Homography-induced ambiguity in Japanese *kanji* and the lexical disambiguating function of *okurigana*

Keisuke HONDA

1 Introduction

In the current Japanese writing system, graphemes of the *kanji* script denote native and Sino-Japanese lexical morphemes.[1] A large number of these graphemes classify as homographic, denoting two or more different morphemes at the same time. This is illustrated in (1), where the grapheme 食 corresponds to a set of five distinct morphemes.[2]

(1) 食$_1$
{*tabe-*, *kuw-*, *kuraw-*, *SYOKU*, *ZIKI*}$_1$

Homographic *kanji* are potentially ambiguous because they contain multiple possibilities of decoding without making any visible differentiation (Sampson 1985,

[1] In the literature, there is no consensus on how to define the term *grapheme*. This paper adopts the conventional conception that grapheme is the basic unit of writing used to represent a certain unit or units of language (Sproat 2000, Coulmas 2003, Rogers 2005, etc.). As regards *kanji*, this term will be used to refer to the so-called 'character' – not the graphic constituent thereof.

Also disputed is the identity of the linguistic entities that are denoted by *kanji* graphemes. Although this issue deserves a full discussion, it is out of the scope of the present paper. For the sake of convenience, this paper tentatively uses (*lexical*) *morphemes* to refer to these entities, although this term is not adequate in some respects.

[2] Throughout this paper, examples will be given in the following format. The first and second lines present graphemes and associated morphemes, respectively. The morphemes are transcribed morphophonemically, with Sino-Japanese items shown in uppercase characters. The grapheme-to-morpheme correspondences are indicated by numerical coindexation. When necessary, a third line will be added to display the morphological structure and English translation of the words encoded.

Kaiser 1995, Coulmas 2003, Kondō 2005). In actual reading, the human reader may be able to infer the intended item from the orthographic, linguistic, textual and other kinds of context. Putting it the other way round, however, the decoding of homographic *kanji* is entirely dependent on the reader's psycholinguistic abilities. As long as matters internal to the writing system are concerned, the problem of homography-induced ambiguity can be thus regarded as a serious constraint on the system's self-sufficiency.

Meanwhile, it has been suggested that *okurigana* provide an effective solution to this problem (Hill 1967, Sampson 1985, Kamei et al. 1996, Vance 2002, Coulmas 2003, Kondō 2005).[3] In their present form, *okurigana* are constituted by *hiragana* graphemes, which encode individual mora-bearing phoneme chunks (hereafter, *morae*). As will be elaborated on in section 2, they are postposed to *kanji* graphemes to spell out a latter section of native words (2a) or elements of native compound words (2b).

(2) a. 食$_1$ べ$_2$ る$_3$ b. 食$_1$ べ$_2$ 過$_3$ ぎ$_4$
tabe$_1$ *be*$_2$ *ru*$_3$ *tabe*$_1$*be*$_2$ *sugi*$_3$ *gi*$_4$
[[*tabe*]*ru*] 'eat' [[*tabe*][*sugi*]] 'overeating'

When applied to homographic *kanji*, *okurigana* indicate the realised phonological form of one particular morpheme (3). This, in effect, specifies the intended item, ruling out all other potential candidates at the same time. The present paper refers to this function as the *lexical disambiguating function* of *okurigana*.

(3) 食$_1$ べ$_2$ る$_3$
{*tabe-*, *kuw-*, *kuraw-*, *SYOKU*, *ZIKI*}$_1$ *be*$_2$ *ru*$_3$

At first glance, the lexical disambiguating function appears to offer a drastic solution

[3] The author is grateful to Prof. Stefan Kaiser for turning his attention to Hill (1967).

to the problem of homography-induced ambiguity. However, few attempts have been made to work out how and how effectively *okurigana* disambiguate homographic *kanji*. Accordingly, it is open to question how much ambiguity is actually removed by the application of *okurigana*.

The present paper makes a critical appraisal of this topic. In section 2, the background will be developed for the discussion that will follow. In section 3, a survey will demonstrate that *okurigana* disambiguate a relatively small proportion of homographic *kanji* that are used today. Finally in section 4, it will be concluded that the lexical disambiguating function of *okurigana* does not provide a radical solution to the problem of homography-induced ambiguity.

2 Background

2.1 (Potential) Homography-Induced Ambiguity in Japanese

Homography is a situation in which a single unit of writing corresponds with two or more units of language (Vance 2002, Rogers 2005).[4] Homographic representations do not alter their form (e.g., replacement of graphic constituents, addition of diacritics, etc.) in accordance with distinctive contrasts that exist between the linguistic entities that they encode (Rogers 2005). Neither do they specify any particular linguistic entity as the intended item in a visible way. Consequently, it is not possible to retrieve the intended item by just looking at homographic representations per se. If a writing system allows a large number of homographic representations, this means that there is a large amount of ambiguity, which, in turn, opacifies the entire system to a considerable degree.

Potentially, this is the case in Japanese. As has been mentioned in section 1, the bulk of *kanji* graphemes classify as homographic, denoting two or more lexical morphemes. In the officially approved *jōyōkanji* set (see section 3.2 for details),

[4] Homography (a.k.a. *polyphony*) is the opposite of *homophony* (a.k.a. *polygraphy*), a situation in which two or more units of writing correspond with a single unit of language. Homography and homophony are subcategories of *polyvalence* (Vance 2002, Coulmas 2003) or *contrastive discrepancies* (Rogers 2005), non-biunique correspondences of orthographic units and linguistic units.

1,249 out of 1,945 graphemes denote two to 12 lexical morphemes (Nomura 1981). In addition, the number of possible readings can be multiplied if the same *kanji* is used to encode morphologically different word-forms of the same lexeme. In short, many *kanji* graphemes embrace the problem of homography-induced ambiguity at the lexical level (4a) and/or the word-form level (4b). Since these graphemes are indispensable elements of the grapheme inventory, it is reasonable to presume that this problem is prevalent in the current Japanese writing system.[5]

(4) a. Lexical ambiguity

食$_1$

{*tabe-*, *kuw-*, *kuraw-*, . . .}$_1$

b. Word-form ambiguity

食$_1$

tabe-ru	*kuw-u*	*kuraw-u*
tabe-nai	*kuw-anai*	*kuraw-anai*
tabe-te	*kuw-te*	*kuraw-te*
. . .	. . .	. . .

$_1$

2.2 The Disambiguating Function of *Okurigana*

As a matter of fact, the addition of *okurigana* to *kanji* graphemes reduces the number of homographic representations and, in turn, the amount of homography-induced ambiguity (Tsukishima and Kabashima 1984, Kondō 2005). At the word-form level, the combination of *kanji* and *okurigana* assigns a unique orthographic representation to each distinct form of inflected words and their derivatives. This effectively prevents homography and the consequent ambiguity altogether. Besides, the disambiguation of word-forms also contributes to disambiguating homographic *kanji* at the lexical level. Moreover, there are certain types of *okurigana* that are designed and used solely for the purpose of lexical disambiguation. These points will be elaborated on below.

[5] In practice, it is possible to prevent this kind of ambiguity by replacing *kanji* with *hiragana* graphemes for an explicit phonographic representation. (The author thanks Assoc. Prof. Jun Ikeda for turning his attention to this point.) However, this is a makeshift patch-up rather than a regularly used alternative.

2.2.1 Word-Form Disambiguation

Okurigana preclude homographic representations entirely at the word-form level. For one primary function, they spell out various endings of inflected words and their derivatives (Tsukishima 1970, Sampson 1985, Abe 1989, Kamei et al. 1996, Vance 2002, Coulmas 2003). In this way, individual word-forms are orthographically differentiated from each other in an unequivocal fashion. For instance, observe the examples (5a, b) and (6a, b).[6]

(5) a. 食$_1$　べ$_2$　る$_3$　　b. 食$_1$　べ$_2$　な$_3$　い$_4$
tabe$_1$ *be*$_2$ *ru*$_3$　　*tabe*$_1$*be*$_2$ *na*$_3$ *i*$_4$
[[*tabe*]*ru*] ‘eat’　　[[[*tabe*]*na*]*i*] ‘not eat’

(6) a. 飲$_1$　む$_2$　　b. 飲$_1$　ま$_2$　な$_3$　い$_4$
nom$_1$ *mu*$_2$　　*nom*$_1$ *ma*$_2$ *na*$_3$ *i*$_4$
[[*nom*]*u*] ‘drink’　　[[[*nom*]*ana*]*i*] ‘not drink’

In the examples above, the general principle is that *okurigana* spell out all word endings, namely the tense suffix (*r*)*u* and the negation suffix *nai*. In many cases, *okurigana* also cover a certain section at the stem-final position.[7] For instance, they spell out the stem-final mora in the V-verb (5a, b), or the combination of the

[6] This paper assumes that the usage of *okurigana* follows regulations stipulated by the current version of *Okurigana no Tsukekata* (Cabinet 1981b), a body of official guidelines first promulgated in 1973 and then revised in 1981. Though these regulations have no legal power, they constitute a de facto standard for *okurigana* orthography, adopted in legal and public documents as well as approved school text books. They are also used in major newspapers and other forms of publication, usually with a greater or lesser number of self-authorised by-laws. Judging from the wide acceptance, it is reasonable to regard these regulations as representing the archetypal usage of *okurigana* in present-day Japan.

