

# 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.<sup>1</sup> 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.<sup>2</sup>

- (1) 食<sub>1</sub>  
 {*tabe-*, *kuw-*, *kuraw-*, *SYOKU*, *ZIKI*}<sub>1</sub>

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

---

<sup>1</sup> 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.

<sup>2</sup> 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.



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).<sup>4</sup> 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),

---

<sup>4</sup> 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.<sup>5</sup>

## (4) a. Lexical ambiguity

食<sub>1</sub>{*tabe-*, *kuw-*, *kuraw-*, ... }<sub>1</sub>

## b. Word-form ambiguity

食<sub>1</sub>

<i>tabe-ru</i>	<i>kuw-u</i>	<i>kuraw-u</i>	}	1
<i>tabe-nai</i>	<i>kuw-anai</i>	<i>kuraw-anai</i>		
<i>tabe-te</i>	<i>kuw-te</i>	<i>kuraw-te</i>		
...	...	...		
...	...	...		

## 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.

---

<sup>5</sup> 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).<sup>6</sup>

- (5) a. 食<sub>1</sub> べ<sub>2</sub> る<sub>3</sub>      b. 食<sub>1</sub> べ<sub>2</sub> な<sub>3</sub> い<sub>4</sub>  
*tabe<sub>1</sub> be<sub>2</sub> ru<sub>3</sub>*      *tabe<sub>1</sub>be<sub>2</sub> na<sub>3</sub> i<sub>4</sub>*  
 [[[*tabe*]*ru*] ‘eat’      [[[[*tabe*]*na*]*i*] ‘not eat’
- (6) a. 飲<sub>1</sub> む<sub>2</sub>      b. 飲<sub>1</sub> ま<sub>2</sub> な<sub>3</sub> い<sub>4</sub>  
*nom<sub>1</sub> mu<sub>2</sub>*      *nom<sub>1</sub> ma<sub>2</sub> na<sub>3</sub> i<sub>4</sub>*  
 [[[*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.<sup>7</sup> For instance, they spell out the stem-final mora in the V-verb (5a, b), or the combination of the

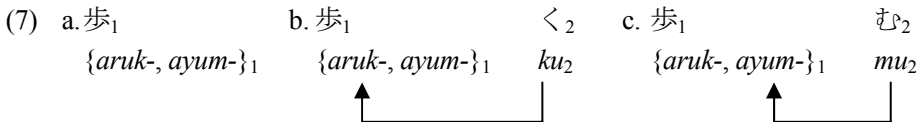
<sup>6</sup> This paper assumes that the usage of *okurigana* follows regulations stipulated by the current version of *Okurigana no Tsukekata* (Cabinet 1981b), a body of official guidelines first promulgated in 1973 and then revised in 1981. Though these regulations have no legal power, they constitute 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.

<sup>7</sup> 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* < *ku* (7b) and む *mu* (7c) suggest the respective verb stems by repeating the final consonant.



With regards to the current orthographic convention stipulated by *Okurigana no Tsume-kata* (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<sub>1</sub> べ<sub>2</sub> る<sub>3</sub>    b. 少<sub>1</sub> な<sub>2</sub> い<sub>3</sub>    c. 穏<sub>1</sub> や<sub>2</sub> か<sub>3</sub> だ<sub>4</sub>  
*tabe<sub>1</sub> be<sub>2</sub> ru<sub>3</sub>*    *sukuna<sub>1</sub>na<sub>2</sub> i<sub>3</sub>*    *odayaka<sub>1</sub>ya<sub>2</sub> ka<sub>3</sub> da<sub>4</sub>*  
 [[*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<sub>2</sub> *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<sub>1</sub> い<sub>2</sub>    b. 遅<sub>1</sub> れ<sub>2</sub> る<sub>3</sub>  
*oso<sub>1</sub> i<sub>2</sub>*    *okure<sub>1</sub> re<sub>2</sub> ru<sub>3</sub>*  
 [[*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 (*ru*), 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<sub>1</sub> ぜ<sub>2</sub> る<sub>3</sub>    b. 混<sub>1</sub> ざ<sub>2</sub> る<sub>3</sub>    c. 混<sub>1</sub> じ<sub>2</sub> る<sub>3</sub>  
*maz(e)<sub>1</sub>ze<sub>2</sub> ru<sub>3</sub>*    *maz(a)<sub>1</sub>za<sub>2</sub> ru<sub>3</sub>*    *maz(i)<sub>1</sub>zi<sub>2</sub> ru<sub>3</sub>*  
 [[[maz]<sub>e</sub>]ru] ‘mix’    [[[maz]<sub>a</sub>]ar]u] ‘mix’    [[[maz]<sub>i</sub>]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).<sup>8</sup>

