
Mitton, Roger and Okada, Takeshi (2007) “The adaptation of an English spellchecker for Japanese writers”. Presented at: *Symposium on Second Language Writing, 15-17 Sept 2007, Nagoya, Japan*.

The adaptation of an English spellchecker for Japanese writers

Roger Mitton and Takeshi Okada

Dr Roger Mitton
School of Computer Science and Information Systems
Birkbeck College
University of London
Malet Street
London
WC1E 7HX United Kingdom

roger@dcs.bbk.ac.uk

Takeshi Okada
Department of Language Education
Division of Language Studies
Graduate School of International Cultural Studies
Tohoku University
41, Kawauchi
Aoba Ward
Sendai City
Miyagi
980-8576 Japan

t-okada@intcul.tohoku.ac.jp

Roger Mitton is a lecturer in Computer Science at Birkbeck, University of London. He has produced a computer-usable dictionary based on the Oxford Advanced Learner's Dictionary and a substantial collection of spelling errors (both are available for research use from the Oxford Text Archive). His research interest is the development of a spellchecker capable of detecting and correcting the kind of errors made by people whose spelling is poor.

Takeshi Okada is a professor at the Graduate School of International and Cultural Studies, Tohoku University. He worked at the University of London on the issue of spelling errors made by Japanese learners of English. His current interest is the application of corpus findings to EFL teaching in Japanese schools.

Abstract

It has been pointed out that the spelling errors made by second-language writers writing in English have features that are to some extent characteristic of their first language, and the suggestion has been made that a spellchecker could be adapted to take account of these features. In the work reported here, a corpus of spelling errors made by Japanese writers writing in English was compared with a corpus of errors made by native speakers. While the great majority of errors were common to the two corpora, some distinctively Japanese error patterns were evident against this common background, notably a difficulty in deciding between the letters b and v, and the

letters l and r, and a tendency to add syllables. A spellchecker that had been developed for native speakers of English was adapted to cope with these errors. A brief account is given of the spellchecker's mode of operation to indicate how it lent itself to modifications of this kind. The native-speaker spellchecker and the Japanese-adapted version were run over the error corpora and the results show that these adaptations produced a modest but worthwhile improvement to the spellchecker's performance in correcting Japanese-made errors.

1. Introduction

The last decade has seen an upsurge of interest in learner corpora [Granger 1998a] – collections of text in a particular language (usually English) gathered from non-native speakers who have been learning the language for some years. The non-standard spellings to be found in these corpora have attracted less research than higher-order features such as grammar and lexis, but some attention needs to be paid to them, at the least to minimise the introduction of spurious errors in the process of data entry [Granger 1998b], but also to prevent them from skewing the results of certain analyses. Granger and Wynne, for example [Granger and Wynne 2000], have shown that the high frequency of variant spellings in learner corpora can distort certain measures of lexical richness, such as type/token ratio. In their study they extracted lists of misspellings from English corpora produced by Dutch, French, Polish and Spanish students and, interestingly for the present paper, they comment that "a mere glimpse at the respective lists shows that each national group has its own specific problems," and that these lists might "prove useful to adapt tools such as spellcheckers to the needs of non-native users."

Mitton had made a similar suggestion some years earlier [Mitton 1996]. In describing the development of a spellchecker for British English, he suggested that, if learners of English made characteristic spelling errors related to their first language, it ought to be possible to adapt the spellchecker to cope with these errors, thus producing for example a version for German speakers, another for Spanish speakers, and so on. Okada [Okada 2005] has assembled and analysed a large corpus of English spelling errors made by speakers of Japanese, and this allows us to put this suggestion to the test.

Although there have been a few attempts to investigate differences in misspellings made by native speakers and non-native speakers, such as [Brown 1970] and [Ziahosseiny and Oller 1970], spelling errors made by Japanese learners of English have been neither compiled into a corpus nor systematically analysed. See [Okada 2005] for further discussion.

Are there any patterns of error characteristic of Japanese writers? If so, can the spellchecker be adapted to cope with them, and, if so, how much improvement does it make to the spellchecker's performance?

2. Characteristically Japanese errors

For this study we used two corpora of English spelling errors. One, collected by Mitton in the 1980's, has been available from the Oxford Text Archive for some years; we used those parts of it produced by native speakers of English. The other is

the one more recently assembled by Okada. Details of these two corpora are presented in Appendix 1.