[7] This description is based on the morphological analysis of Japanese verbs, which differs in several respects from the traditional analysis underpinning the current *okurigana* orthography. One notable difference lies in the delineation of verb stems. For instance, the morphological analysis of the verbs *taberu* (5a) and *nomu* (6a) is *tabe-ru* and *nom-u*, respectively, where *tabe* and *nom* constitute verb stems. In the traditional analysis, the same verbs are treated as *ta-be-ru* and *no-mu*, respectively, where *ta* and *no* constitute verb stems.

stem-final consonant and the suffix-initial vowel in the C-verb (6a, b). There exist several other factors which further complicate the actual application of *okurigana*. Some of these will be mentioned in section 2.2.2. For the purpose of the present discussion, though, it suffices to observe that *okurigana* preclude homographic representations of different word-forms, thus blocking homography-induced ambiguity at the word-form level completely.

2.2.2 Lexical Disambiguation

In addition to word-form disambiguation, several researchers have suggested that *okurigana* serve to disambiguate *kanji* graphemes at the lexical level (Hill 1967, Sampson 1985, Kamei et al. 1996, Vance 2002, Coulmas 2003, Kondō 2005). In some cases, lexical disambiguation can be characterised as a by-product of word-form disambiguation. As Sampson (1985) maintains, the spelling out of the stem-final consonant in C-verbs may contribute to distinguishing between different verb stems. Coulmas (2003) echoes this point more affirmatively, saying that *okurigana* "*usually* reduce the choice of possible interpretations of the preceding character [i.e., *kanji*] to one" (p. 182; emphasis added by the author). This is illustrated by the grapheme 歩 (7a), which denotes (among others) two near-synonym C-verb stems, namely *aruk* and *ayum* 'walk'. The *okurigana* く *ku* (7b) and む *mu* (7c) suggest the respective verb stems by repeating the final consonant.

(7) a. 歩$_1$ — {*aruk-*, *ayum-*}$_1$

b. 歩$_1$ く$_2$ — {*aruk-*, *ayum-*}$_1$ *ku*$_2$

c. 歩$_1$ む$_2$ — {*aruk-*, *ayum-*}$_1$ *mu*$_2$

With regards to the current orthographic convention stipulated by *Okurigana no Tsukekata* (Cabinet 1981b) (see footnote 6 for details), lexical disambiguation of this sort can be extended further. For Sampson (1985) and Coulmas (2003), the basic idea is that *okurigana* repeat a certain section at the stem-final position. In fact, this holds true not only for C-verbs but also for many inflected words in general. This is

exemplified below, where *okurigana* repeat stem-final elements of the V-verb (8a), adjective (8b) and adjectival verb (8c).

(8) a. 食$_1$　べ$_2$　る$_3$　b. 少$_1$　な$_2$　い$_3$　c. 穏$_1$　や$_2$　か$_3$　だ$_4$
tabe$_1$　*be*$_2$　*ru*$_3$　*sukuna*$_1$*na*$_2$　*i*$_3$　*odayaka*$_1$*ya*$_2$　*ka*$_3$　*da*$_4$
[[*tabe*]*ru*] ‘eat’　[[*sukuna*]*i*] ‘few/little’　[[*odayaka*]*da*] ‘gentle’

At times, *okurigana* spell out only suffixes, leaving out the stem-final portion. Even in such cases, *okurigana* allow lexical disambiguation indirectly. As they partially represent the realised form of a particular morpheme, lexical disambiguation can be achieved in relation to other possible candidates. To cite a case, the *okurigana* い *i* in (9a) covers only the inflectional suffix *i*. Still, it provides sufficient information for distinguishing the intended reading *osoi* ‘slow’ from other possibilities such as *okureru* ‘be late’ (9b).

(9) a. 遅$_1$　い$_2$　b. 遅$_1$　れ$_2$　る$_3$
oso$_1$　*i*$_2$　*okure*$_1$　*re*$_2$　*ru*$_3$
[[*oso*]*i*] ‘slow’　[[*okure*]*ru*] ‘be late’

In other instances, *okurigana* are used exclusively for lexical disambiguation. This is particularly visible in words belonging to the same word family, exemplified by the three words shown in (10a-c). Morphologically, they share the common root *maz*, which forms the respective verb stems together with the transitive marker *e* (10a) or the intransitive markers *ar* (10b) or *ir* (10c). Here, *okurigana* are applied in such a way that they cover not only the inflectional suffix (*r*)*u*, but also the stem-final elements where contrasts occur. This is ascribed to an orthographic rule designed to distinguish between morphemes while maintaining the maximal uniformity in the orthographic representation.

(10) a. 混$_1$ ぜ$_2$ る$_3$ b. 混$_1$ ざ$_2$ る$_3$ c. 混$_1$ じ$_2$ る$_3$
maz(*e*)$_1$*ze*$_2$ *ru*$_3$ *maz*(*a*)$_1$*za*$_2$ *ru*$_3$ *maz*(*i*)$_1$ *zi*$_2$ *ru*$_3$
[[[*maz*]*e*]*ru*] ‘mix’ [[[*maz*]*ar*]*u*] ‘mix’ [[[*maz*]*ir*]*u*] ‘mix’

In addition to inflected words, *okurigana* are also used for many uninflected non-nominal words (11a) and nominals derived from inflected words (11b, c).[8]

(11) a. 但$_1$ し$_2$ b. 食$_1$ べ$_2$ で$_3$ c. 飲$_1$ み$_2$
tadasi$_1$*si*$_2$ *tabe*$_1$ *be*$_2$ *de*$_3$ *nom*$_1$ *mi*$_2$
[*tadasi*] ‘provided that’ [[*tabe*]*de*] ‘plenty to eat’ [[*nom*]*i*] ‘binge’

To generalise, *okurigana* usually spell out a certain portion at the final position of lexical morphemes (i.e., stems or free forms), with or without suffixes in inflected words or derivatives thereof.

At a more fundamental level, Kaiser (1995) points out that the presence and absence of *okurigana* often indicate the distinction of native and Sino-Japanese morphemes, respectively. As has been mentioned in section 1, according to the current orthographic convention, *okurigana* are used only for native items of the lexicon. It follows from this fact that if *okurigana* are added to a homographic *kanji*, they implicate that the intended reading is native, not Sino-Japanese. To illustrate this, observe the examples in (12a-c). Shown here are three near-synonyms for ‘food’, namely native *tabemono* (12a), native colloquial *kuimono* (12b), and Sino-Japanese *syokumotu* (12c). The presence of *okurigana* in the first two examples automatically precludes the possibility of the Sino-Japanese reading.

(12) a. 食$_1$ べ$_2$ 物$_3$ b. 食$_1$ い$_2$ 物$_3$ c. 食$_1$ 物$_2$
tabe$_1$ *be*$_2$ *mono*$_3$ *kuw*$_1$ *i*$_2$ *mono*$_3$ *syoku*$_1$ *motu*$_2$
[[*tabe*][*mono*]] ‘food’ [[[*kuw*]*i*][*mono*]] ‘food’ [[*syoku*][*motu*]] ‘food’

[8] Historically, the term *sutegana* (a.k.a. *sukegana* or *soegana*) is reserved for *kana* graphemes used for uninflected words (Kamei et al. 1996).

In the light of the observations thus far, the lexical disambiguating function can be summarised as follows. First, the presence of *okurigana* indicates that the *kanji* grapheme encodes a native lexical morpheme. Second, *okurigana* specify the intended reading by spelling out a final section of one particular morpheme, with or without the following inflectional and derivational suffixes. In theory, this rules out other possibilities of decoding and hence disambiguates homographic *kanji* at the lexical level.

2.3 Restrictions on the Lexical Disambiguating Function

In the previous section, it was shown that *okurigana* serve to disambiguate *kanji* graphemes at two levels. At the word-form level, they provide a unique representation to each distinct word-form. This prevents homography-induced ambiguity from occurring at this level. At the lexical level, on the other hand, *okurigana* can specify the intended morpheme. Notwithstanding, a closer inspection reveals that the lexical disambiguating function is heavily restricted by several conditions. These restrictions give rise to an empirical question about the actual effectiveness of disambiguation at the lexical level.