- (11) a. 但<sub>1</sub> し<sub>2</sub>                      b. 食<sub>1</sub> べ<sub>2</sub> で<sub>3</sub>                      c. 飲<sub>1</sub> み<sub>2</sub>  
*tadasi<sub>1</sub>si<sub>2</sub>*                              *tabe<sub>1</sub> be<sub>2</sub> de<sub>3</sub>*                              *nom<sub>1</sub> mi<sub>2</sub>*  
 [[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<sub>1</sub> べ<sub>2</sub> 物<sub>3</sub>    b. 食<sub>1</sub> い<sub>2</sub> 物<sub>3</sub>    c. 食<sub>1</sub> 物<sub>2</sub>  
*tabe<sub>1</sub> be<sub>2</sub> mono<sub>3</sub>*    *kuw<sub>1</sub> i<sub>2</sub> mono<sub>3</sub>*    *syoku<sub>1</sub> motu<sub>2</sub>*  
 [[[tabe][mono]] ‘food’    [[[kuw]i][mono]] ‘food’    [[syoku][motu]] ‘food’

<sup>8</sup> 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.<sup>9</sup>

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

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

---

<sup>9</sup> 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’.<sup>10</sup> 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<sub>1</sub>

{*waru*-, *AKU*, *O*}<sub>1</sub>

c. i. 悪<sub>1</sub>    𠄎<sub>2</sub>

*waru*<sub>1</sub> *i*<sub>2</sub>

[[*waru*]*i*] ‘bad’

ii. 悪<sub>1</sub>

*AKU*<sub>1</sub>

[*aku*] ‘evil’

iii. 悪<sub>1</sub>

*O*<sub>1</sub>

[*o*] ‘bad’

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

<sup>10</sup> 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<sub>1</sub>

{*ak-*, *aker-*, *sora*, *kara*, *KUU*}<sub>1</sub>

c.i. 空<sub>1</sub>

<<sub>2</sub>

ii. 空<sub>1</sub>

iii. 空<sub>1</sub>

*ak*<sub>1</sub>

*ku*<sub>2</sub>

*sora*<sub>1</sub>

*kara*<sub>1</sub>

[[*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<sub>1</sub>

{*tatakaw-*, *ikusa*, *SEN*}<sub>1</sub>

c.i. 戦<sub>1</sub>

ゝ<sub>2</sub>

ii. 戦<sub>1</sub>

iii. 戦<sub>1</sub>

*tatakaw*<sub>1u2</sub>

*ikusa*<sub>1</sub>

*SEN*<sub>1</sub>

[[*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).<sup>11</sup> 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 <sub>1</sub> < <sub>2</sub>      | ii. | 開 <sub>1</sub> い <sub>2</sub> て <sub>3</sub>      |
|           | ak <sub>1</sub> ku <sub>2</sub>    |     | ak <sub>1</sub> i <sub>2</sub> te <sub>3</sub>    |
|           | [[ak]u] ‘open’                     |     | [[[ak]i]te] ‘open-GERUND’                         |
| b.i.      | 開 <sub>1</sub> <                   | ii. | 開 <sub>1</sub> い <sub>2</sub> て <sub>3</sub>      |
|           | hirak <sub>1</sub> ku <sub>2</sub> |     | hirak <sub>1</sub> i <sub>2</sub> te <sub>3</sub> |
|           | [[hirak]u] ‘open’                  |     | [[[hirak]i]te] ‘open-GERUND’                      |
- 
- |           |                                    |     |  |
|-----------|------------------------------------|-----|--|
| (20) a.i. | 行 <sub>1</sub> < <sub>2</sub>      | ii. | 行 <sub>1</sub> つ <sub>2</sub> て <sub>3</sub>       |
|           | ik <sub>1</sub> ku <sub>2</sub>    |     | ik <sub>1</sub> Q <sub>2</sub> te <sub>3</sub>     |
|           | [[ik]u] ‘go’                       |     | [[ik]te] ‘go-gerund’                               |
| b.i.      | 行 <sub>1</sub> う <sub>2</sub>      | ii. | 行 <sub>1</sub> つ <sub>2</sub> て <sub>3</sub>       |
|           | okonaw <sub>1</sub> u <sub>2</sub> |     | okonaw <sub>1</sub> Q <sub>2</sub> te <sub>3</sub> |
|           | [[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;