These were *not* corpora drawn from carefully matched samples under controlled conditions to facilitate comparison. On the contrary, each corpus is a collection of subcorpora gathered in different ways, at different times, by different people from different sorts of writers. We would argue that it can still be instructive to compare them. If a strong pattern emerges which is clearly not derived from some small portion of the data, we can be reasonably confident that we are looking at a genuine finding.

Naturally, poorer spellers make a disproportionate contribution to error corpora. Some of the Japanese writers were junior high-school students with only a limited command of English, while among the contributors to the native-speaker corpus were nine-year-old children and adult-literacy students. Since some of the writers had severe spelling difficulties, some of the misspellings are pretty remote from their target words.

Let it be said first of all that the great majority of errors made by Japanese writers are indistinguishable from those made by native speakers. Compare, for example, misspellings of the word *disappoint* from the native-speaker corpus (on the left) and the Japanese-speaker corpus (on the right):

Native-speaker corpus	**Frequency**	**Japanese-speaker corpus**	**Frequency**
disapoint	100	disapoint	96
dissapoint	81	dissapoint	42
dissappoint	32	dissappoint	8
dessapoint	29	desapoint	5
disopoint	5	dicapoint	5
dissapiont	5	dispoint	3
dissipoint	3	diccapaint	1
diserpoint	2	dicepoint	1
dissopoint	2	disapaint	1
deapoit	1	disapint	1
deinpent	1	disapoind	1
derepoint	1	disapointment	1
desapoint	1	disapoit	1
desappoint	1	disappint	1
desippoint	1	disappo	1
dessipoint	1	disepoint	1
disapint	1	diserpoint	1
disapiont	1	disipoint	1
disapont	1	dissaopint	1
disappiont	1	dissaspoint	1
disappointment	1	disserpoint	1
disepoit	1	disspoint	1
dispoint	1	dusapoint	1
dispont	1	thispoint	1

dispot	1		
disserpoint	1		
dissippoint	1		
dissopite	1		
dissypoint	1		

Table 1. Misspellings of *disappoint*

Both sides of the table show the characteristic skewed distribution of spelling-test results – a single most popular misspelling, then three or four less popular and then a long tail of variations each produced by just one person. Obviously the Japanese writers had difficulty with the same aspects of *disappoint* that the native speakers did – the single *s* and the double *p* and, to a lesser extent, the vowels. Of the 177 Japanese attempts at *disappoint*, 156 (88 per cent) also occurred in the native-speaker list.

Against this background, however, three types of error stand out in the corpus as distinctively Japanese. They are:

1) the substitution of *l* for *r* and vice-versa,
2) the substitution of *b* for *v* and vice-versa,
3) the insertion of extra syllables, particularly with the vowels *o* and *u*.

That Japanese writers may be inclined to make these errors is well known and is hardly surprising. The phonemes /l/ and /r/ are not distinguished in spoken Japanese, nor are /b/ and /v/; the *Romazi* system for rendering Japanese words in the Roman alphabet makes do with the letter *r* for the Japanese l/r sound and with *b* for the b/v sound and does not use the letters *l* or *v* at all. Consequently a Japanese writer has more difficulty in choosing between *l* and *r* (or *b* and *v*) in English than a native speaker does [Okada 2005]. Similarly, spoken Japanese is formed predominantly of consonant+vowel syllables. Given an English consonant cluster such as *br*, *dr* or *tr*, a Japanese writer is inclined to insert a vowel. The word *library* illustrates all these problems, and the Japanese corpus contains misspellings such as *libelary, liberary, liburally, liburary, liveraly, liverary* and *liverely* (and many more).

For Table 2, we computed the number of times that the letters *l, r, b* or *v* should have occurred. For example, there were forty different misspellings of *library* in the Japanese corpus; if all forty had been correctly spelt, there would have been forty *l*'s and eighty *r*'s. We counted the number of times that *l* was written in place of *r*, and vice-versa, in the misspellings; *libelary*, for example, would add one to the number of times that *l* was written in place of *r*. We could then compute the proportion of *l*'s and *r*'s that were substituted, and likewise for *b*'s and *v*'s. Base totals are in parentheses.