To begin with, there are certain restrictions on the applicability of *okurigana*. As has been mentioned in the previous sections, *okurigana* are applied to *kanji* graphemes only when these graphemes encode native morphemes. Seen from the other side, *okurigana* are powerless over *kanji* graphemes representing Sino-Japanese morphemes (13). This is a major drawback because these morphemes constitute the majority of the vocabulary items.[9]

(13)　*Okurigana* are not used for Sino-Japanese morphemes.

Similarly, the lexical disambiguating function is irrelevant to *kanji* graphemes

[9] Miyajima et al. (1987) cites a survey of number of headwords in a general dictionary published in 1969. There, 31,839 (52.9%) out of 60,218 headwords classified as Sino-Japanese. Even though these figures include a certain proportion of non-standard or rarely used words, they do illustrate the size of Sino-Japanese items in the vocabulary of present-day Japanese.

representing native nouns, except for those derived from inflected words such as verbs and adjectives (14) (for examples, see (11b) and (11c) above). Although *Okurigana no Tsukekata* (Cabinet 1981b) permits a dozen or so exceptions, they do not undermine the general principle of '*okurigana* for non-nominals'.[10] Again, this limits the amount of homography-induced ambiguity removed by *okurigana*.

(14) *Okurigana* are not used for native nouns that are not derivatives of inflected words.

The restrictions (13) and (14) limit the applicability of *okurigana* according to the type of morphemes that are denoted by homographic *kanji*. They generate more restrictions when combined with another factor, namely the number of morphemes. First, it follows from (13) that *kanji* graphemes are not fully disambiguated if they have two or more Sino-Japanese morphemes, irrespective of the number of native morphemes (15a). This is because *okurigana* cannot distinguish a Sino-Japanese *X* from other Sino-Japanese *Y*, *Z* and so forth. In other words, these graphemes inevitably produce homographic representations (15b, c).

(15) a. *Okurigana* cannot fully disambiguate *kanji* graphemes denoting two or more Sino-Japanese morphemes.

b. 悪$_1$
{*waru-*, *AKU*, *O*}$_1$

c. i. 悪$_1$ い$_2$ ii. 悪$_1$ iii. 悪$_1$
waru$_1$ *i*$_2$ *AKU*$_1$ *O*$_1$
[[*waru*]*i*] 'bad' [*aku*] 'evil' [o] 'bad'

Similarly, it follows from (14) that *kanji* graphemes are not fully disambiguated if

[10] These exceptions are the following: 辺り, 哀れ, 勢い, 幾ら, 後ろ, 傍ら, 幸い, 幸せ, 互い, 便り, 半ば, 情け, 斜め, 独り, 誉れ, 自ら, 災い. Numerals and an interrogative ending with *tu* also take *okurigana* (e.g., 一つ, 二つ, 三つ, 幾つ).

they have two or more native nouns (except for derivatives and the words shown in footnote 10), irrespective of the number of Sino-Japanese items and native non-nominals (16a). Again, these items necessarily create homographic representations (16b, c).

(16) a. *Okurigana* cannot fully disambiguate *kanji* graphemes denoting two or more native nominal morphemes, excluding nominalised items and other exceptions.

b. 空$_1$

{*ak-*, *aker-*, *sora*, *kara*, *KUU*}$_1$

c. i. 空$_1$　く$_2$　　ii. 空$_1$　　iii. 空$_1$

ak$_1$　*ku*$_2$　　*sora*$_1$　　*kara*$_1$

[[*ak*]*u*] ‘open’　　[*sora*] ‘sky’　　[*kara*] ‘empty’

Furthermore, it follows from (13) and (14) that homography occurs in *kanji* graphemes which denote one Sino-Japanese morpheme and one or more native nouns at the same time (17a). Because of the Sino-Japanese item, there is no room left for another orthographic representation without *okurigana* (17b, c).

(17) a. *Okurigana* cannot fully disambiguate *kanji* graphemes denoting one Sino-Japanese and one or more native non-derivative nouns that do not take *okurigana*.

b. 戦$_1$

{*tatakaw-*, *ikuasa*, *SEN*}$_1$

c. i. 戦$_1$　う$_2$　　ii. 戦$_1$　　iii. 戦$_1$

tatakaw$_1$*u*$_2$　　*ikusa*$_1$　　*SEN*$_1$

[[*tatakaw*]*u*] ‘battle’ (v.)　　[*ikusa*] ‘battle’ (n.)　　[*SEN*] ‘battle’

In addition, there are cases in which *okurigana* do accompany *kanji* graphemes and yet fail to prevent homographic representations (18). Some *kanji-okurigana*

combinations always produce identical forms (19a, b), while others become homographic in certain word-forms (20a, b).[11] In *Okurigana no Tsukekata* (Cabinet 1981b), there are some by-laws deliberately designed to differentiate between such pairs to avoid confusion. However, these are individual patch-ups and do not eliminate homographic representations completely.

(18) *Okurigana* cannot fully disambiguate certain *kanji* graphemes which produce two or more homographic representations even when *okurigana* are added.

(19) a.i. 開$_1$ く$_2$ ii. 開$_1$ い$_2$ て$_3$
ak$_1$ *ku*$_2$ *ak*$_1$ *i*$_2$ *te*$_3$
[[*ak*]*u*] ‘open’ [[[*ak*]*i*]*te*] ‘open-GERUND’
b.i. 開$_1$ く ii. 開$_1$ い$_2$ て$_3$
hirak$_1$ *ku*$_2$ *hirak*$_1$ *i*$_2$ *te*$_3$
[[*hirak*]*u*] ‘open’ [[[*hirak*]*i*]*te*] ‘open-GERUND’

(20) a.i. 行$_1$ く$_2$ ii. 行$_1$ っ$_2$ て$_3$
ik$_1$ *ku*$_2$ *ik*$_1$ *Q*$_2$ *te*$_3$
[[*ik*]*u*] ‘go’ [[*ik*]*te*] ‘go-gerund’
b.i. 行$_1$ う$_2$ ii. 行$_1$ っ$_2$ て$_3$
okonaw$_1$*u*$_2$ *okonaw*$_1$*Q*$_2$ *te*$_3$
[[*okonaw*]*u*] ‘do’ [[*okonaw*]*te*] ‘do-GERUND’

To sum up, *okurigana* cannot fully disambiguate *kanji* graphemes which violate any of the conditions in (21a-c).

(21) *Kanji* graphemes cannot be lexically disambiguated by *okurigana* if they:
a. denote two or more Sino-Japanese morphemes;

[11] In the examples (19) and (20), morphophonological alternations are omitted in the transcriptions. The symbol *Q* stands for consonant length, which is moraic in Japanese.

b. denote two or more native nominal morphemes (excluding nominalised items and other exceptions);
c. denote one Sino-Japanese morpheme and one or more native nominal morphemes (excluding nominalised items and other exceptions); or
d. produce two or more homographic representations even when accompanied by *okurigana*.

The existence of these restrictions indicate that *okurigana* cannot always carry out the lexical disambiguate function. The question now arises: How effective is the lexical disambiguating function of *okurigana*, or, in other words, what is the amount of homographic representations actually blocked by this function? In the next section, this point will be discussed on the basis of empirical data obtained from a survey of *kanji* graphemes that are used today.

3 Survey

3.1 Aim

The author conducted a survey to measure the effectiveness of the lexical disambiguating function with regards the *kanji* graphemes that are currently in use. The survey was aimed to estimate the relative proportion of homographic *kanji* that are feasible to lexical disambiguation. For this purpose, it was necessary to sort out, from a set of commonly used *kanji*, only those graphemes which would not violate any of the four restrictions (21a-d) formulated in section 2.3.

3.2 Material and Method

The *kanji* graphemes and the associated morphemes were taken from *Jōyōkanji-hyō* (Cabinet 1981a), an inventory of officially approved 1,945 *kanji* graphemes known as the *jōyōkanji*. The inventory defines the form of each grapheme and the way in which it should be used, together with examples and exceptional usages. *Jōyōkanji-hyō* is by no means an exhaustive list of all *kanji* graphemes that are used today. However, it is a de facto standard adopted in various types of publication, just

as *Okurigana no Tsukekata* (Cabinet 1981b) is for *okurigana* orthography. It is therefore appropriate to regard the *jōyōkanji* as constituting a representative set of the current *kanji* graphemes.

As Nomura (1981) mentions, *Jōyōkanji-hyō* (Cabinet 1981a) contains 696 graphemes denoting only one morpheme, be it native or Sino-Japanese. These graphemes were identified in a preliminary survey (Table 1). They were irrelevant to the present survey because they relate to individual morphemes through biunique mappings, and are thus do not produce homography-induced ambiguity at the lexical level. Leaving these out, 1,249 graphemes were included in the 'check table'.