<sup>11</sup> 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	扱芋虞蚊貝垣渦且株刈繰崎咲皿芝杉瀬滝但棚塚坪峠箱肌姫堀又岬娘夕杖
1	0	664	<p>           亜愛圧案以医委威為胃尉意維緯域老逸姻員院韻員英衛液馱悅謁閱宴援演王凹央応往欧翁億            憶乙恩可佳科菓貨禍寡箇譜賀雅餓介拐界械階効害涯慨該概拉核郭較閣嚇穫括活喝褐轄刊缶            完官看勘喚敢楛款閑寬感漢欲監憾還館環簡觀艦鑑頭岐希汽奇季紀規規揮棋棄騎宜義儀擬機            議菊喫却旧級糾給巨距凶享協況峽局斤均菌禁繫吟銀区句具偶遇屈訓勳軍郡刑系徑啓溪景慶            警芸劇傑件券具儉圈檢憲謙顯玄孤孤個午吳娛碁護孔后坑孝抗拘肯侯恒洪郊校航康項鉦醇稿            衡講購号拷剛豪克穀酷獄昆婚紺壘佐查許才宰栽齋債材劑咋索策錯察棧算贊暫士司史祉肢師            視詞嗣詩資誌滋磁璽式識軸疾舍赦邪勺尺積廢朱珠需儒樹囚週酬鈿叔淑肅塾術俊句准殉純循            順準遵処庶署諸序叙徐匠抄肖尚昭將症祥称涉章紹訟掌晶硝註証獎彰衝實礁冗条狀淨剩壞壞            錠囑職信娠紳審迅陣帥粹睡隨隨枢崇寸是齊制征性聖製稅斥折隻席績續拙窵撰仙宣栓旋旋銃            線遷織禪漸祖租措塑壯莊曹創僧層綵槽燥像臟即則俗族屬賊卒他妥墮惰駄胎泰逮隊態第題宅            扞卓拓託濯諾達丹單胆誕段談痴稚畜逐秩空嫡宙忠抽衷駐貯庁帳脹腸傲勅朕陳貨墜呈廷抵邸            亭貞帝訂通停偵軀適迭哲鉄徹撤典点展電斗徒途奴到党陶塔搭塔糖騰騰胴堂銅匣特督德篤毒            凸屯忒肉尿妊寧念能腦農把派霸婆肺俳排輩倍陪媒陪伯舶漠爆伐闊版班班呼般販搬頒範繁藩晚            番蠻盤妃批披非碑罷微百票評標秒賓頻敵瓶扶府附婦符普膚賦譜部服副復復復稜墳丙陸塀幣            弊遍弁勉舖簿邦胞俸砲肪某剖帽棒質朴僕撲沒奔盆摩魔每枚摸抹慢慢未魁密脈妙盟銘盲猛絞            厄約偷輸癒幽悠郵猶裕融予洋容庸陽擁翌羅酪覽濫欄吏理痢隘陸陸陸陸厲慮了兩料獵偵領察            療厘倫累墨類令零隸齡歷烈烈廉鍊炉芳郎浪廊樓錄綠瀆         </p>
<b>Total:</b>			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 transliteration. 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.<sup>12</sup>

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

---

<sup>12</sup> 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.<sup>13</sup> 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

---

<sup>13</sup> 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 (cont.)

