Tentative Analysis of Japanese Writing Errors Observed within Responses Collected for a Word Association Survey[1]

Terry Joyce

Abstract. As a source of insights into the organization of orthographic representations within the Japanese mental lexicon, this paper presents a tentative analysis of Japanese writing errors observed within responses collected for a word association survey conducted as part of the ongoing construction of the large-scale Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007a; Joyce & Miyake, 2008). In contrast to the 'informal' or 'spontaneous' writing situations from which previous studies have collected writing errors (e.g., Hatta, Kawakami, & Hatasa, 1997; Hatta, Kawakami, & Tamaoka, 1998), the writing error data collected in this study would seem to be more instructive about the types of information utilized when an individual tries to retrieve and write a kanji that they are less than confident about. In total, 1,078 writing errors (924 kanji orthography writing errors and 154 kana orthography writing errors) are classified according to a newly extended classification scheme. The classification results for kanji orthography writing errors are consistent with the notion that the activation of phonological information is an important factor in the production of kanji writing errors. However, the present results also indicate that even when native Japanese writers are not confident about the correct strokes for a particular kanji character, they often have a partial visual image for the target kanji's components or have a vague visual outline for its overall form, suggesting that the activation of orthographical information is also an important factor in the production of kanji writing errors.

Keywords: Writing errors, orthographic representation, kanji, kana, word association survey, Japanese Word Association Database (JWAD)

Introduction

As Joyce (2008, in press) observes, there is a pervasive impression amongst scholars of writing systems that the Japanese writing system is excessively complex. Indeed, it has been described as being the *most complex* (Sproat, 2000; Kess, 2005; Gnanadesikan, 2009), the *most complicated* (Coulmas, 1989; Fischer, 2001; Robinson, 1995), the "hardest to force into any taxonomic mold" (Sproat, 2000, p. 132), the *most intricate* (Coulmas, 1989; Kess, 2005), the *most onerous* (Unger, 2004), as being *without inferiors* (Sansom, 1928), or simply as "one of the worst overall systems of writing ever created" (DeFrancis, 1989, p. 138). Smith (1996), however, provides a rarer appraisal of the Japanese writing system that is more balanced in nature. On the one hand, Smith highlights two aspects that undoubtedly contribute to the system's overall sense of complexity. The first is the dual system of kanji readings—with both on-readings, borrowed Sino-Japanese pronunciations, and kun-readings, Native Japanese pronunciations—while the second aspect is its multi-script nature—with kanji (literally, 'Chinese characters'), hiragana and katakana, which are two native syllabaries, and, increasingly, rōmaji (the Roman alphabet), as well as Arabic numerals being used in largely separate and complementary ways as components of an overall system. On the

other hand, Smith also suggests that there has been a tendency to overstate the sense of complexity.[2] In defense of the Japanese writing system, Smith points to the fact that it is capable of meeting the functional needs of a modern, literate society.

Smith (1996) is noteworthy for taking a more objective stance towards the Japanese writing system, but in seeking to develop a more accurate and more appropriately nuanced understanding of the system, it is probably well to also recognize that appeals to literacy levels within Japanese society are not totally without problems.[3] That said, however, while a paper that considers Japanese writing errors may have some justification for sounding a tiny cautionary note against uncritically accepting claims about literacy levels for Japanese, the present paper does not attempt to pursue the implications of its writing error data for larger issues related to Japanese literacy.[4] Rather, this paper takes a cognitive psychology approach to the writing error data. Acknowledging that handwriting is in itself, irrespective of the particular writing system employed, a cognitively highly demanding task, it seems reasonable to expect that any large-scale sample of handwriting will contain at least a few writing errors. As recognized within psycholinguistics, linguistics errors—including speech errors, writing errors and signing errors—can provide important information about the nature and organization of the mental lexicon (Carroll, 2004; Fromkin, 1971, 1973; Harley, 2001). In a similar vein, this paper presents a tentative analysis of the Japanese writing errors observed within responses collected for a word association survey conducted as part of the ongoing construction of the large-scale Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007a; Joyce & Miyake, 2008), as a promising approach for obtaining interesting insights into the organization of orthographic representations within the Japanese mental lexicon.