Table 1 *Jōyōkanji* Graphemes Denoting Only One Morpheme

# of SJ	# of NJ	# of Kanji	Kanji
0	1	32	扱芋虞蚊貝垣潟且株刈繰崎咲皿芝杉瀬滝但棚塚坪峠箱肌姫堀又岬娘匁枠
1	0	664	亜愛圧案以医委威為胃尉意維緯域壱逸姻員院韻宇英衛液駅悦謁閲宴援演王凹央応往欧翁億憶乙恩可佳科菓貨禍寡箇課賀雅餓介拐界械階劾害涯慨該概拡核郭較閣嚇穫括活喝褐轄刊缶完官看勘喚敢棺款閑寛感漢歓監憾還館環簡観艦鑑頑岐希汽奇季紀軌規揮棋棄騎宜義儀擬犠議菊喫却旧級糾給巨距凶享協況峡局斤均菌禁緊吟銀区句具偶遇屈訓勲軍郡刑系径啓渓景慶警芸劇傑件券県倹圏検憲謙顕玄孤弧個午呉娯碁護孔后坑孝抗拘肯侯恒洪郊校航康項鉱酵稿衡講購号拷剛豪克穀酷獄昆婚紺墾佐査詐才宰栽斎債材剤昨索策錯察桟算賛暫士司史祉肢師視詞嗣詩資誌滋磁璽式識軸疾舎赦邪勺尺釈爵朱珠需儒樹囚週酬銃叔淑粛塾術俊旬准殉純循順準遵処庶署諸序叙徐匠抄肖尚昭将症祥称渉章紹訟掌晶硝粧証奨彰衝賞礁冗条状浄剰壌嬢錠嘱職信娠紳審迅陣帥粋睡随髄枢崇寸是斉制征牲聖製税斥析隻席績籍拙窃摂仙宣栓旋践銑線遷繊禅漸祖租措塑壮荘曹創僧層総槽燥像臓即則俗族属賊卒他妥堕惰駄胎泰逮隊態第題宅択卓拓託濯諾達丹単胆誕段談痴稚畜逐秩窒嫡宙忠抽衷駐貯庁帳脹腸徴勅朕陳賃墜呈廷抵邸亭貞帝訂逓停偵艇適迭哲鉄徹撤典点展電斗徒途奴到党陶塔搭痘糖謄騰胴堂銅匿特督徳篤毒凸屯弐肉尿妊寧念能脳農把派覇婆肺俳排輩倍陪媒賠伯舶漠爆伐閥版班畔般販搬頒範繁藩晩番蛮盤妃批披非碑罷微百票評標秒賓頻敏瓶扶府附婦符普膚賦譜部服副復福複雰墳丙陛塀幣弊遍弁勉舗簿邦胞俸砲肪某剖帽棒貿朴僕撲没奔盆摩魔毎枚膜抹慢漫未魅密脈妙盟銘盲猛紋厄約愉輸癒幽悠郵猶裕融予洋容庸陽擁曜翌羅酪覧濫欄吏理痢陸略隆硫虜慮了両料猟僚領寮療厘倫累塁類令零隷齢歴列烈廉錬炉労郎浪廊楼録論湾
Total:		696	

(Abbreviations: SJ = Sino-Japanese; NJ = native)

The check table was developed from Tamaoka et al. (2000), a computer-legible database prepared in Microsoft Excel 2000, which can be downloaded free of charge from the Oxford Text Archive (http://ota.ahds.ac.uk/). This database describes various properties of the *jōyōkanji* graphemes using 27 variables, such as the form,

associated readings, their English translation, and so forth. The author downloaded the database in June 2007, and extracted from it the following data: (i) the graphemes; (ii) the number of native readings; (iii) the number of Sino-Japanese readings; (iv) the native readings in romanised transliteration; and (v) the Sino-Japanese readings in romanised transliteraion. In the original database, parentheses are used to indicate the portion of *kanji* readings which are covered by *okurigana*. This information was particularly useful for the present survey.

The accuracy of the data was checked against hardcopy databases provided by Nomura (1981) and Kai and Shinozaki (2006). It was discovered that Tamaoka et al. (2000) incorrectly described the grapheme 気 as having one native and two Sino-Japanese morphemes. However, the database did not show any native item fitting this description. In fact, *Jōyōkanji-hyō* (Cabinet 1981a) assigns only Sino-Japanese morphemes, namely *ki* and *ke*. Apart from some typographic errors, this was the only flaw and was corrected accordingly by the author. The data were then reordered according to the number of Sino-Japanese morphemes and that of native morphemes.

Using the check table, the author checked each of the 1,249 graphemes against the four restrictions (21a-d), marking and counting every violation. For the restriction (21d), *kanji* graphemes which would always produce identical orthographic forms (marked as *Lex*(*ical*)) were distinguished from those which would become homographic only in certain word-forms (*W*(*ord*)-*F*(*orm*)). Both cases were assumed to be violating the restriction (21d). The checking covered all and only those items cited in *Jōyōkanji-hyō* (Cabinet 1981a); although a number of readings are marked as 'special or of limited use' in the inventory, they were simply treated as separate items on a par with their unmarked counterparts.[12]

The graphemes which did not commit any violation were identified as feasible for lexical disambiguation through *okurigana*. The total number of these graphemes was

[12] To cite an instance, *Jōyōkanji-hyō* (Cabinet 1981a) assigns two distinct native items to the grapheme 雨, namely *ame* and *ama*. Linguistically, the latter is an allomorph of the former, which can occur only as a prefix (e.g., *ame* + *oto* → *amaoto* 'sound of rain').

divided by the total number of the population (i.e., 1,249) to calculate their relative proportion within the entire set.

3.3 Result and Discussion

As a result of the survey, 569 graphemes were identified as feasible for lexical disambiguation (for details, see Appendix: Survey Result). They amounted for 45.5% of the homographic *jōyōkanji* graphemes. At present, there are no independent grounds to evaluate these figures as large or small. Still, the result clearly demonstrates that *okurigana* remove homography-induced ambiguity in less than a half of homographic *jōyōkanji*. The implication is that the disambiguating function is far less capable of "usually" reducing "the choice of possible interpretations of the preceding character to one" as Coulmas (2003: 182) claims. Rather, it only provides a safeguard for the reader to recheck his or her decision about the intended reading for a limited number of *kanji* graphemes.

There were 755 cases of violation, with some graphemes committing double or triple violations.[13] Of these, 728 (96.4%) were violations against the restrictions (21a), (21b) and/or (21c). This is significant because these three restrictions all stem out of the more fundamental constraints: no *okurigana* for Sino-Japanese items (13) and non-derivative native nouns (14). Interestingly, these two restrictions have been in place for centuries. The survey revealed that it is these traditional conventions which impose a tight constraint on the effectiveness of the lexical disambiguating function.

4 Conclusion

This paper has discussed the effectiveness of *okurigana* as devices for disambiguating homographic *kanji*. It was shown that *okurigana* disambiguate *kanji* graphemes at two levels, namely the word-form level and the lexical level. It was pointed out that unlike word-form disambiguation, lexical disambiguation is subject to several strong restrictions. On the basis of a corroborative survey, it was

[13] Note that this should not be confused with the number of graphemes committing violations.

demonstrated that the effectiveness of the lexical disambiguating functions is severely limited by these restrictions.

To conclude, *okurigana* do not provide a radical solution to the problem of homography-induced ambiguity at the lexical level. Even though they do remove a certain amount of ambiguity, they by no means bring about a large-scale reduction of the problem. As a result, *kanji* remain largely homographic, keeping the Japanese writing system highly ambiguous, and thus limiting the system's self-sufficiency to a considerable degree.

Acknowledgements

The author is greatly obliged to Prof. Stefan Kaiser for his valuable comments and suggestions, without which the present paper would have never been completed. Special thanks go to Prof. Tadayuki Yuzawa and Assoc. Prof. Jun Ikeda, who have spared much of their precious time to discuss with the author many of the issues explored here. It is a pleasure to also acknowledge Dr. Ruth Vanbaelen and Sunthara Teerawut, who have been generous enough to accommodate the author's very demanding requests. Finally, heartfelt thanks are due to fellow students and many friends who have kindly provided many insightful suggestions.

References

Abe, Seiya (1989) Jōyōkanji no okurigana. In: Kiyoji Satō (ed.) *Kanji to Kokugo Mondai* (*Kanji Kōza*, 11) Tokyo: Meiji Shoin.

Cabinet (1981a) Jōyōkanji-hyō: Cabinet notification No. 1, October 1, 1981. *Kanpō: Gōgai* October 1 edition, 1981. Tokyo: Ōkurashō Insatsukyoku.