ID	JRH ID	Kanji	2+SJ	2+ SJ/N	IS&R J/N/N	Homography Ex. W/F	ID	JRH ID	Kanji	2+SJ	2+ SJ/N	IS&R J/N/N	Homography Ex. W/F	ID	JRH ID	Kanji	2+SJ	2+ SJ/N	IS&R J/N/N	Homography Ex. W/F	
201	339	牛					251	409	隅					301	483	巖					
202	340	去	X		X		252	411	罷					302	485	厚					
203	342	居					253	413	君			X		303	487	駿	X				
204	343	臣					254	416	罷					304	488	縣	X				
205	344	視	X		X		255	419	群			X		305	489	幻					
206	345	差					256	420	兒	X				306	491	言	X				
207	346	虛	X		X		257	422	形					307	492	吟					
208	347	壯					258	425	莖			X		308	493	限					
209	349	魚			X		259	426	係			X		309	494	原					
210	350	御	X		X		260	427	珣					310	495	理					X
211	351	漁					261	428	殺					311	496	滅					
212	353	世			X		262	429	壯					312	497	滅					
213	354	聖					263	430	盛	X				313	498	盛	X				
214	355	住					264	432	得					314	499	已	X				
215	357	京	X				265	434	徒	X		X		315	500	戶					X
216	358	供	X				266	435	強					316	501	古					
217	362	疾					267	436	歌					317	502	呼					
218	363	疾					268	438	駭					318	503	固					X
219	364	空					269	439	頓					319	506	故					X
220	365	恭					270	440	携					320	507	社					
221	366	胸			X		271	441	權			X		321	509	庫	X				
222	367	費					272	443	觀					322	510	湖					X
223	368	強	X				273	445	礎			X		323	511	厓					
224	369	敦					274	447	迎					324	512	陸					X
225	370	額	X				275	448	觸			X		325	513	駁					
226	371	增	X				276	450	擊					326	514	腫					
227	372	様			X		277	451	濼			X		327	515	丑					X
228	373	様					278	452	欠					328	516	丑					
229	374	様					279	453	穴			X		329	519	後	X				
230	375	號	X				280	454	穴					330	521	磨					X
231	376	響					281	455	欠			X		331	523	語					
232	377	響					282	456	喘					332	524	誤					
233	378	印	X				283	458	孫					333	526	口	X				
234	379	曉			X		284	459	月	X				334	527	工	X				
235	380	蒙	X				285	460	大			X		335	528	公					X
236	381	蒙					286	462	見					336	530	功	X				
237	382	曲					287	464	厓			X		337	531	巧					X
238	384	榎	X				288	465	建					338	532	仄					X
239	385	玉					289	466	研			X		339	533	交	X				
240	388	正					290	469	兼					340	534	交					
241	389	金	X		X		291	470	鉅			X		341	535	光	X				X
242	391	動	X				292	471	鉅					342	536	向					
243	392	翠			X		293	472	鏤			X		343	538	好	X				
244	393	藍			X		294	473	鏤			X		344	539	江					X
245	396	體					295	475	藍					345	540	考					
246	397	體			X		296	477	體	X				346	541	行	X				
247	402	苦					297	478	體	X				347	545	政					
248	403	藍					298	479	體			X		348	546	夏					X
249	405	藍					299	480	運					349	547	効					
250	406	空			X		300	481	釋	X				350	548	幸					X

Appendix: Survey Result (cont.)

ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.	ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.	ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.
401	634	業						451	702	啞						501	776	弱					
402	635	業			X			452	703	夜			X			502	777	弱	X				
403	636	業			X			453	706	姿			X			503	778	姿	X				
404	637	歲						454	707	手			X			504	779	手	X		X		
405	639	灌						455	708	指			X			505	780	手	X		X		
406	640	歲	X					456	709	施						506	781	守	X		X		X
407	641	載						457	711	輩			X			507	783	取					X
408	642	載			X			458	712	載			X			508	784	取					X
409	643	載						459	713	載			X			509	785	載					
410	646	財	X					460	716	財			X			510	786	載					
411	647	罪			X			461	719	款			X			511	788	消	X				X
412	649	作	X					462	721	理			X			512	789	理					
413	650	削						463	722	理			X			513	790	理					X
414	655	醉			X			464	724	曉			X			514	791	券			X		X
415	656	醉						465	725	節						515	792	受			X		X
416	658	卮	X					466	726	示			X			516	793	授					
417	659	礼			X			467	727	示			X			517	797	收					X
418	660	副						468	728	寺			X			518	799	州			X		X
419	661	殺						469	729	次	X		X			519	800	舟			X		X
420	663	殺						470	730	耳			X			520	801	秀					
421	664	殺						471	731	自	X					521	802	朋					
422	665	變	X					472	732	似						522	803	完	X				
423	667	三			X			473	733	里	X					523	804	給	X				X
424	668	山			X			474	734	畢	X					524	805	款					X
425	669	參						475	735	佳			X			525	806	款					X
426	671	參			X			476	736	佳			X			526	807	修	X				X
427	672	修	X					477	737	理			X			527	808	款					
428	673	理			X			478	739	理			X			528	809	理					
429	674	理			X			479	740	理			X			529	811	就	X				
430	675	取						480	746	七			X			530	812	衆	X				X
431	677	歎						481	747	失						531	813	衆					
432	679	歎						482	748	至						532	814	衆					
433	682	子	X					483	750	軌	X					533	816	醜					X
434	683	支						484	751	醜						534	817	醜					X
435	684	止						485	752	質			X			535	818	十	X				X
436	685	氏			X			486	753	質			X			536	819	汁					
437	686	仕	X					487	754	質			X			537	820	充					
438	689	四			X			488	756	質			X			538	821	住					
439	690	市			X			489	757	社	X		X			539	822	兼	X				
440	691	市			X			490	758	重	X		X			540	823	兼	X				
441	692	完			X			491	760	差			X			541	824	從	X				
442	693	完			X			492	761	射			X			542	825	誤					
443	694	余			X			493	762	得			X			543	827	歐			X		X
444	695	至						494	764	斜			X			544	828	祝			X		X
445	696	同						495	765	產			X			545	830	祝	X				X
446	697	志	X					496	766	蓋			X			546	831	產					
447	698	私	X					497	767	射			X			547	834	箭					X
448	699	便			X			498	769	射	X					548	836	熟					X
449	700	副						499	770	副						549	837	出	X				X
450	701	結						500	838	述						550	838	述					