Despite the potential for illuminating various issues relating to the relative contributions of orthographic, phonological and semantic information in the lexical representation and lexical retrieval processes involved in the Japanese mental lexicon, there have been very few studies of writing errors made by normal Japanese speakers. Two notable exceptions are the closely related studies by Hatta, Kawakami, and Hatasa (1997) and Hatta, Kawakami, and Tamaoka (1998). Study 1 of Hatta, et al. (1997) analyzes 374 kanji writing errors within two-kanji compound words made by Japanese college students in informal writing situations.[5] According to Hatta, et al. (1997), these writing errors were collected over a seven-year period from 2,200 Japanese students attending five colleges,[6] although they provide no information about the actual size of the written corpus from which the errors were identified. As the context in which their writing errors were made is markedly different from the context in which the present writing error data was collected, it is important to clarify what Hatta, et al. (1997) refer to as 'informal writing situations' and Hatta, et al. (1998) describe as 'spontaneous sentence writing situations'. Hatta, et al. (1997) stress that, although the writing errors "were collected mainly from essays and letters to teachers on examination answer sheets", such "essays and letters to teachers were only written if students had ample time remaining in their 100 minute examinations" (p. 406). Hatta, et al. (1997), therefore, claim that the prevailing writing situations were free from time pressures. Hatta, et al. (1998) are somewhat more explicit in spelling out their basic assumptions when they suggest that, firstly, the students "were under no obligation to use kanji in their essays and letters to their teacher", because they could always transcribe a word in kana orthography if they were uncertain about their recall of the appropriate kanji

character, and, secondly, because "no student would knowingly include errors in an essay" but still they "voluntarily chose to write the kanji form we must conclude that they thought their versions were correct" (p. 461). If we can accept that the student participants in those studies were truly free of any time pressures when committing the writing errors, then it would be reasonable to regard them as natural 'slips of the pen' or unintended writing errors that potentially reflect the normal lexical retrieval processes involved in producing handwriting.

The classification of writing errors utilized by both Hatta, et al. (1997) and Hatta, et al. (1998) is essentially the same as that proposed by Hatta, et al. (1997).[7] It is based primarily on three kinds of basic substitutions; substitutions based on a shared reading or phonological kanji (P type), substitutions based on orthographically similar kanji (O type), substitutions based on semantically similar kanji (S type). The classification also includes combinations of the three basic types (yielding P+O, P+S, O+S, and P+O+S types), order reversals, substitutions with non-kanji (combinations of kanji strokes or components that do not actually exist as Japanese kanji), and errors that do not readily match to the other categories. Thus, in total, the classification consists of ten writing error categories.

Given that Hatta, et al. (1998) appears to provide the more coherent presentation of the classification results from the two studies, the present paper relies on their interpretations.[8] Hatta, et al. (1998) discuss first the relative frequencies of writing errors involving the three basic substitutions, or single factor error categories, noting that orthographically-related errors (O type) accounted for 17.9% of the writing errors, followed by phonologically-related errors (P type) accounting for 9.1%, and semantically-related errors (S type) at 6.0%. After commenting that the orthographically-related errors were the most frequent of the basic substitution errors, however, Hatta, et al. (1998) also suggest that the pattern of writing errors was rather different when the multiple-factor error categories were also considered. The total frequency for all the error categories involving phonologically-related errors (i.e., the P, P+S, P+O, and P+O+S types) was 60.2%, while the total for all error categories involving orthographically-related errors (i.e., the O, P+O, O+S, and P+O+S types) was 43.6%, and the total for all of the semantically-related errors (i.e., the S, P+S, O+S, and P+O+S types) was 29.7%. The results promoted Hatta, et al. (1998) to suggest that phonologically-related errors are a major factor in kanji writing errors (p. 463). It should also be noted that 15.0% of the writing errors were classified under the non-kanji category.

To the extent that we may accept Hatta and colleagues assumptions that the informal or spontaneous writing situations under which their writing error data was collected were free from time pressures and that their participants could always resort to writing a word in kana orthography if they were uncertain about their recall of the appropriate kanji character, their findings indicate that the activation of phonological information during the lexical retrieval processes involved in handwriting kanji characters may occasionally induce kanji writing errors. Drawing generally on Levelt's (1995) well-known model of speech production, we may conceive of the cognitive processes involved in handwriting production as consisting of the four stages of conceptualization, formulation, hand motor execution, and self-monitoring.[9] In the inital stage of conceptualization, the writer determines at a conceptual level the message or content that they wish to express in writing. The formulation stage may be understood as

encompassing a number of linguistic subcomponents, such a syntactic structuring and lexical selection. In the case of handwriting production in Japanese, following on from lexical selection, we may assume that there is usually also an orthography selection process, where the writer should decide on the most appropriate orthographic representation of target words; whether to write them in kanji orthography or kana orthography.[10] If the writer decides that kanji orthography would be most appropriate, their task would then be to retrieve the relevant kanji from long-term memory. This lexical retrieval process would be driven by both semantic (the target meaning) and phonological (the sound of the target word) information, but given the extremely high incidences of homophomes in Japanese, the phonological information is likely to activate many kanji orthographic candidates. The findings from Hatta and colleagues seem to indicate that native Japanese writers will occasionally err in selecting the wrong kanji from amongst the set of phonologically similar candidate representations that have been activated.