	Native-speaker corpus	**Japanese-speaker corpus**
l written in place of *r*	0.5% (16126)	7.0% (2708)
r written in place of *l*	0.8% (13532)	16.5% (1698)
v written in place of *b*	0.7% (2906)	10.0% (982)
b written in place of *v*	0.9% (2350)	18.1% (171)

Table 2. *r/l* and *b/v* substitutions

The tendency of Japanese speakers to create extra syllables by the insertion of *o* or *u* is shown in Table 3. It is not always possible to infer, from the spelling alone, how many syllables a word has – compare *covenant* with *pavement* or *naked* with *raked*. So counting the number of syllables in a misspelling is, to some extent, a matter of conjecture. But it is generally possible to make a reasonable guess. The spellchecker that forms the topic of the next section contains a module (described in [Mitton 1996]) that makes an estimate of the number of syllables in a misspelling, so we used that for this analysis.

	Native-speaker corpus	**Japanese-speaker corpus**
Misspelling has more syllables than the target, caused by the insertion of *o* or *u*	1.2%	3.9%
Misspelling has more syllables than target, for other reasons	8.3%	7.7%
Misspelling has same number of syllables as target	68.0%	75.1%
Misspelling has fewer syllables than target	22.5%	13.3%
Base totals	(33740)	(4962)

Table 3. Numbers of syllables in misspellings compared with target words

3. Adapting the spellchecker

Spellcheckers cope poorly with these distinctively Japanese errors. For example, out of the seven attempts at *library* listed earlier, only for two of them - *liberary* and *liburary* - does Microsoft Word succeed in offering *library*. In particular, an *l/r* or *b/v* substitution near the beginning of a word is likely to mislead a spellchecker into searching a completely wrong part of the dictionary.

The spellchecker that was referred to in the introduction can be adapted fairly easily to cope with particular patterns of error, but, to see how, we need to give a brief sketch of its normal mode of operation. (Readers familiar with the literature on string matching will recognise that it uses a version of minimum-edit distance [Wagner 1974], adapted in a similar way to that described in [Veronis 1988].)

At the heart of its procedure is an algorithm that takes a misspelling and a possible target word and compares them letter by letter from left to right. At each point on its progress, it is comparing a letter of the misspelling (let's call it LM) with a letter of the target (let's call it LT) and it has three options: it can decide that LM has been wrongly inserted, or that LT has been wrongly omitted, or that LM has been substituted for LT. (If LM is the same as LT, it can accept LM as correct – a special case of substitution.)

Suppose, for example, that it is considering *phone* as a possible target for the misspelling *fown*. (It would also consider *frown* and *town* and many other words in the same sort of way.) What do you have to do to *phone* to end up with *fown*? The algorithm might decide that *f* has been substituted for *p*, that the *h* has been omitted, that the *o* is correct, that the *w* is an insertion, that the *n* is correct, and that the *e* has been omitted. It could conclude that, to get from *phone* to *fown*, you need at least four

editing operations (one substitution, one insertion and two omissions) – an "edit-distance" of 4.

This simple counting of operations is a bit too simple, however. Consider *stone* as a possible target. By the logic of the previous paragraph, *fown* and *stone* also have an edit-distance of 4. But surely someone who writes *fown* is much more likely to be trying to write *phone* than *stone*? We would like the algorithm to tell us that *fown* and *phone* are closer than *fown* and *stone*.

We can achieve this by putting information into the spellchecker's dictionary. We can prime the entry for *phone* so that *f* is accepted as a plausible substitution for *ph* (the algorithm picks up information from the entry for *phone* so that it counts the substitution of *f* for *p* and the omission of *h* at a reduced rate). Similarly we can arrange that *own* may be accepted as a possible alternative to the *one* of *phone* (since there are similar-sounding words such as *own*, *grown*, *blown* etc). The effect of this is to reduce the edit-distance of *fown* from *phone*. By contrast the entry for *stone* says nothing about accepting an *f* for the *st*, so the substitution of *f* for *s* and the omission of the *t* would be counted at the full rate (though the replacement of *one* by *own* would be counted at a reduced rate for the same reasons as for *phone*), so the algorithm concludes that *fown* is closer to *phone* than to *stone*.