Cabinet (1981b) Okurigana no Tsukekata: Cabinet notification No. 3, October 1, 1981. *Kanpō: Gōgai* October 1 edition, 1981. Tokyo: Ōkurashō Insatsukyoku.

Coulmas, Florian (2003) *Writing systems: An introduction to their linguistic analysis* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Hill, Archibald A. (1967) The typology of writing systems. In: William M. Austin (ed.) *Papers in Linguistics in Honor of Léon Dostert*, 92-99. The Hague: Mouton.

Kai, Mutsurō and Yoshiko Shinozaki (2006) Shiryō "Jōyōkanji-hyō no onkun" ni tsuite. *Nihongogaku* 25 (11): 245-282.

Kaiser, Stefan (1995) Sekai no moji, Chūgoku no moji, Nihon no moji: Kanji no ichizuke saikō. *Sekai no Nihongo Kyōiku* 5: 155-167.

Kamei, Takashi, Rokurō Kōno and Eiichi Chino (eds.) (1996) *Jutsugohen* (Gengogaku Daijiten, 6) Tokyo: Sanseidō.

Kondō, Takako (2005) Kanji to okurigana. In: Tomiyoshi Maeda and Masaki Nomura (eds.) *Kanji to Nihongo*, 146-163. Tokyo: Asakura Shoten.

Miyajima, Tatsuo, Masaaki Nomura, Kiyoshi Egawa, Hiroshi Nakano, Sinji Sanada and Hideo Satake (eds.) (1987) *Zusetsu Nihongo: Gurafu de miru kotoba no sugata*. Second edition. Tokyo: Kadokawa Shoten.

Nomura, Masaaki (1981) Jōyōkanji no onkun. *Keiryō Kokugogaku* 13 (1): 27-33.

Rogers, Henry (2005) *Writing systems: A linguistic approach*. Malden, MA: Blackwell Publishing.

Sampson, Geoffrey (1985) *Writing systems*. Stanford, CA: Stanford University Press.

Sproat, Richard (2000) *A computational theory of writing systems* (Studies in Natural Language Processing) Cambridge: Cambridge University Press.

Tamaoka, Katsuo, Kim Kirsner, Yūshi Yanase, Yayoi Miyaoka and Masahiro Kawakami (2000) *Database for the 1,945 basic Japanese kanji (J-1945D)*. http://ota.ahds.ac.uk (produced on May 1, 2000).

Tsukishima, Hiroshi (1970) Gendaigo no seishohō ni tsuite: "Tōyōkanji kaitei onkun-hyō" oyobi "Kaitei okurigana no tsukekata" o chūshin ni. *Gengo Seikatsu* 228: 18-26.

Tsukishima, Hiroshi and Tadao Kabashima (1984) Okurigana. In: Kokugo Gakkai (ed.) *Kokugogaku Daijiten*. Fourth edition. 88-89. Tokyo: Tōkyōdō Shuppan.

Vance, Timothy J. (2002) The exception that proves the rule: Ideography and Japanese kun'yomi. In: Mary S. Erbaugh (ed.) *Difficult Characters: Interdisciplinary Studies of Chinese and Japanese Writing*, 177-193. Columbus, OH: National East Asian Language Resource Center, Ohio State University.

Appendix: Survey Result

Abbriviations: ID = identification number; JKH ID = ID in *Jōyōkanji-hyō* (Cabinet 1981a); 2+SJ = 2 or more Sino-Japanese; 2+NJ N = 2 or more native nouns; 1SJ&1NJ N = 1 Sino-Japanese and 1 native noun; Lex = lexical; W-F = word-form. **Note:** Native nouns do not include nominalised items and other exceptions.

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography	
						Lex	W-F
1	2	哀					
2	4	悪	x				
3	5	握					
4	8	安					
5	10	暗					
6	12	衣			x		
7	13	位			x		
8	14	囲					
9	16	依	x				
10	18	易	x				
11	23	異			x		
12	24	移					
13	25	偉					
14	27	違					
15	29	慰					
16	30	遺	x				
17	33	育					
18	34	一	x				
19	38	引					
20	39	印			x		
21	40	因					
22	44	陰			x		
23	45	飲					
24	46	隠					
25	48	右	x				
26	50	羽		x	x		
27	51	雨		x	x		
28	52	運					
29	53	雲			x		
30	54	永					
31	55	泳					
32	57	映					
33	58	栄				x	
34	59	営					
35	60	詠					
36	61	影			x		
37	62	鋭					
38	64	役	x				
39	65	疫	x				
40	66	益	x				
41	70	越					
42	73	円					
43	74	延					
44	75	沿					
45	76	炎			x		
46	79	園			x		
47	80	煙			x		
48	81	猿			x		
49	82	遠	x				
50	83	鉛			x		

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography	
						Lex	W-F
51	84	塩			x		
52	86	縁			x		
53	87	汚				x	
54	93	押					
55	95	殿					
56	96	桜			x		
57	98	奥			x		
58	99	横			x		
59	100	屋			x		
60	105	卸					
61	106	音	x	x			
62	108	温					
63	109	穏					
64	110	下	x	x			
65	111	化	x				
66	112	火		x	x		
67	113	加					
68	115	仮	x				
69	116	何		x	x		
70	117	花			x		
71	118	価			x		
72	120	果					
73	121	河			x		
74	123	架					
75	124	夏	x				
76	125	家	x	x			
77	126	荷			x		
78	127	華	x				
79	131	渦			x		
80	132	過					
81	133	嫁			x		
82	134	暇			x		
83	136	靴			x		
84	138	歌			x		
85	140	稼					
86	142	我		x	x		
87	143	画	x				
88	144	芽			x		
89	149	会	x				
90	150	回	x				
91	151	灰			x		
92	152	快					
93	153	戒					
94	154	改					
95	156	怪					
96	158	悔					
97	159	海			x		
98	161	皆			x		
99	163	絵	x				
100	164	開				x	

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography	
						Lex	W-F
101	166	解	x				
102	167	塊			x		
103	168	壊					
104	169	懐			x		
105	170	外	x	x			
106	174	街	x				
107	179	各			x		
108	180	角		x	x		
109	182	革			x		
110	183	格	x				
111	185	殻			x		
112	187	覚					
113	189	隔					
114	191	確					
115	192	獲					
116	195	学					
117	196	岳			x		
118	197	楽	x				
119	198	額			x		
120	199	掛					
121	205	渇					
122	206	割			x		
123	207	滑					
124	212	干					
125	214	甘					
126	215	汗			x		
127	218	肝			x		
128	220	冠			x		
129	221	巻			x		
130	223	陥					
131	224	乾					
132	226	患					
133	227	貫					
134	228	寒					
135	230	堪					
136	231	換					
137	235	間	x	x			
138	237	勧					
139	239	幹			x		
140	242	慣					
141	243	管			x		
142	244	関			x		
143	247	緩					
144	256	丸			x		
145	257	元	x				
146	258	含					
147	259	岸			x		
148	260	岩			x		
149	261	眼	x				
150	263	顔			x		

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography	
						Lex	W-F
151	264	願					
152	265	企					
153	266	危					
154	267	机			x		
155	268	気	x				
156	271	忌					
157	274	祈					
158	278	既					
159	279	帰					
160	280	記					
161	281	起					
162	282	飢					
163	283	鬼			x		
164	284	基		x	x		
165	285	寄					
166	287	喜					
167	288	幾			x		
168	290	期	x				
169	292	貴				x	
170	294	旗			x		
171	295	器			x		
172	296	輝					
173	297	機			x		
174	299	技			x		
175	301	偽			x		
176	302	欺					
177	304	疑					
178	306	戯					
179	311	吉	x				
180	313	詰					x
181	315	客	x				
182	316	脚	x				
183	317	虐					
184	318	逆			x		
185	319	九	x				
186	320	久	x				
187	321	及					
188	322	弓			x		
189	323	丘			x		
190	325	休					
191	326	吸					
192	327	朽					
193	328	求					
194	329	究					
195	330	泣					
196	331	急					
197	334	宮	x				
198	335	救					
199	336	球			x		
200	338	窮					