However, the results from Hatta, et al. (1997) and Hatta, et al. (1998) would seem to tell us very little about the types of information that might be activated and utilized when trying to retrieve and write a kanji when the individual is less than confident about its orthographic representation. Although the writer may have identified the target word in terms of its meaning and phonology, the activated semantics and phonology may fail to activate a sufficiently distinct orthographic representation of the word—that is, fail to achieve lexical retrieval of the orthographic form from long-term memory—and so the individual would either experience complete retrieval failure (no orthographic representation is activated) or a lack of confidence for only vaguely or partially activated orthographic representations. As Hatta and colleagues suggest, in some cases, when the writer is uncertain about their recall of the appropriate kanji character, they could resort to writing the word in kana orthography, but that option is not always available because excessive usage of kana orthography is associated with varying levels of social stigma depending on the circumstances. When kanji orthography is really called for, but the native Japanese writer is uncertain about the correct kanji, whenever possible, they will consult a dictionary, or, increasingly, rely on electronic devices, such as the ubiquitous Japanese cell phone. However, in constract the informal or spontaneous writing situations from which Hatta and colleagues collected the corpus of essays and letters that yielded their writing error data, it is probably not so uncommon for some native Japanese writers to be in situations where they need to write a word in kanji orthography, but are unable to retrieve with certainty the proper orthographic information from long-term memory. It seems reasonable to assume that many of the respondents, who participated in the traditional paper-based word association survey conducted by Joyce (2005), sometimes found themselves in such realistic situations. Accordingly, the present writing error data which from questionnaire answer sheets would seem to be more instructive about the types of information utilized when an individual tries to retrieve and write a kanji that they are less than confident about. Also, while the studies by Hatta and his colleagues primarily focused only on kanji writing errors for two-kanji compound words, because the present study assumes that the lexical retrieval processes for kanji orthography are similar irrespective of the overall structure of the target word, all the kanji orthography writing errors identified in the data are analyzed together, yielding a much larger data set.

Method

Materials. The writing error data analyzed in the present study consists of the 1,078 Japanese writing errors identified within the written responses collected for a word association survey. During the initial stages of the ongoing construction of the large-scale Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007a; Joyce & Miyake, 2008), Joyce (2005) conducted two surveys using traditional pen-and-paper questionnaires.[11] Although the two surveys varied in terms of their coverage of the JWAD survey corpus, the basic procedures were the same. In total, 1,486 native Japanese university students (934 males and 552 females; average age 19.03, SD = 0.97) participated in the surveys as volunteers. The questionnaires were mostly distributed at various English language classes at a national university catering for students from many different majors. Generally, the participants took about 15 to 20 minutes to complete the word association questionnaires.

Each participant made word association responses on questionnaire sheets that consisted of 100 items. Each respondent list was printed with 10 items per page, and the items were printed in 18pt Mincho font beside an underlined blank space for the response (e.g., 本 _____). The word association task asked the participants to read the printed items (basic Japanese kanji and words) and to write down the first semantically-related word that came to mind. The questionnaire also included a couple of instructions relating to aspects of the Japanese writing system. The first instruction related to the multiple scripts of Japanese, and asked the participants to respond with what they considered to be the most natural orthographic representation of the association word. For example, on seeing 本 /hon/ 'book' and thinking of the related word of /manga/ 'comic book', the participants were being asked to select the most appropriate orthographic representation; in kanji as 漫画, in hiragana as まんが, or in katakana as マンガ. Of even greater relevance for our present discussions, because there was a concern that participants might respond with a different word if they found that they could not recall how to correctly write the kanji of the first response they thought of, the questionnaires also included an instruction requesting that in such circumstances, rather than thinking of another response, they should write as much of the word as they could, but to indicate that they were not confident of the correct strokes by providing the word's pronunciation in a hiragana gloss above the word. This instruction encouraged the respondents to attempt to write down words that they were not confident about being able to write correctly, in order to improve the reliability of the word association data.[12]

Writing error classification. The present study analyzes and classifies all the 1,078 writing errors, irrespective of the orthographic form of the target word. Thus, present classification also distinguishes between writing errors for kanji orthography and writing errors for kana orthography. The classification of writing errors for kanji orthography words essentially draws on the classification of writing errors for two-kanji compound words proposed by Hatta and colleagues. Although writing error data analyzed in Hatta, et al (1998) only consisted of two-kanji compound words, most of the basic categories can be applied to any kanji writing errors. However, one important departure from the classification proposed by Hatta and colleagues is in the handling of non-kanji items. Although Hatta, et al. (1998) treated non-kanji as a separate category that accounted for 15.0% of their data, in the present study, all kanji

writing errors are classified according to their orthographic, phonological and semantic similarity with the target kanji. Accordingly, the present classification includes nine categories for kanji orthography errors

In extending the classification scheme to include writing errors for kana orthography, four new categories are added. The first new category relates to okurigana errors in the case of mixed kanji and hiragana orthography words, where there were either too many or too few kana characters. The second new category is for kana spelling errors where a mora is represented by an inappropriate character. The third category relates to nigori placement errors, where nigori 'voicing diacritics' marks were either omitted or added in error. The final new for kana orthography words is phonological transcription errors, where the phonological representation deviates from the standard transcription.