For a more detailed treatment of this topic than is appropriate in this short paper, the reader is referred to [Mitton 1996].

Having detected a misspelling, the spellchecker retrieves a few hundred words from its dictionary that look as though they might be the target and then compares each of these with the misspelling, assessing how good a match each one is by the above algorithm. Having computed an edit-distance for each candidate word, it ranks them in order, with the best matches (the shortest edit-distance) at the top, and offers the top few to the user.

The adaptation for Japanese writers is simple. Just as the spellchecker is already primed to expect, for example, an *f* instead of a *ph* in misspellings of a word like *phone*, we arrange for it also to anticipate a *v* instead of a *b* in misspellings of a word like *library*. Likewise *b* for *v*, *l* for *r*, and *r* for *l*.

4. The effect of the adaptations

Given a misspelling, the spellchecker generates a list of a few hundred suggestions, ranked in order with its best guess at the top. If we know what the target was, we can see if the target appears in the list and, if so, how near the top. Our corpora contain thousands of misspellings, with the target recorded for each one. We can therefore run the spellchecker over the corpora, generating a list of suggestions for each misspelling and recording where the target appears in the list. (We excluded from this exercise those misspellings where the target word was not in the spellchecker's dictionary since adaptations to the spellchecker could obviously make no difference in those cases.)

We have two corpora and two versions of the spellchecker – the original one designed for native speakers and the version adapted for Japanese writers. This gives us four sets of results, presented in Table 4.

	Native-speaker corpus, native-speaker spellchecker	Native-speaker corpus, Japanese spellchecker	Japanese corpus, native-speaker spellchecker	Japanese corpus, Japanese spellchecker
First	54.2%	53.3%	61.2%	65.8%
Top three	67.9%	67.3%	73.3%	78.7%
Top six	73.4%	73.0%	77.9%	83.5%
Base total	35612	35612	4848	4848

Table 4. The performance of the two spellcheckers on the two corpora

The first column of the table shows that when the original version of the spellchecker was run on the native-speaker corpus, it produced the target at the top of its list for 54.2% of the misspellings; for 67.9% of the misspellings it offered the target in the top three of its suggestions, and for 73.4% in the top six.

If its success rate seems poor, remember that the base totals are of types, not tokens. In other words, each misspelling occurs just once in the corpus; so succeeding with *disapoint* (which it does), for example, counts as only one success, even though this misspelling is a common one, whereas failing (which it does) with *deapoit* (another misspelling of *disappoint)* counts as one failure, even though this particular misspelling was written by just one person.

As you would expect, the Japanese version of the spellchecker (column two) did not perform better than the original one for the native-speaker corpus; if anything it performed slightly worse. The original spellchecker (column three) performed fairly well with the Japanese corpus, as one would expect given that the majority of the Japanese-made errors were very much like native-speaker ones. But the fourth column is the important one. The Japanese version of the spellchecker performed appreciably better than the original when run on the Japanese corpus.

The superior performance of the Japanese version on the Japanese errors is almost entirely due to the adaptations for *l/r* and *b/v*. Adaptations were also made so that it would cope better with the extra syllables of the Japanese misspellings, but these made hardly any difference. The reason is, simply, that the original spellchecker is already very forgiving of extra vowels, paying much more attention to consonants.

As we made clear earlier, the Japanese corpus, like the native-speaker one, is made up of several subcorpora (see Appendix 1), and the contributors to these subcorpora varied in their command of English. Some will have been more prone to make spelling mistakes and, in particular, to make the sort of spelling mistakes that we have focused on. And, obviously, they were not all trying to spell the same words. Consequently the effect of the Japanese-adapted spellchecker shows up more with some of the subcorpora than with others. Appendix 2 presents these results. The effect shows up, to varying degrees, in all of the constituent subcorpora, though in some the difference is very slight.

Appendix 3 compares the performance of the two versions of the spellchecker on one part of the Japanese corpus, namely 111 misspellings of the word *albatross*, a hard word for Japanese writers.

5. Conclusion

This experiment has shown that there are distinctive features in Japanese misspellings of English, that a spellchecker designed originally for native speakers of English could be easily adapted to cope with these features, and that these adaptations made a modest but worthwhile improvement to the spellchecker's performance when dealing with Japanese-made errors.