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
201	339	牛			x		
202	340	去	x				
203	342	居					
204	343	拒					
205	344	拠	x				
206	345	挙					
207	346	虚	x				
208	347	許					
209	349	魚		x	x		
210	350	御	x				
211	351	漁	x				
212	353	共			x		
213	354	叫					
214	355	狂					
215	357	京	x				
216	358	供	x				
217	362	挟					
218	363	狭					
219	364	恐					
220	365	恭					
221	366	胸		x	x		
222	367	脅				x	
223	368	強	x				
224	369	教					
225	370	郷	x				
226	371	境	x				
227	372	橋			x		
228	373	矯					
229	374	鏡			x		
230	375	競	x				
231	376	響					
232	377	驚					
233	378	仰	x				
234	379	暁			x		
235	380	業	x				
236	381	凝					
237	382	曲					
238	384	極	x				
239	385	玉			x		
240	388	近					
241	389	金	x	x			
242	391	勤	x				
243	392	琴			x		
244	393	筋			x		
245	396	謹					
246	397	襟			x		
247	402	苦					
248	403	駆					
249	405	愚					
250	406	空		x	x		
251	409	隅			x		
252	411	掘					
253	413	君			x		
254	416	薫					
255	419	群			x		
256	420	兄	x				
257	422	形	x	x			
258	425	茎			x		
259	426	係			x		
260	427	型			x		
261	428	契					
262	429	計					
263	430	恵	x				
264	432	掲					
265	434	経	x				
266	435	蛍			x		
267	436	敬					
268	438	軽					
269	439	傾					
270	440	携					
271	441	継					
272	443	憩					
273	445	鶏			x		
274	447	迎					
275	448	鯨			x		
276	450	撃					
277	451	激					
278	452	欠					
279	453	穴			x		
280	454	血			x		
281	455	決					
282	456	結					
283	458	潔					
284	459	月	x				
285	460	犬			x		
286	462	見					
287	464	肩			x		
288	465	建	x				
289	466	研					
290	469	兼					
291	470	剣			x		
292	471	軒			x		
293	472	健					
294	473	険					
295	475	堅					
296	477	嫌	x				
297	478	献	x				
298	479	絹			x		
299	480	遣					
300	481	権	x				
301	483	賢					
302	485	繭			x		
303	487	験	x				
304	488	懸	x				
305	489	幻			x		
306	491	言	x				
307	492	弦			x		
308	493	限					
309	494	原			x		
310	495	現					
311	496	減					
312	497	源			x		
313	498	厳	x				
314	499	己	x				
315	500	戸			x		
316	501	古					
317	502	呼					
318	503	固					
319	506	故			x		
320	507	枯					
321	509	庫	x				
322	510	湖			x		
323	511	雇					
324	512	誇					
325	513	鼓			x		
326	514	顧					
327	515	五			x		
328	516	互					
329	519	後	x	x			
330	521	悟					
331	523	語					
332	524	誤					
333	526	口	x				
334	527	工	x				
335	528	公			x		
336	530	功	x				
337	531	巧					
338	532	広					
339	533	甲	x				
340	534	交					
341	535	光			x		
342	536	向					x
343	538	好					
344	539	江			x		
345	540	考					
346	541	行	x			x	x
347	545	攻					
348	546	更			x		
349	547	効					
350	548	幸			x		
351	552	厚					
352	555	皇	x				
353	556	紅	x	x			
354	557	荒					
355	559	香	x				
356	560	候			x		
357	562	耕					
358	564	貢	x				
359	565	降					
360	566	高			x		
361	568	控					
362	569	黄	x	x			
363	570	慌					
364	571	港			x		
365	572	硬					
366	573	絞					
367	575	溝			x		
368	577	構					
369	578	綱			x		
370	581	興	x				
371	583	鋼			x		
372	587	合	x				
373	592	告					
374	593	谷			x		
375	594	刻					
376	595	国			x		
377	596	黒			x		
378	600	骨			x		
379	601	込					
380	602	今	x				
381	603	困					
382	605	恨					
383	606	根			x		
384	608	混					
385	610	魂			x		
386	612	懇					
387	613	左			x		
388	616	砂	x				
389	617	唆					
390	618	差					
391	620	鎖			x		
392	621	座					
393	623	再	x				
394	624	災					
395	625	妻			x		
396	626	砕					
397	629	彩					
398	630	採					
399	631	済					
400	632	祭					

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
401	634	細					
402	635	菜			x		
403	636	最					
404	637	裁					
405	639	催					
406	640	歳	x				
407	641	載					
408	642	際			x		
409	643	在					
410	646	財	x				
411	647	罪			x		
412	649	作	x				
413	650	削					
414	655	酢			x		
415	656	搾					
416	658	冊	x				
417	659	札			x		
418	660	刷					
419	661	殺	x				
420	663	撮					
421	664	擦					
422	665	雑	x				
423	667	三			x	x	
424	668	山			x		
425	669	参					
426	671	蚕			x		
427	672	惨	x				
428	673	産			x		
429	674	傘			x		
430	675	散					
431	677	酸					
432	679	残					
433	682	子	x				
434	683	支					
435	684	止					
436	685	氏			x		
437	686	仕	x				
438	689	四		x	x	x	
439	690	市			x		
440	691	矢			x		
441	692	旨			x		
442	693	死					
443	694	糸			x		
444	695	至					
445	696	伺					
446	697	志			x		
447	698	私			x		
448	699	使					
449	700	刺					
450	701	始					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
451	702	姉			x		
452	703	枝			x		
453	706	姿			x		
454	707	思					
455	708	指			x		
456	709	施	x				
457	711	紙			x		
458	712	脂			x		
459	713	紫			x		
460	716	雀			x		
461	719	試					
462	721	雌		x	x		
463	722	飼					
464	724	賜					
465	725	諮					
466	726	示	x				
467	727	字			x		
468	728	寺			x		
469	729	次	x				
470	730	耳			x		
471	731	自	x				
472	732	似					
473	733	児	x				
474	734	事	x				
475	735	侍			x		
476	736	持					
477	737	時			x		
478	739	慈					
479	740	辞					
480	746	七		x	x		
481	747	失					
482	748	室			x		
483	750	執	x				
484	751	湿					
485	752	漆			x		
486	753	質	x				
487	754	実			x		
488	756	写					
489	757	社			x		
490	758	車			x		
491	760	者			x		
492	761	射					
493	762	捨					
494	764	斜					
495	765	煮					
496	766	遮					
497	767	謝					
498	769	蛇	x				
499	772	借					
500	773	酌					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
501	776	若	x				
502	777	弱					
503	778	寂	x				
504	779	手		x	x		
505	780	主	x	x			
506	781	守	x				x
507	783	取					
508	784	狩					
509	785	首			x		
510	786	殊			x		
511	788	酒		x	x		
512	789	種			x		
513	790	趣			x		
514	791	寿			x		
515	792	受					
516	793	授					
517	797	収					
518	799	州			x		
519	800	舟		x	x		
520	801	秀					
521	802	周					
522	803	宗	x				
523	804	拾	x				
524	805	秋			x		
525	806	臭					
526	807	修	x				
527	808	終					
528	809	習					
529	811	就	x				
530	812	衆	x				
531	813	集					
532	814	愁					
533	816	醜					
534	817	襲					
535	818	十	x	x			
536	819	汁			x		
537	820	充					
538	821	住					
539	822	柔	x				
540	823	重	x				
541	824	従	x				
542	825	渋			x		
543	827	獣			x		
544	828	縦			x		
545	830	祝	x				
546	831	宿			x		
547	834	縮					
548	836	熟					
549	837	出	x				x
550	838	述					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
551	841	奉			x		
552	842	瞬					
553	844	巡					
554	845	盾			x		
555	852	潤					
556	855	初		x	x		x
557	856	所			x		
558	857	書					
559	859	暑					
560	861	緒	x				
561	863	女	x	x			
562	864	如	x				
563	865	助			x		
564	869	除	x				
565	870	小		x	x		
566	871	升			x		
567	872	少					
568	873	召					
569	875	床		x	x		
570	879	承					
571	880	招					
572	881	昇					
573	882	松			x		
574	883	沼			x		
575	885	宵			x		
576	887	消					
577	891	笑					
578	892	商					
579	893	唱					
580	898	勝					x
581	901	焼					
582	902	焦					
583	906	詔			x		
584	907	象	x				
585	908	傷			x		
586	910	照					
587	911	詳					
588	913	障					
589	916	償					
590	918	鐘			x		
591	919	上	x	x			
592	920	丈			x		
593	924	乗					
594	925	城			x		
595	928	常		x	x		
596	929	情	x				
597	930	場			x		
598	931	畳			x		
599	932	蒸					
600	933	縄			x		

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
601	937	譲					
602	938	醸					
603	939	色	x				
604	940	食	x				
605	941	植					
606	942	殖					
607	943	飾					
608	944	触					
609	946	織	x				
610	948	辱					
611	949	心			x		
612	950	申					
613	951	伸					
614	952	臣	x				
615	953	身			x		
616	954	辛					
617	956	侵					
618	957	津			x		
619	958	神	x	x			
620	960	振					
621	961	浸					
622	962	真			x		
623	963	唇			x		
624	964	針			x		
625	965	深					
626	967	進					
627	968	森			x		
628	969	診					
629	970	寝					
630	971	慎					
631	972	新			x		
632	974	震					
633	975	薪			x		
634	976	親			x		
635	977	人	x				
636	978	刃			x		
637	979	仁	x				
638	980	尽					
639	982	甚					
640	984	尋					
641	985	図	x				
642	986	水			x		
643	987	吹					
644	988	垂					
645	989	炊					
646	992	衰					
647	993	推					
648	994	酔					
649	995	遂					
650	997	穂			x		