Results

Of the 1,486 participants in the word association surveys conducted by Joyce (2005), 762 (51.2%) committed at least one writing error, with the average number of errors being 1.41 (SD = 0.7), and the maximum number of errors being 5.

In total, 1,078 writing errors were identified amongst the word association responses. Table 1 presents a breakdown of the writing errors as a function of the orthographic form of the target word. Table 1 indicates that kanji orthography writing errors occurred most frequently in the context of attempting to write two-kanji compound words (47.9%) as target association responses.[13]

Table 1. Breakdown of writing errors as a function of target word orthography

Orthographic form of the target word	Numbers	%
Single kanji character	51	4.7
Single kanji + okurigana	173	16.0
Two-kanji compound words	516	47.9
Two kanji + okurigana	66	6.1
Three-kanji compound words	102	9.5
Three kanji + okurigana	16	1.5
Four-kanji compound words	71	6.6
Hiragana	27	2.5
Katakana	32	3.0
Others (including mixed orthography phrases)	24	2.2
Total	1,078	100.0

In classifying the writing error data according to the present classification scheme, the first division was between errors relating to kanji orthography and those relating to kana orthography. This revealed that there were 924 writing errors for kanji orthography and 154 writing errors for kana orthography. These two main categories were further classified according to the type of error. The complete classification of writing errors as a function of orthography and error type is presented in Table 2, on the next page. Examples of the major classification categories (scanned images of some of the writing errors) are presented in the Appendix.

Table 2. Classification of writing errors as a function of orthography and error type

Error category	Numbers	%
Kanji orthography writing errors		
Phonologically-related writing errors (P type)	309	33.4
Orthographically-related writing errors (O type)	254	27.5
Semantically-related writing errors (S type)	28	3.0
Phonologically and orthographically-related (P+O)	171	18.5
Phonologically and semantically-related (P+S)	79	8.5
Orthographically and semantically-related (O+S)	17	1.8
Phonologically, orthographically, semantically-related (P+O+S)	8	0.9
Reversals	32	3.5
Others	26	2.8
Total	924	100.0
Kana orthography writing errors		
Okurigana errors	89	57.8
Kana spelling errors	11	7.1
Nigori placement errors	13	8.4
Phonological transcription errors	10	6.5
Others	31	20.1
Total	154	100.0

Discussion

While it should be noted that 1,078 writing errors represent less than approximately 1% of all the word association responses collected in the word association surveys conducted by Joyce (2005),[14] still, the finding that 51.2% of the participants committed at least one writing error would seem to support the notion that making written responses to a traditional pen-and-paper questionnaire format is very different from the informal or spontaneous writing situations that Hatta and colleagues drew on in collecting their writing error data. It would seem that many of the respondents in the word association survey experienced a situation that is probably not so uncommon for some native Japanese writers; of wanting to write a word in kanji orthography but, at that moment, finding themselves unable to retrieve with certainty the proper orthographic information from long-term memory.

Turning to the error categories for the kanji orthography writing errors, it is appropriate to first look at the three basic, single factor, substitution errors. As presented in Table 2, 33.4% were phonologically-related errors (P type), while 27.5% were orthographically-related writing errors, with comparatively very few semantically-related writing errors at just 3.0%. Thus, in the present findings, phonologically-related errors were a little more frequent than orthographically-related errors. However, in comparison with the single factor substitution errors reported by Hatta, et al. (1998), the most striking difference would seem to be in the relative ratios of these three basic substitutions. Although there was clearly some variation in the range of frequencies for the three basic substitutions (P type = 17.9%, O type = 9.1%, and S type = 6.0%) in Hatta, et al. (1998), the difference in the frequencies of phonologically-related errors (33.4%) and orthographically-related errors (27.5%) is much smaller in

the present error data, suggesting that writing errors are almost as likely to be induced by the activation of orthographic information, when a native Japanese writer is attempting to write a word in kanji orthography and may be experiencing some uncertainty about the proper form of the kanji in question. To be fair, we should acknowledge that Hatta, et al.'s (1998) suggestion that phonologically-related errors are a major factor in kanji writing errors was based primarily on comparisons of summed frequencies for all phonologically-related errors, for all orthographically-related errors, and all semantically-related errors. However, the most important multiple-factor error category for the present data is writing errors that reflect both phonological and orthographical similarities (P+O type), which is the third largest category overall at 18.5%. Thus, the difference between the P and P+O types combined (51.9%) and the O and P+O types combined (46.0%) would still seem to be relatively small.