References

Brown, D. H. (1970). Categories of spelling difficulty in speakers of English as a first and second language. *Journal of Verbal Learning and Verbal Behavior, 9,* 232-236.

Furugouri, T. & Hiranuma, K. (1987). Statistical characteristics of English sentences written by the Japanese and detecting and correcting spelling-errors. *Mathematical Linguistics, 16,* 16-26.

Granger, S. (Ed.) (1998). *Learner English on computer.* London: Longman.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London: Longman.

Granger, S. & Wynne, M. (2000). Optimising measures of lexical variation in EFL learner corpora. In J. M. Kirk (Ed.), *Corpora galore* (pp.249-258). Amsterdam: Rodopi.

Mitton, R. (1996). *English spelling and the computer.* London: Longman. (now available electronically from `eprints.bbk.ac.uk`)

Okada, T. (2003, May). *Spelling errors made by Japanese EFL writers: A corpus-based approach.* Paper presented at Colchester Second Language Acquisition Workshop 2003, University of Essex, Colchester, UK.

Okada, T. (2004). A corpus analysis of spelling errors made by Japanese EFL writers. *Yamagata English Studies, 9,* 17-36.

Okada, T. (2005). Spelling errors made by Japanese EFL writers: with reference to errors occurring at the word-initial and the word-final position. In V. Cook & B. Bassetti (Ed.), *Second language writing systems* (pp.164-183). Clevedon: Multilingual Matters.

Veronis, J. (1988). Computerized correction of phonographic errors. *Computers and the humanities, 22,* 43-56.

Wagner, R. A. & Fischer, M. J. (1974). The string-to-string correction problem. *Journal of the A C M, 21(1),* 168-73.

Ziahosseiny, S. M. & Oller, J. W. Jr. (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning, 20(2),* 183-189.

Appendix 1: The corpora

The Japanese corpus consists of 5060 different misspellings representing over 12000 attempts at 1184 target words. Note that the misspellings in the corpus are types, not tokens; *hight*, for example, was written 181 times for *height*, but it has only one entry. The corpus is an amalgamation of the following seven subcorpora:

1. AEMH-error.txt

Misspellings extracted from English essays handwritten in class by 244 Japanese university students, 201 of them majoring in English. There were 20299 running words in total; 393 of these were misspelt, which, after removal of duplicates, gives us 296 misspellings of 234 target words. For further details of the raw material, refer to the URL's of the original source files:
`http://www.lb.u-tokai.ac.jp/lcorpus/data/asao01/`
`http://www.lb.u-tokai.ac.jp/lcorpus/data/asao02/`
`http://www.lb.u-tokai.ac.jp/lcorpus/data/shitara01/`

2. EXAMS-error.txt

162 misspellings of 151 target words, taken from the Japanese part of EXAMS.DAT included in the Birkbeck Spelling Error Corpus (`http://ota.ahds.ac.uk/`). This contains 213 attempts generated by 49 Japanese writers. The misspellings are taken from compositions written in examinations for the Cambridge First Certificate in English.

3. HELC-JR-error.txt

Junior high-school students were given sentences in class to translate from Japanese into English. There were 286 target sentences and the students produced 85120 running words in total. The number of subjects per target sentence varied from 20 to 120. The subcorpus contains 1921 misspellings of 431 target words (3366 attempts). The original source is maintained as *Hiroshima English Learners' Corpus No.1* by Shogo Miura at Hiroshima University, Japan.

4. HELC-SR-error.txt

Similar to the previous corpus except with senior high-school students. There were 68 target sentences and 40638 running words. The number of subjects per target sentence varied from 40 to 120. This subcorpus contains 346 misspellings of 187 target words (673 attempts). The original source is maintained as *Hiroshima English Learners' Corpus No.2*. This and the previous subcorpus are described at:

`http://home.hiroshima-u.ac.jp/d052121/eigo1.html`

5. SAMANTHA-error.txt

Japanese university students were given a test of 53 English words. For each word, they were given a written definition in Japanese and an approximation in katakana to the English pronunciation. 333 people sat the test. 7418 of their attempts were incorrect, giving 2071 misspellings of 53 target words. The original error corpus, named SAMANTHA Error Corpus, is maintained by Takeshi Okada at Tohoku University, Japan.

http://www.intcul.tohoku.ac.jp/okada/corpora/Samantha/Samantha-top.html

6. SUZUKI-error.txt

Personal collection of misspellings made by unspecified number of Japanese high-school students in their classroom activities or in short tests. Collected by Michiaki Suzuki at Nan'yo High School, Yamagata, Japan. This subcorpus does not contain frequency information. There are 46 misspellings of 43 target words.