Appendix: Survey Result (cont.)

ID	JRH ID	Kanji	2+SJ	2+N	IS&K JMN	Homography Ex	WF	ID	JRH ID	Kanji	2+SJ	2+N	IS&K JMN	Homography Ex	WF	ID	JRH ID	Kanji	2+SJ	2+N	IS&K JMN	Homography Ex	WF	
601	937	藤						651	998	睡						701	074	専						
602	938	藤						652	1003	奴	X					702	075	專						
603	939	倉	X					653	1004	根						703	076	淺						
604	940	倉		X				654	1007	飲						704	077	洗					X	
605	941	糧						655	1010	井	X					705	078	染					X	
606	942	糧						656	1011	世						706	079	頤					X	
607	943	軸						657	1012	正	X					707	082	粘		X				
608	944	軸						658	1013	生	X					708	083	職					X	
609	946	礎	X					659	1014	成	X					709	085	競					X	
610	948	礎						660	1015	配						710	087	澗					X	
611	949	巾					X	661	1016	庫						711	089	漚						
612	950	巾						662	1019	性	X					712	091	漚						
613	951	巾						663	1031	性						713	093	絳					X	
614	952	臣	X					714	094	全						714	094	全						
615	953	身					X	715	095	前						715	095	前					X	
616	954	字						716	096	看	X					716	096	然					X	
617	956	字						667	1026	眉	X					717	097	然		X				
618	957	律						668	1027	堦						718	100	隴						
619	958	律	X					669	1028	滑	X					719	101	隴						
620	960	振						670	1029	盛	X					720	104	糞		X				
621	961	浸						671	1030	婿						721	106	粗		X				
622	962	眞					X	672	1031	婿						722	107	粗					X	
623	963	眞					X	673	1032	勢						723	108	疎					X	
624	964	針					X	674	1034	誠						724	109	誣						
625	965	浸						675	1035	精	X					725	111	礎					X	
626	967	進						676	1037	雙						726	112	又		X				
627	968	泰					X	677	1038	勝						727	114	早		X				
628	969	泰						678	1039	謙	X					728	115	争						
629	970	望						679	1040	駁						729	116	走						
630	971	抵						680	1042	力						730	117	蔡						
631	972	新					X	681	1044	右	X					731	118	租		X				
632	974	映						682	1045	赤						732	119	星					X	
633	975	新					X	683	1046	非	X					733	121	芝						
634	976	觀						684	1050	惟						734	122	倉					X	
635	977	仁	X					685	1051	貫						735	123	理					X	
636	978	乃					X	686	1052	跡						736	124	押						
637	979	仁	X					687	1053	積						737	125	桑					X	
638	980	尺						688	1056	切						738	126	掃						
639	982	基						689	1057	折	X					739	128	巢					X	
640	984	基						690	1060	掖						740	129	窓						
641	985	函	X					691	1061	設						741	131	廊						
642	986	水					X	692	1062	筆						742	132	綵						
643	987	忒						693	1064	箭	X					743	133	粧		X				
644	988	垂						694	1065	設	X					744	135	裝		X				
645	989	世					X	695	1066	手						745	138	漚						
646	992	世						696	1067	蝮						746	140	漚					X	
647	993	推						697	1068	千						747	142	漚					X	
648	994	鼎					X	698	1069	川						748	143	鼎					X	
649	995	遂						699	1071	占						749	144	鼎					X	
650	997	睡					X	700	1072	先						750	145	造						

Appendix: Survey Result (cont.)

ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.	ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.	ID	JRH ID	Kanji	2+SJ	2+ SJN	IS&K JMN	Homography Lex	W.F.
801	1233	啞						951	1470	背						951	1470	背					
802	1234	學						952	1479	靛						952	1479	靛					
803	1236	鹿	X					953	1473	配						953	1473	配					
804	1237	鹿		X				954	1475	敗						954	1475	敗					
805	1238	池			X			955	1476	莖						955	1476	莖					
806	1239	知						956	1478	莖						956	1478	莖					
807	1240	治		X				957	1480	莖						957	1480	莖					
808	1241	廣			X			958	1481	莖						958	1481	莖					
809	1242	歌			X			959	1484	莖						959	1484	莖					
810	1243	歌				X		960	1486	白						960	1486	白					
811	1244	遊						961	1488	追						961	1488	追					
812	1247	理						962	1489	拍						962	1489	拍					
813	1248	竹			X			963	1490	拍						963	1490	拍					
814	1251	端						964	1492	啤						964	1492	啤					
815	1252	案						965	1493	啤						965	1493	啤					
816	1255	案						966	1494	麥						966	1494	麥					
817	1256	香		X				967	1496	麥						967	1496	麥					
818	1258	中			X			968	1499	烟						968	1499	烟					
819	1259	中						969	1501	八						969	1501	八					
820	1260	虫						970	1502	蘇						970	1502	蘇					
821	1261	虫						971	1503	蘇						971	1503	蘇					
822	1265	注						972	1504	蘇						972	1504	蘇					
823	1266	良			X			973	1506	拉						973	1506	拉					
824	1267	柱						974	1507	拉						974	1507	拉					
825	1269	柱						975	1509	豆						975	1509	豆					
826	1271	毒						976	1510	豆						976	1510	豆					
827	1273	毒						977	1511	兜						977	1511	兜					
828	1275	兆						978	1512	兜						978	1512	兜					
829	1276	吋			X			979	1513	件						979	1513	件					
830	1277	吋						980	1514	件						980	1514	件					
831	1278	桃						981	1515	坂						981	1515	坂					
832	1280	勞						982	1516	坂						982	1516	坂					
833	1281	駝						983	1522	飯						983	1522	飯					
834	1282	駝						984	1524	飯						984	1524	飯					
835	1283	預			X			985	1533	比					985	1533	比						
836	1284	約						986	1534	皮						986	1534	皮					
837	1285	鳥						987	1536	云						987	1536	云					
838	1286	鳥			X			988	1538	破						988	1538	破					
839	1288	超						989	1540	體						989	1540	體					
840	1290	體						990	1542	血						990	1542	血					
841	1292	漁						991	1543	燕						991	1543	燕					
842	1295	調			X			992	1544	殼						992	1544	殼					
843	1294	調						993	1545	殼						993	1545	殼					
844	1295	聽						994	1546	殼						994	1546	殼					
845	1296	聽						995	1547	悲						995	1547	悲					
846	1297	直			X			996	1548	庫						996	1548	庫					
847	1299	沈						997	1549	庫						997	1549	庫					
848	1300	顏						998	1464	馮						998	1464	馮					
849	1304	顏						999	1553	雞						999	1553	雞					
850	1305	追						1000	1554	美						1000	1554	美					

Appendix: Survey Result (cont.)