The outline of the cognitive processes involved in Japanese handwriting production, sketched out earlier, suggested the inclusion of an orthography selection process, where a writer must decide on the most appropriate orthographic representation of target words; either to write in kanji orthography or kana orthography. The word association questionnaire included an instruction that explicitly asked the participants to consider this selection. However, the word association questionnaire also included an instruction that effectively requested the participants to forgo the default strategy of employing kana orthography in cases where they were not confident about correctly writing the relevant kanji character. As included in the writing error examples in the Appendix, one participant attempted to write as a word association response the word /itazura/ 'mischief, prank'. Although the target kanji orthography of 悪戯 is an example of jukujikun,[15] the participant decided that kanji orthography was appropriate. It would, however, seem that on trying to retrieve the proper kanji characters from long-term memory, the individual could not recall the correct form of the second component kanji 戯 in its entirety. Still, in writing down "虚?", it is clear that the individual had a sense for the overall form of the target kanji character, and while they could not recall the 戈 component on the right-hand side, they could retrieve the left-hand side component of 虚 and they were aware that it was the left-side component and that some other component was necessary for the right-hand side.

Thus, while the present classification results for the kanji orthography writing errors are generally consistent with the notion that the activation of phonological information is an important factor in the production of kanji writing errors (Hatta, et al., 1998), they also indicate that the activation of orthographical information is also an important factor in the production of kanji writing errors. Even when native Japanese writers are not confident about the correct strokes for a particular kanji character, they often have a partial visual image for the target kanji's components or have a vague visual outline for its overall form.

Finally, it is interesting to note, while the kana orthography writing errors represent only 14.3% of the writing error data, as many as 57.8% of those are okurigana errors. Given that the survey participants were primarily university freshmen who would have received similar educations concerning okurigana usage rules, this finding may reflect the fact that the usage rules are not so straightforward and so young Japanese adults may not have consistent knowledge of okurigana usage.

In summary, this paper has presented a tentative analysis of Japanese writing errors observed within responses collected for a word association survey conducted as part of

the ongoing construction of the large-scale Japanese Word Association Database (JWAD) (Joyce, 2005, 2006, 2007a; Joyce & Miyake, 2008), as a source of insights into the organization of orthographic representations within the Japanese mental lexicon.

Notes

[1.] Aspects of this research were presented at the 71st Annual Meeting of the Japanese Psychological Association, 18-20th September, 2007, held at Toyo University, Tokyo, Japan. This research was supported by the 21st Century Center for Excellence Program "Framework for Systematization and Application of Large-scale Knowledge Resources" headed by Prof Furui at Tokyo Institute of Technology, Tokyo. It was also supported by a Grant-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science (Research project: 18500200. "Daikibo Nihongo Rensōgo Dētabēsu no Kōchiku/ Riyō ni Yoru Goi Chishiki no Mappingu" (Mapping lexical knowledge through the construction and application of a large-scale database of Japanese word associations).

[2.] Joyce (2008, in press) argues that portrayals of the Japanese writing system as being highly complex may be more a reflection of limitations and misunderstandings with typologies of writing systems, and, specifically, a consequence of the common failure to fully acknowledge the morphographic principle—where orthographic units represent morphemes—within linguistic classifications of writing systems Joyce (2008, in press) also suggests that the conceptual shift that comes with clearly recognizing that it is *more precise* to refer to kanji as being morphographic in nature, rather than logographic which is highly misleading, has profound consequences for our understanding of writing systems and writing, as well as the kinds of questions that researchers might ask about the nature and organization of the literate mental lexicon.

[3.] While the issue of Japanese literacy undoubtedly warrants far more research that it has received to data, questions of just how functional literacy within Japanese contexts might be defined are not touched on in the present paper. However, research into the topic could undoubtedly benefit from drawing on Hatano's (1995) attempt to expand on and enrich the 'practice account of literacy', as proposed by Scribner and Cole (1981), by applying it to the question of literacy in standard Japanese orthography (see also Akita & Hatano, 1999; Makita, 1968; Stevenson, Lee, Stigler, Kitamura, Kimura, & Kato, 1986; Taylor & Taylor, 1995). In that context, it is worth a moment's pause to also reflect on the comments of Kess and Miyamoto (1999) regarding the impact on kanji usage due to their easy accessibility with electronic devices. They write, "there may even come a time in Japanese when there will be two character sets in actual practice; one will be for reading comprehension, a read-one set, while the other set will be a write-only set, the set that literate Japanese will have to be able to produce in handwriting tasks … Perhaps the reality is that the Japanese are already there" (p. 30). Kess and Miyamoto's perceptive comments would seem to have particular resonance both for the recently much-publicized reading mistakes of the Japanese Prime Minister Aso (Yamaguchi, 2009) and for the proposals currently under discussion to increase the numbers of official Jōyō Kanji from the present level of 1,945 to 2,131 kanji characters (Sayama & Sekine, 2009).