7. FRGRI-error.txt

366 misspellings of 324 target words obtained from compositions written by 88 Japanese university freshmen not majoring in English. The students also submitted translations of their compositions. The list is given in an article written in Japanese [Furugouri and Hiranuma 1987]. This subcorpus also does not contain frequency information.

For comparison, a corpus of errors made by native speakers was created by combining several files from the Birkbeck Spelling Error Corpus, obtainable from the Oxford Text Archive. The result was a corpus of over 35,000 misspellings of nearly 6000 target words, representing over 220,000 attempts. Table 1 summarises the constituent files. In the case of the three American files, words whose spellings differed from British English were excluded. The remaining subcorpora were British.

File name	Source	Target wds	Attempts	Misspellings
CHES	202 10-year-old children	30	2474	1364
FAWTH1	Printed American sources	739		809
FAWTH2	3 adult poor spellers	484	1084	557
GATES	Pupils in New York schools	3390	144179	4401
MASTERS	American school + university	264	43755	13020
NFER1	83 Adult literacy students	40	838	495
PERIN1	42 Sec school + adult lit stds	61	807	640
PERIN2	6 adult-literacy students	538	658	625
PERIN3	176 14- and 15-year olds	40	1678	901
PETERS1	156 children at 9, 10 and 11	290	18304	10556
PETERS2	925 15-year-olds	1618	4147	2576
UPWARD	163 15-year-olds	576	1073	753
WING	40 univ entrance candidates	185	237	191

Table A1. The component files making up the native-speaker corpus

For further details of each original error file, refer to the description file that accompanies the Birkbeck Spelling Error Corpus from the Oxford Text Archive (http://ota.ahds.ac.uk/).

The comparative spelling error corpora are available from the web page below.
http://www.intcul.tohoku.ac.jp/okada/corpora/Atsuo-Henry/

Appendix 2: The performance of the two spellcheckers on the Japanese subcorpora

The following table compares the performance of the native-speaker and Japanese-adapted spellcheckers for each of the Japanese subcorpora. The figures show the percentage of errors for which the spellchecker proposed the correct word as the first in its list of suggestions. The superior performance of the adapted spellchecker can be seen in all the subcorpora, though it varies from quite large (SUZUKI and SAMANTHA) to barely perceptible (FRGRI and HELC-JR).

Corpus	Native-speaker spellchecker	Japanese-adapted spellchecker	N of errors
SUZUKI	80.4	89.1	46
SAMANTHA	55.2	63.9	1953
AEMH	71.2	74.5	274
HELC-SR	73.4	75.5	327
EXAMS	81.5	83.4	151
HELC-JR	59.5	61.2	1878
FRGRI	82.9	83.2	368

Table A2. The percentage of errors, in each of the Japanese subcorpora, for which the spellcheckers proposed the correct word first in the list of suggestions

A **t** test for dependent scores was carried out on each subcorpus; for each of the spellcheckers, the position of the correct word in its top ten suggestions for each of the misspellings counted as its "score" for that misspelling. A score of 11 was given if the correct word did not appear in the top ten. The value of **t** is affected, of course, by the number of errors in the subcorpus as well as by the amount of difference in the scores. For SAMANTHA, AEMH, HELC-SR and HELC-JR, the differences are statistically significant at the 0.05 level or beyond; for SUZUKI, because of the small number of errors, they are significant only at the 0.1 level. For EXAMS and FRGRI, the differences fail to reach statistical significance, though they are in the same direction as for the other subcorpora and contribute to the same overall picture.

Appendix 3: Japanese misspellings of *albatross*

The following table shows how the two versions of the spellchecker coped with a large number of Japanese misspellings of the word *albatross*, a difficult word for Japanese writers since it contains *l, b* and *r,* and has two places – *lb* and *tr* – where it is tempting to insert a vowel.