ID	JRH ID	Kanji	2+SI	2+ N/A	IS&K FN/A	Homogeneity		JRH ID	Kanji	2+SI	2+ N/A	IS&K FN/A	Homogeneity		JRH ID	Kanji	2+SI	2+ N/A	IS&K FN/A	Homogeneity		
						Exc	WF						Exc	WF						Exc	WF	
1001	1555	備						1101	697	志					1151	1780	間					
1002	1557	備						1102	698	防					1152	1782	問					x
1003	1558	匹						1103	699	勇					1153	1784	夜					x
1004	1559	必						1104	702	並					1154	1785	野					x
1005	1560	証						1105	1704	紡					1155	1788	記					x
1006	1561	証						1106	1705	紡					1156	1789	証					x
1007	1564	水						1107	1706	傍					1157	1790	腫					x
1008	1565	表						1108	1710	暴					1158	1791	由					x
1009	1566	使						1109	1711	彫					1159	1792	油					x
1010	1569	使						1110	1712	駭					1160	1794	論					x
1011	1571	並						1111	1713	北					1161	1797	雅					x
1012	1573	並						1112	1714	北					1162	1798	女					x
1013	1574	裡						1113	1716	故					1163	1799	宥					x
1014	1575	品						1114	1718	裏					1164	1800	勁					x
1015	1576	品						1115	1722	裏					1165	1806	勁					x
1016	1577	賦						1116	1724	朝					1166	1807	殊					x
1017	1578	賦						1117	1725	九					1167	1808	殊					x
1018	1583	不						1118	1727	醉					1168	1809	憂					x
1019	1584	夫						1119	1729	醉					1169	1811	憂					x
1020	1585	夫						1120	1732	妹					1170	1812	女					x
1021	1586	付						1121	1734	埋					1171	1814	奈					x
1022	1587	布						1122	1735	葬					1172	1815	奈					x
1023	1590	布						1123	1738	葬					1173	1816	預					x
1024	1592	魚						1124	1740	万					1174	1817	幼					x
1025	1593	私						1125	1741	溝					1175	1818	用					x
1026	1594	學						1126	1745	味					1176	1819	用					x
1027	1597	章						1127	1751	臣					1177	1821	苦					x
1028	1599	蔽						1128	1752	腥					1178	1824	穢					x
1029	1600	蔽						1129	1753	腥					1179	1825	穢					x
1030	1604	蔽						1130	1754	莠					1180	1826	莠					x
1031	1605	武						1131	1755	莠					1181	1828	淫					x
1032	1607	興						1132	1756	夢					1182	1829	淫					x
1033	1608	打						1133	1757	夢					1183	1830	醜					x
1034	1609	風						1134	1759	名					1184	1831	醜					x
1035	1610	た						1135	1760	命					1185	1832	醜					x
1036	1613	幅						1136	1761	命					1186	1833	醜					x
1037	1616	覆						1137	1762	泚					1187	1835	抑					x
1038	1618	覆						1138	1765	咄					1188	1837	抑					x
1039	1619	弘						1139	1766	滅					1189	1838	欲					x
1040	1620	弘						1140	1767	免					1190	1839	欲					x
1041	1621	弘						1141	1768	免					1191	1841	欲					x
1042	1622	物						1142	1769	綿					1192	1842	風					x
1043	1623	紗						1143	1770	茂					1193	1844	蒙					x
1044	1624	紗						1144	1771	茂					1194	1845	蒙					x
1045	1626	襪						1145	1772	襪					1195	1846	襪					x
1046	1628	襪						1146	1773	襪					1196	1847	襪					x
1047	1629	襪						1147	1775	襪					1197	1848	襪					x
1048	1630	分						1148	1777	網					1198	1850	品					x
1049	1631	文						1149	1778	網					1199	1851	品					x
1050	1632	開						1150	1779	駭					1200	1856	料					x



Appendix: Survey Result (cont.)

ID	JKH ID	Kemp	2-SI	2+		Homogeneity	
				2-N	2-N	Lex	W.F.
1201	1857	里					
1202	1860	取				X	
1203	1861	腹				X	
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	並				X	
1211	1871	勢				X	
1212	1874	旅				X	
1213	1879	兵					
1214	1881	浜					
1215	1883	隆				X	
1216	1884	重					
1217	1889	種	X				
1218	1890	力	X				
1219	1891	緑	X				
1220	1892	林				X	
1221	1895	輪				X	
1222	1896	磯				X	
1223	1897	磯					
1224	1898	頤				X	
1225	1903	乳	X				
1226	1904	冷					
1227	1905	助					
1228	1906	医					
1229	1907	例					
1230	1908	新	X				
1231	1910	筆	X				
1232	1913	麗					
1233	1914	厝				X	
1234	1917	劣					
1235	1919	裂					
1236	1920	亞				X	
1237	1921	漣					
1238	1923	碑					
1239	1926	駁				X	
1240	1927	駁	X				
1241	1928	老					
1242	1931	朝					
1243	1935	理					
1244	1936	六			X	X	X
1245	1939	和	X				
1246	1940	話				X	
1247	1941	餅					
1248	1943	莖					
1249	1945	院				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。