[4.] In the absence of suitable quantitative data from which to make appropriate judgments and comparisons of scale, it is, for example, very difficult to judge just how realistic are the 'societal' expectations voiced by Hatta, Kawakami, and Tamaoka (1998). After noting that "children usually begin to learn the writing system of their mother tongue as soon as they start school and are expected to master the rules which govern handwriting", and that because "more than 95 percent of 15-year-old students go on to senior high school (Grades 10-12), most Japanese students learn Japanese for 12 years", Hatta, et al. comment that "it is strongly expected that Japanese university students will achieve mastery of the kanji writing system. In actual fact, the true situation differs somewhat from the assumed norm: university students do occasionally experience difficult in kanji reading and writing" (p.457). However, the astute reader will immediately recognize that the expression of "occasionally experience" is simply too vague to have any real meaning. A moment of reflection should suffice for us to reject the implied assumption that university students of other nationalities never experience difficulties in reading and writing the standard orthographic forms of their mother tongues. It is, of course, impossible to appropriately contextualize such comments without proper scales of reference. Hatta, Kawakami, and Hatasa (1997) and Hatta, Kawakami, and Tamaoka (1998) both draw on a corpus of 374 writing errors involving two-kanji compound words, and while neither study notes the actual size of the corpus of essays and letters in which

the writing errors were found, Hatta et al. (1997) acknowledge that it consists of writings collected from 2,200 students over a seven-year period. Based on the number of students and the substantial period, an incidence of just 374 writing errors would seem to be incredibly low.

[5.] Hatta, et al.'s (1997) Study 1 also analyzed 62 kanji writing errors within three-kanji compound words, but the present paper will focus on just the two-kanji compound words, because Hatta, et al. (1998) only refer to the 374 writing errors for two-kanji compound words (albeit without acknowledging the existence of Hatta, et al. (1997)). The present author assumes that both studies are referring to one identical set of errors, because the numbers and the breakdowns according their error classifications match exactly.

[6.] Hatta, et al. (1997) note that these included "two nurse training colleges, one college of kindergarten workers, one college for the care of the aged, and one teacher training university" (p. 405).

[7.] Although substitutions based on orthographically similar kanji are referred to as morphographically similar (M type) in Hatta, et al. (1997), the modified term of orthographic in Hatta, et al. (1998) is more appropriate.

[8.] Hatta, et al. (1997) seems to make a mistake in calculating the percentage of writing errors that involved semantically-related substitutions, leading them to overemphasize the importance of semantic similarities as a source of kanji writing errors.

[9.] In the case of Levelt's (1995) speech production model, the third stage is usually referred to as articulation. Hatta, et al. (1998) suggest a cognitive model of kanji writing that involves the three main stages of concept formation, kanji lexicon, and motor act, which is structurally quite similar.

[10.] Although Hatta, et al. (1998) posit the choice of kanji orthography versus kana orthographic at their concept formation stage, for the present author, it seems more natural to assume it to be a sub-process of the formulation stage.

[11.] A web-based version of the word association survey has subsequently been developed in order to facilitate the collection of large-scale quantities of word association responses. The survey is currently available at http://gn.valdes.titech.ac.jp/1free/index.jsp.

[12.] Although hiragana glosses do not accompany all the kanji writing errors, the frequent inclusion of glosses on the questionnaire sheets (for both correctly and incorrectly written kanji orthography association responses) indicates that most of the participants took this instruction seriously. The author also witnessed respondents consulting electronic dictionaries and cell phones to check on their written responses.

[13.] This would seem to be consistent with the claim that two-kanji compound words are the most common word structure in Japanese (Nomura, 1988).

[14.] The figure of 1,078 is 0.72% of 148,600, which is the total number of questionnaire items (100 items x 1,486 participants). However, the questionnaire item total should be adjusted slightly for a small number of non-responses due to either not completing the questionnaire to the end or inadvertently skipping a page.

[15.] Jukujikun can be understood as orthographic representations of words, where the meanings associated with the component kanji may relate to the overall meaning of the word, but where there is no correspondence between the pronunciations associated with the component kanji and the pronunciation of the whole word.

References

Akita, K., & Hatano, G. (1999). Learning to read and write in Japanese. In M. Harris & G. Hatano (Eds.), *Learning to read and write: A cross-linguistics perspective* (pp. 214-234). Cambridge: Cambridge University Press.

Carroll, D. W. (2004). *Psychology of language*. (4th ed.). Belmont, CA: Wadsworth/Thompson Learning.

Coulmas, F. (1989). *The writing systems of the world*. Oxford: Basil Blackwell.

