNS of Japanese process kanji with a disposition towards phonetic association (similar to NS) when writing?

   a.   Does this disposition change based on the typed or hand-written domain?

6.   Do JFL learners show more accuracy with typed input over hand-written input?

For research question 1, this question will be assessed through testing JFL learners on hand-written and typed writing tasks, and the results of these tasks will be analyzed based on the error analysis criteria established by Horodeck (1987) and further refined by Hatta, Kawamura, and Tamaoka (1998). These rates of accuracy will be assessed generally (i.e., in terms of overall accuracy in percentage), and Question 1.1 will be addressed based on the classification criteria outlined in Horodeck (1987) and Hatta, Kawamura, and Tamaoka (1998).

Question 4 will be answered based on a comparison of their hand-written samples and their typed samples to determine if kanji is more frequently used (if there is a higher degree of kanji density [% ratio of kana to kanji]), and if a wider variety of more "difficult" kanji (kanji of a higher grade level, kanji that is rated as being less-frequently used, as based on the statistical tendency for kanji to be used in that situation to express that word, etc.) is being used as a result of the facilitating effects of computer-mediated text input.

Question 5 would then be determined based on the overall accuracy based on the analysis criteria outlined by Horodeck (1987) and further refined by Hatta, Kawamura, and Tamaoka (1998), which would be compared based on the typed and hand-written domains.

Methodology:

Essay Collection

The above hypotheses and research questions are examined through several steps.

The first step is the collection of typed essay assignments in Japanese on a given topic (for example, writing about a time when you were made to do something by your parents as a child) and then comparing the production of these texts with their corresponding production on the essay section of a chapter test in which the same writing topic is assigned to the students. The amount of kanji used, accuracy, and kanji grade level are analyzed to catch emergent trends in the handwritten and typed domains.

Dictation Task:

The task will ask participants to listen to sentences containing vocabulary that has been

covered previously in their study of the language. They will then be asked to write these sentences in Japanese, first by hand, then using a computer, using kanji where appropriate. Certain sentences that will come up later in the task will introduce vocabulary that the participant would most likely not be familiar with, to determine if the participants would "put faith" in their own handwriting ability or computer-based text input to write words and phrases in kanji even if they are not familiar with these words, or to see if using a computer-based writing system allows the participants to "accidentally" produce the correct characters without actually knowing their meanings. By having the students transcribe words that they might not know how to write, the task could also explore the students' ability to use semantic clues to make inferences about what the right kanji combination might be in a given situation. These transcriptions would present novel words that participants are likely not to know that are made up of kanji that they should know, with sentences that hint at what the kanji should be (nevertheless, these kanji should be phonologically ambiguous on a word processor).

Assessment of the accuracy and usage of kanji (handwritten/free & dictated) will take place based on general graphical correctness, with additional assessment of native speakers as to whether kanji should be used in that situation as opposed to kana (for example, take the phrase 有り難うございます - a phrase most commonly written only in hiragana, although a computer would give the writer the option of writing it in its archaic, stilted kanji form - this would be considered pragmatically incorrect by most native readers of Japanese, resulting in an overuse error in the analysis).

Think-aloud and review:

Participants are asked to use a think-aloud protocol during their completion of the dictation task, and are then asked compare their performance between the handwritten and typed tasks and to freely comment on them. Discrepancies in the two domains will be pointed out to the participant for comment.

Interview:

Participants are interviewed on their feelings about kanji study based on their reactions to the dictation task.

Questionnaire:

Participants are given a 60-question questionaire regarding their attitudes, strategies, and approaches towards kanji study. This questionaire is administered via computer using the school's online course management system.

Sample:

32 students of second-year Japanese from a large midwestern university participated in every part of this study. Participants were all native speakers of English who were currently enrolled in their fourth semester of college Japanese. The age of participants ranged from 19 to 29, and included 14 females and 18 males.

Results:

Data collection for this research continued until May 7th of 2010 (coincidentally the due date for the submission of this paper for the PJPF proceedings), therefore detailed analysis of the results of this data would be premature at this time, although this analysis will be made public as part of the completion of my doctoral thesis in language education at Indiana University, Bloomington. Please feel free to contact me at mhdixon@indiana.edu if you would be interested in knowing more about the results of this research.

References

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu: University of Hawaii Press.

Flaherty, M. (1991). "Do second-language learners of Japanese process kanji in the same way as Japanese children?" Sekai no Nihongokyooiku. 1, March, 183-200.

Gottlieb, N. (2000). Waapuro no shakaigengo-gaku (in Japanese) (Sociolinguistics in the Word processor). *Nihogogaku*. 19 (12), pp.52-57.

Hatta, T., Kawakami, A., Tamaoka, K.(1998) "Writing Errors in Japanese Kanji: A Study with Japanese Students and Foreign Learners of Japanese." Reading and Writing: An Interdisciplinary Journal, v10 n3-5 p457-70.

Horodeck, R. (1987). *The role of sound in reading and writing kanji*. Ph.D. thesis, Cornell University, Ithaca, NY.

Ishii, K. (2000). Computer to Kanji Kyooiku (in Japanese) (Computer and Kanji Education). *Nihongogaku*. 19 (9), pp.40-48.

Kess, J., & Miyamoto, T. (1999). *The Japanese mental lexicon: Psycholinguistic studies of kana and kanji porcessing*.Amsterdam: Benjamins.

Law, S.-P., & Caramazza, A. (1995). "Cognitive processes in writing Chinese characters: Basic issues and some preliminary data." in de Gelder, B. & Morais, J. (Eds.). *Speech and Reading: A Comparative Approach*. Sussex: Erlbaum.

Matsunaga, S. (1994). *The linguistic and psycholinguistic nature of kanji: Do kanji represent and trigger only meanings?* Ph.D. thesis, University of Hawaii, Honolulu, HI.

Naucler, K. (1983). "Connections between spoken and written language: Evidence from three investigations of normal and pathological speech performance." *Journal of Pragmatics* 7.595-602.

Rozin, P., S. Poritsky, & R. Sotsky. 1971. "American children with reading problems can easily learn to read Chinese characters." *Science* 171.1264-1267.

   Steinberg, D. D. & Yamada, J. (1978-1979). "Are whole word kanji easier to learn than syllable kana?" *Reading Research Quarterly* 14.88-99.

   Takase, K. (2001). Digital echnologies and literacy in Japanese http://www.ioe.ac.uk/ccs/dowling/studentswork/takase/takase_lit.htm

Taylor, I. and Taylor, M.M. (1995). *Writing and Literacy in Chinese, Korean and Japanese*. John and Benjamins Publishing Co.

Tajima, K. (1989). Kikai-jisho no kosei. In S. Mizutani (ed.), Asakura Nihongo koza 1: Moji, hyoki to go-kosei, pp. 20-76. Tokyo: Asakura Shoten.

Tzeng, A. K.-Y. (1994). *Comparative studies on the word preceptions of Chinese and English: Evidence against an orthographic-specific hypothesis*. Ph.D. thesis, University of California, Riverside.

Tzeng, O. J. L., Lin, Z. H., Hung, D. L., & Lee, W. L. (1995). Learning to be a conspirator: a tale of becoming a good Chinese reader. In B. Gelder, & J. Morais (Eds.), *Speech and reading: a comparative approach* (pp. 227–244). Hove: Erlbaum.

Watanabe, Y. (1991). "Latency in the transcription, transliteration, and pronunciation of Japanese ideographs and syllabic letters." *Tohoku Psychologica Folia* 50.25-34.