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
651	998	錘			x		
652	1003	数	x				
653	1004	据					
654	1007	畝		x			
655	1010	井	x				
656	1011	世	x				
657	1012	正	x				
658	1013	生	x	x			
659	1014	成	x				
660	1015	西	x				
661	1016	声	x	x			
662	1019	姓	x				
663	1021	性	x				
664	1022	青	x				
665	1023	政	x				
666	1024	星	x				
667	1026	省	x				
668	1027	逝					
669	1028	清	x				
670	1029	盛	x			x	
671	1030	婿			x		
672	1031	晴					
673	1032	勢					
674	1034	誠			x		
675	1035	精	x				
676	1037	誓					
677	1038	静	x				
678	1039	請	x				
679	1040	整					
680	1042	夕			x		
681	1044	石	x				
682	1045	赤	x				
683	1046	昔	x				
684	1050	惜					
685	1051	責					
686	1052	跡			x		
687	1053	積					
688	1056	切	x				
689	1057	折			x		
690	1060	接					
691	1061	設					
692	1062	雪			x		
693	1064	節	x				
694	1065	説	x				
695	1066	舌			x		
696	1067	絶					
697	1068	千			x		
698	1069	川			x		
699	1071	占					
700	1072	先			x		

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
701	1074	専					
702	1075	巣			x		
703	1076	浅					
704	1077	洗					
705	1078	染					
706	1079	扇			x		
707	1082	船		x	x		
708	1083	戦			x		
709	1085	銭			x		
710	1087	潜					
711	1089	選					
712	1091	薦					
713	1093	鮮					
714	1094	全					
715	1095	前			x		
716	1096	善					
717	1097	然	x				
718	1100	禱					
719	1101	阻					
720	1104	素	x				
721	1106	粗					
722	1107	組			x		
723	1108	疎					
724	1109	訴					
725	1111	礎			x		
726	1112	双			x		
727	1114	早	x				
728	1115	争					
729	1116	走					
730	1117	奏					
731	1118	相	x				
732	1119	草			x		
733	1121	送					
734	1122	倉			x		
735	1123	捜					
736	1124	挿					
737	1125	桑			x		
738	1126	掃					
739	1128	巣			x		
740	1129	窓			x		
741	1131	喪			x		
742	1132	葬					
743	1133	装	x				
744	1135	想	x				
745	1138	遭					
746	1140	操			x		
747	1142	霜			x		
748	1143	騒					
749	1144	藻			x		
750	1145	造					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
751	1147	増					
752	1148	憎					
753	1149	蔵			x		
754	1150	贈	x				
755	1153	束			x		
756	1154	足			x		
757	1155	促					
758	1157	息			x		
759	1158	速					
760	1159	側			x		
761	1160	測					
762	1165	続					
763	1167	率	x				
764	1168	存	x				
765	1169	村			x		
766	1170	孫			x		
767	1171	尊				x	
768	1172	損					
769	1174	多					
770	1175	打					
771	1180	太	x				
772	1181	体	x				
773	1182	対	x				
774	1183	耐					
775	1184	待					
776	1185	怠					
777	1187	退					
778	1188	帯			x		
779	1190	袋			x		
780	1192	替					
781	1194	貸					
782	1195	滞					
783	1197	大	x				
784	1198	代	x	x			
785	1199	台	x				
786	1205	沢			x		
787	1211	濁					
788	1214	脱					
789	1215	奪					
790	1218	担					
791	1220	炭			x		
792	1222	探					
793	1223	淡					
794	1224	短					
795	1225	嘆					
796	1226	端		x	x		
797	1228	鍛					
798	1229	団	x				
799	1230	男	x				
800	1232	断					x

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
801	1233	弾			x		
802	1234	暖					
803	1236	壇	x				
804	1237	地	x				
805	1238	池			x		
806	1239	知					
807	1240	治	x				
808	1241	値		x	x		
809	1242	恥			x		
810	1243	致					
811	1244	遅					
812	1247	置					
813	1248	竹			x		
814	1251	蓄					
815	1252	築					
816	1255	茶	x				
817	1256	着	x				
818	1258	中			x		
819	1259	仲			x		
820	1260	虫			x		
821	1261	沖			x		
822	1265	注					
823	1266	昼			x		
824	1267	柱			x		
825	1269	鋳					
826	1271	著					
827	1273	弔					
828	1275	兆					
829	1276	町			x		
830	1277	長					
831	1278	挑					
832	1280	張					
833	1281	彫					
834	1282	眺					
835	1283	頂			x		
836	1284	釣					
837	1285	鳥			x		
838	1286	朝			x		
839	1288	超					
840	1290	跳					
841	1292	澄					
842	1293	潮			x		
843	1294	調					
844	1295	聴					
845	1296	懲					
846	1297	直	x				
847	1299	沈					
848	1300	珍					
849	1304	鎮					
850	1305	追					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
851	1307	通	x				x
852	1308	痛					
853	1310	漬					
854	1312	丁	x				
855	1313	低					
856	1316	弟	x				
857	1317	定	x				
858	1318	底			x		
859	1325	庭			x		
860	1329	堤			x		
861	1330	提					
862	1331	程			x		
863	1333	締					
864	1334	泥			x		
865	1335	的			x		
866	1336	笛			x		
867	1337	摘					
868	1338	滴			x		
869	1340	敵			x		
870	1346	天		x	x		
871	1348	店			x		
872	1351	添					
873	1352	転					
874	1353	田			x		
875	1354	伝					
876	1355	殿	x	x			
877	1358	吐					
878	1361	都	x				
879	1362	渡					
880	1363	塗					
881	1364	土	x				
882	1366	努					
883	1367	度	x				
884	1368	怒				x	
885	1369	刀			x		
886	1370	冬			x		
887	1371	灯			x		
888	1372	当					
889	1373	投					
890	1374	豆	x				
891	1375	東			x		
892	1377	逃					
893	1378	倒					
894	1380	凍					
895	1381	唐			x		
896	1382	島			x		
897	1383	桃			x		
898	1384	討					
899	1385	透					
900	1386	悼					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
901	1387	盗					
902	1391	棟		x	x		
903	1392	湯			x		
904	1394	登	x				
905	1395	等					
906	1396	答					
907	1397	筒			x		
908	1398	統					
909	1399	稲		x	x		
910	1400	踏					
911	1402	頭	x	x			
912	1404	闘					
913	1406	同					
914	1407	洞			x		
915	1409	動					
916	1411	童			x		
917	1412	道	x				
918	1413	働					
919	1415	導					
920	1419	得					
921	1424	独					
922	1425	読	x				
923	1427	突					
924	1428	届					
925	1430	豚			x		
926	1431	鈍					
927	1432	曇					
928	1433	内	x				
929	1434	南	x				
930	1435	軟					
931	1436	難					
932	1437	二			x		
933	1438	尼			x		
934	1441	日	x	x			
935	1442	入				x	
936	1443	乳		x	x		
937	1445	任					
938	1447	忍					
939	1448	認					
940	1450	熱					
941	1451	年			x		
942	1453	粘					
943	1454	燃					
944	1455	悩					
945	1456	納	x				
946	1460	濃					
947	1462	波			x		
948	1464	破					
949	1466	馬		x	x		
950	1468	拝					

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
951	1469	杯			x		
952	1470	背		x	x		
953	1473	配					
954	1475	敗					
955	1476	廃					
956	1478	売					
957	1480	梅			x		
958	1481	培					
959	1484	買					
960	1486	白	x	x			
961	1488	迫					
962	1489	拍	x				
963	1490	泊					
964	1492	博	x				
965	1493	薄					
966	1494	麦			x		
967	1496	縛					
968	1499	畑		x			
969	1501	八		x	x	x	
970	1502	鉢	x				
971	1503	発	x				
972	1504	髪			x		
973	1506	抜					
974	1507	罰	x				
975	1509	反	x				
976	1510	半					
977	1511	犯					
978	1512	帆			x		
979	1513	伴	x				
980	1514	判	x				
981	1515	坂			x		
982	1516	板	x				
983	1522	飯			x		
984	1524	煩	x				
985	1533	比					
986	1534	皮			x		
987	1536	否			x		
988	1538	彼		x	x		
989	1540	肥			x		
990	1542	卑					
991	1543	飛					
992	1544	疲					
993	1545	秘					
994	1546	被					
995	1547	悲					
996	1548	扉			x		
997	1549	費					
998	1552	避					
999	1553	尾			x		
1000	1554	美					