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

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47,* 27-52.

Fromkin, V. A. (Ed.). (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Gnanadesikan, A. E. (2009). *The writing revolution: Cuneiform to the internet*. Oxford: Wiley-Blackwell.

Fischer, S. R. (2001). *A history of writing*. London: Reaktion Books.

Harley, T. A. (2001). *The psychology of language: From data to theory*. (2nd ed.). Hove: Taylor & Francis.

Hatano, G. (1995). The psychology of Japanese literacy: Expanding 'the practice account'. In L. M. W. Martin, K. Nelson, & E. Tobach (Eds.), *Sociocultural psychology: Theory and practice of doing and knowing* (pp. 250-275). Cambridge: Cambridge University Press.

Hatta, T., Kawakami, A., & Hatasa, Y. (1997). Kanji writing errors in Japanese college students and American Japanese learners. In H.-C. Chen (ed.). Cognitive processing of Chinese and related Asian languages. Sha Tin, N.T., Hong Kong: Chinese University Press.

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, 10, 457-470.

Joyce, T. (2005). Constructing a large-scale database of Japanese word associations, In K. Tamaoka, (Ed.), *Corpus Studies on Japanese Kanji*. (Glottometrics 10). pp. 82-98. Hituzi Syobo: Tokyo, Japan and RAM-Verlag: Lüdenschied, Germany.

Joyce, T. (2006). Mapping word knowledge in Japanese: Constructing and utilizing a large-scale database of Japanese word associations. In *Proceedings of the International Symposium on Large-Scale Knowledge Resources (LKR2006)* (pp.155-158), Tokyo: Tokyo Institute of Technology.

Joyce, T. (2007a). Mapping word knowledge in Japanese: Coding Japanese word associations. *Proceedings of the Symposium on Large-Scale Knowledge Resources (LKR2007)* (pp. 233-238), Tokyo: Tokyo Institute of Technology.

Joyce, T. (2007b). Rensōgo chōsa no hannō de kansatsu sareta kakimachigai no kentō [Study of the writing errors observed in responses to a word association survey]. Poster presentation given at the *71th Annual Conference of Japanese Psychological Association*. 18-20 September, Toyo University, Tokyo.

Joyce, T. (2008). The significance of the morphographic principle for the classification of writing systems. Oral presentation given at *Typology of Writing Systems: Sixth International Workshop on Writing Systems*, 18-19 September, Braunschweig, Germany.

Joyce, T. (in press). The significance of the morphographic principle for the classification of writing systems. *Written Language and Literacy* (Special issue on Typology of Writing Systems edited by S. Borgwaldt & T. Joyce).

Joyce, T., & Miyake, M. (2008). Capturing the structures in association knowledge: Application of network analyses to large-scale databases of Japanese word associations. In A. Ortega & T. Tokunaga (Eds.), *The 3rd International Conference on Large-scale Knowledge Resources (LKR 2008)*. (Lecture Notes in Computer Science). (pp. 116-131). Berlin: Springer-Verlag.

Kess, J. F. (2005). On the history, use, and structure of Japanese kanji. In K. Tamaoka, (Ed.), *Corpus Studies on Japanese Kanji*. (Glottometrics 10). pp. 1-15. Hituzi Syobo: Tokyo, Japan and RAM-Verlag: Lüdenschied, Germany.

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

Levelt, W. J. M. (1995). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Makita, K. (1968). The rarity of reading disability in Japanese children. *American Journal of Orthopsychiatry, 38*, 599-614.

Nomura, M. (1988). Niji kango no kōzō [The structure of 2 kanji Sino-Japanese words]. *Nihongogaku, 7*, 5, 44-55.

Robinson, A. (1995). *The story of writing: Alphabets, hieroglyphs & pictograms*. London: Thames & Hudson.

Sansom, G. (1928). *A historical grammar of Japanese*. Oxford: Oxford University Press.

Sayama, M., & Sekine, K. (2009). Kisō kara wakaru jōyō kanji: "Shiyō no meyasu" minaoshichū. *Yomiuri Shimbun*, 25 February.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Smith, J. S. (1996). Japanese writing. In P. T. Daniels & W. Bright (Eds.), *The world's writing systems* (pp. 209-217). New York: Oxford University Press.

Sproat, R. (2000). *A computational theory of writing systems*. Stanford: Cambridge University Press.

Stevenson, H. W., Lee, S.-Y., Stigler, J., Kitamura, S., Kimura, S., & Kato, T. (1986) Learning to read Japanese. In H. Stevenson, M. Azuma, & K Hakuta, (Eds.), *Child development and education in Japan* (pp. 217-235). New York: Freeman.

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean and Japanese*. Amsterdam: John Benjamins.

Unger, J. M. (2004). *Ideogram: Chinese characters and the myth of disembodied meaning*. Honolulu: University of Hawai'i Press.

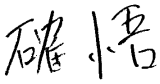
Yamaguchi, M. (2009). Aso's kanji conundrums spur self-reflection, textbook sales. *Japan Times*, Saturday 7 March, available online at http://search.japantimes.co.jp/cgi-bin/nn20090307f3.html.

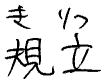
Appendix: Writing error examples

Kanji orthography writing errors

Phonologically-related writing errors (P type) (n = 309)

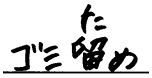
地獄 /jigoku/ 

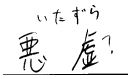
覚悟 /kakugo/ 

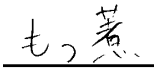
規律 /kiritsu/ 

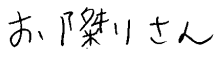
強調 /kyōchō/ 

Orthographically-related writing errors (O type) (n = 254)

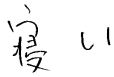
ゴミ溜め /gomitame/ 

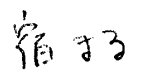
悪戯 /itazura/ 

もつ煮 /motsuni/ 

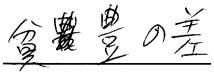
お隣さん /otonarisan/ 

Semantically-related writing errors (S type) (n = 28)

眠い /nemui/ 

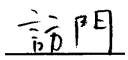
泊まる /tomaru/ 

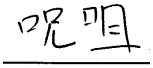
暗室 /anshitsu/ 

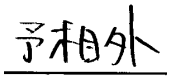
貧富の差 /hinpu no sa/ 

Phonologically and orthographically-related (P+O) (n = 171)

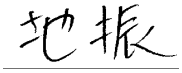
疑問 /gimon/ 

訪問 /hōmon/ 

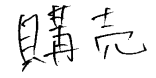
呪詛 /juso/ 

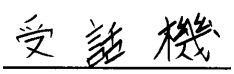
予想外 /yosōgai/ 

Phonologically and semantically-related (P+S) (n = 79)

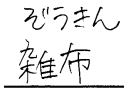
地震 /jishin/ 

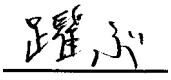
騒音 /sōon/ 

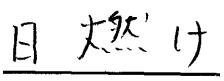
購買 /kōbai/ 

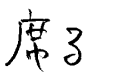
受話器 /juwaki/ 

Orthographically and semantically-related (O+S) (n = 17)

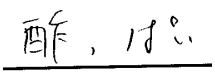
雑巾 /zōkin/ 

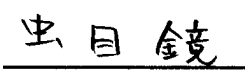
跳ぶ /tobu/ 

日焼け /hiyake/ 

座る /suwaru/ 

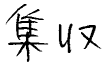
Phonologically, orthographically, semantically-related (P+O+S) (n = 8)

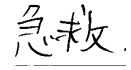
酸っぱい /suppai/ 

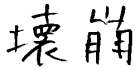
虫眼鏡 /mushimegane/ 

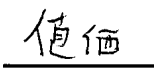
Kanji orthography writing errors: Continued

Reversals (n = 32)

収集 /shūshū/ 

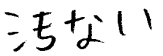
救急 /kyūkyū/ 

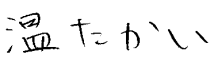
崩壊 /hōkai/ 

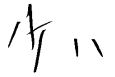
価値 /kachi/ 

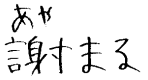
Kana orthography writing errors

Okurigana errors (n = 89)

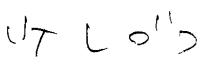
汚い /kitanai/ 

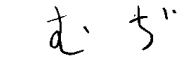
暖かい /atatakai/ 

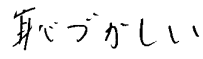
少ない /sukunai/ 

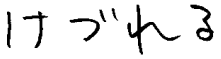
謝る /ayamaru/ 

Kana spelling errors (n = 11)

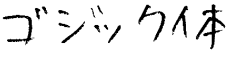
少しずつ /sukoshi zutsu/ 

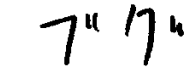
むじ /muji/ 

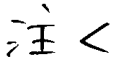
恥ずかしい /hazukashii/ 

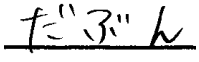
けずれる /kezureru/ 

Nigori placement errors (n = 13)

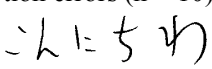
ゴシック体 /goshikkutai/ 

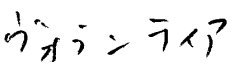
フグ /fugu/ 

注ぐ /sosogu/ 

たぶん /tabun/ 

Phonological transcription errors (n = 10)

こんにちは /konnichwa/ 

ボランティア /borantia/ 

それでは /soredewa/