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
1001	1555	備			x		
1002	1557	鼻			x		
1003	1558	匹			x		
1004	1559	必					
1005	1560	泌	x				
1006	1561	筆			x		
1007	1564	氷		x	x		
1008	1565	表			x		
1009	1566	俵			x		
1010	1569	漂					
1011	1571	苗		x	x		
1012	1573	病	x				
1013	1574	描					
1014	1575	猫			x		
1015	1576	品			x		
1016	1577	浜			x		
1017	1578	貧	x				
1018	1583	不	x				
1019	1584	夫	x				
1020	1585	父			x		
1021	1586	付					
1022	1587	布			x		
1023	1590	怖					
1024	1592	負					
1025	1593	赴					
1026	1594	浮					
1027	1597	富	x				
1028	1599	腐					
1029	1600	敷					
1030	1604	侮					
1031	1605	武	x				
1032	1607	舞			x		
1033	1608	封	x				
1034	1609	風	x	x			
1035	1610	伏					
1036	1613	幅			x		
1037	1616	腹			x		
1038	1618	覆					x
1039	1619	払					
1040	1620	沸					
1041	1621	仏			x		
1042	1622	物	x				
1043	1623	粉		x	x		
1044	1624	紛					
1045	1626	噴					
1046	1628	憤					
1047	1629	奮					
1048	1630	分	x				
1049	1631	文	x				
1050	1632	聞	x				
1051	1634	平	x				
1052	1635	兵	x				
1053	1636	併					
1054	1637	並			x		
1055	1638	柄		x	x		
1056	1640	閉					
1057	1644	米	x				
1058	1645	壁			x		
1059	1646	癖			x		
1060	1647	別					
1061	1648	片			x		
1062	1649	辺			x		
1063	1650	返					
1064	1651	変					
1065	1652	偏					
1066	1654	編					
1067	1656	便	x				
1068	1658	歩	x				
1069	1659	保					
1070	1660	捕					
1071	1661	浦			x		
1072	1662	補					
1073	1664	母			x		
1074	1665	募					
1075	1666	墓			x		
1076	1667	慕					
1077	1668	暮					
1078	1670	方			x		
1079	1671	包					
1080	1672	芳					
1081	1674	奉	x				
1082	1675	宝			x		
1083	1676	抱				x	
1084	1677	放					
1085	1678	法	x				
1086	1679	泡			x		
1087	1681	倣					
1088	1683	峰			x		
1089	1685	崩					
1090	1686	訪					
1091	1687	報					
1092	1688	豊					
1093	1689	飽					
1094	1690	褒					
1095	1691	縫					
1096	1692	亡	x				
1097	1693	乏					
1098	1694	忙					
1099	1695	坊	x				
1100	1696	妨					
1101	1697	忘					
1102	1698	防					
1103	1699	房			x		
1104	1702	冒					
1105	1704	紡					
1106	1705	望	x				
1107	1706	傍					
1108	1710	暴	x				
1109	1711	膨					
1110	1712	謀	x				
1111	1713	北			x		
1112	1714	木	x	x			
1113	1716	牧			x		
1114	1718	墨			x		
1115	1722	本			x		
1116	1724	翻					
1117	1725	凡	x				
1118	1727	麻			x		
1119	1729	磨					
1120	1732	妹			x		
1121	1734	埋					
1122	1735	幕	x				
1123	1738	末	x				
1124	1740	万	x				
1125	1741	満					
1126	1745	味			x		
1127	1751	民			x		
1128	1752	眠					
1129	1753	矛			x		
1130	1754	務					
1131	1755	無	x				
1132	1756	夢			x		
1133	1757	霧			x		
1134	1759	名	x				
1135	1760	命	x				
1136	1761	明	x				
1137	1762	迷					
1138	1765	鳴					
1139	1766	滅					
1140	1767	免					
1141	1768	面		x	x		
1142	1769	綿			x		
1143	1770	茂					
1144	1771	模	x				
1145	1772	毛			x		
1146	1773	妄	x				
1147	1775	耗	x				
1148	1777	網			x		
1149	1778	目	x	x			
1150	1779	黙					
1151	1780	門			x		
1152	1782	問			x		
1153	1784	夜		x	x		
1154	1785	野			x		
1155	1788	訳			x		
1156	1789	薬			x		
1157	1790	躍					
1158	1791	由	x				
1159	1792	油			x		
1160	1794	論					
1161	1797	唯	x				
1162	1798	友			x		
1163	1799	有	x				
1164	1800	勇					
1165	1806	遊	x				
1166	1807	雄		x	x		
1167	1808	誘					
1168	1809	憂					x
1169	1811	優					
1170	1812	与					
1171	1814	余					
1172	1815	誉					
1173	1816	預					
1174	1817	幼					
1175	1818	用					
1176	1819	羊			x		
1177	1821	要					
1178	1824	揚					
1179	1825	揺					
1180	1826	葉			x		
1181	1828	溶					
1182	1829	腰			x		
1183	1830	様			x		
1184	1831	踊					
1185	1832	窯			x		
1186	1833	養					
1187	1835	謡			x		
1188	1837	抑					
1189	1838	浴					
1190	1839	欲					
1191	1841	翼			x		
1192	1842	裸			x		
1193	1844	来				x	
1194	1845	雷			x		
1195	1846	頼					
1196	1847	絡					
1197	1848	落					
1198	1850	乱					
1199	1851	卵			x		
1200	1856	利					

Appendix: Survey Result (*cont.*)

ID	JKH ID	Kanji	2+SJ	2+ NJ N	1SJ& 1NJ N	Homography Lex	Homography W-F
1201	1857	里			x		
1202	1860	裏			x		
1203	1861	履					
1204	1862	離					
1205	1864	立	x				
1206	1865	律	x				
1207	1867	柳			x		
1208	1868	流	x				
1209	1869	留	x				
1210	1870	竜			x		
1211	1871	粒			x		
1212	1874	旅			x		
1213	1879	良					
1214	1881	涼					
1215	1883	陵			x		
1216	1884	量					
1217	1889	糧	x				
1218	1890	力	x				
1219	1891	緑	x				
1220	1892	林			x		
1221	1895	輪			x		
1222	1896	隣			x		
1223	1897	臨					
1224	1898	涙			x		
1225	1903	礼	x				
1226	1904	冷					
1227	1905	励					
1228	1906	戻					
1229	1907	例					
1230	1908	鈴	x				
1231	1910	霊	x				
1232	1913	麗					
1233	1914	暦			x		
1234	1917	劣					
1235	1919	裂					
1236	1920	恋			x		
1237	1921	連					
1238	1923	練					
1239	1926	路			x		
1240	1927	露	x				
1241	1928	老					
1242	1931	朗					
1243	1935	漏					
1244	1936	六		x	x	x	
1245	1939	和	x				
1246	1940	話			x		
1247	1941	賄					
1248	1943	惑					
1249	1945	腕			x		
Number of Violation:			262	66	400	16	11

多読性に由来する漢字の曖昧性と送り仮名の語彙的明確化機能

本田 啓輔

現在の日本語表記体系で用いられる漢字は、一字に対し複数の語彙的形態素が対応する多読性 (i.e., 複数読み) のものが大半を占める。多読性の漢字には、形態素間の差異を明示したり、特定の形態素を適切な読みとして指定したりすることができない。このため、これらの漢字は本来的に曖昧性を有するといえる。また、適切な読みの判定は、読み手の心理言語学的能力に大きく依存すると考えられる。従って、多読性の漢字を多く含む日本語表記体系は、自律性が著しく制限された体系であるといえよう。

一方、先行研究では、送り仮名が多読性の漢字の読みを明確化し、曖昧性の緩和に貢献するとの主張が見られる (Hill 1967、Sampson 1985、Kamei et al. 1996、Vance 2002、Coulmas 2003、Kondō 2005 等)。その骨子は、送り仮名によって語の一部が明示されることで、結果的に適切な読みが指定されるというものである。しかし、送り仮名による明確化機能は十分に研究されているとはいえず、特にその有効性についてはほとんど議論されていない。

本研究では、送り仮名が語形レベルと語彙レベルの両方で漢字を明確化し得ることを示した上で、後者における明確化機能 (語彙的明確化機能) の有効性を実証的に検討する。具体的には、まず語彙的明確化機能の輪郭を、語形レベルにおける明確化との関係に注目しながら描き出す。次に、先行研究ではほとんど注目されてこなかったいくつかの制約を取り上げ、分析する。その上で、常用漢字を対象として行った調査の結果を提示し、送り仮名の語彙的明確化機能の有効性が、それらの制約によって大きく制限されていることを示す。これを踏まえ、送り仮名の語彙的明確化機能が、多読性に由来する漢字の曖昧性に対し、抜本的な解決をもたらすものではないことを論じる。