The first column contains the misspelling itself and the second column shows how many of the people who contributed to the corpus produced this particular misspelling. The third and fourth columns record the suggestions of the native-speaker spellchecker. The third shows where the target word *albatross* came in the list of suggestions; a zero indicates that the target did not appear in the list at all. The fourth column shows the word that came first in the list of suggestions for this misspelling (where the target word was offered first in the list, then this word is, obviously, *albatross*). The fifth and sixth columns give comparable data for the Japanese version of the spellchecker.

The main difference is that the Japanese version performed much better for misspellings beginning *ar*. (Only for one of these – *arbatross* – did the native-speaker version match the Japanese version's performance; this is because the spellchecker has a special routine for handling misspellings that differ from the target by just one letter.)

		Native-speaker spellchecker		*Japanese spellchecker*	
Misspelling	*Frequency*	*Pos of target.*	*First suggestion*	*Pos of target.*	*First suggestion*
albatros	56	1	albatross	1	albatross
arbatros	28	0	arbiters	1	albatross
albatoros	27	1	albatross	1	albatross
arbatross	16	1	albatross	1	albatross
albatrous	9	1	albatross	1	albatross
albatlos	7	1	albatross	1	albatross
arbatoros	7	0	arbiters	1	albatross
arbatrous	7	0	abattoirs	1	albatross
arubatros	7	0	abattoirs	1	albatross
alvatros	6	1	albatross	1	albatross
alubatros	5	1	albatross	1	albatross
albatoross	4	1	albatross	1	albatross
albatolos	3	1	albatross	1	albatross
albatorous	3	1	albatross	1	albatross
albutros	3	1	albatross	1	albatross
arbatolos	3	0	abattoirs	1	albatross
arubatoros	3	0	arbiters	1	albatross
arvatros	3	0	avatars	1	albatross
albatloss	2	1	albatross	1	albatross
albatorose	2	1	albatross	1	albatross
albatos	2	1	albatross	1	albatross
albatrose	2	1	albatross	1	albatross
albertros	2	1	albatross	1	albatross
albertross	2	1	albatross	1	albatross
albertrous	2	1	albatross	1	albatross
aldatoros	2	1	albatross	1	albatross
alvatoros	2	3	elevators	1	albatross
alvatross	2	1	albatross	1	albatross
alvatrous	2	1	albatross	1	albatross
arbatloss	2	0	abattoirs	1	albatross
arubatoross	2	0	arbiters	1	albatross
arubatross	2	0	arbiters	1	albatross
ulbatross	2	1	albatross	1	albatross
albadoross	1	1	albatross	1	albatross
albathross	1	1	albatross	1	albatross
albatlas	1	1	albatross	1	albatross
albatlaus	1	1	albatross	1	albatross
albatlus	1	1	albatross	1	albatross
albatorsu	1	1	albatross	1	albatross
albatous	1	1	albatross	1	albatross
albatrce	1	1	albatross	1	albatross

albatris	1	1	albatross	1	albatross
albatroth	1	1	albatross	1	albatross
albatrros	1	1	albatross	1	albatross
albattlos	1	1	albatross	1	albatross
albattros	1	1	albatross	1	albatross
albattrous	1	1	albatross	1	albatross
albertolose	1	3	albatrosses	1	albatross
albertrose	1	1	albatross	1	albatross
albtoros	1	1	albatross	1	albatross
albtros	1	1	albatross	1	albatross
albtross	1	1	albatross	1	albatross
albtrus	1	1	albatross	1	albatross
albutoros	1	1	albatross	1	albatross
albutross	1	1	albatross	1	albatross
albutrous	1	1	albatross	1	albatross
aldatolos	1	1	albatross	1	albatross
allatoroce	1	26	illiteracy	35	illiteracy
allbutoloss	1	1	albatross	1	albatross
allubatros	1	1	albatross	1	albatross
alubatolos	1	1	albatross	1	albatross
alubatoros	1	1	albatross	1	albatross
alubatorose	1	1	albatross	1	albatross
alubatoross	1	1	albatross	1	albatross
alubatorous	1	1	albatross	1	albatross
alubatos	1	1	albatross	2	elevators
alubatras	1	1	albatross	1	albatross
alubatrosu	1	1	albatross	1	albatross
alvatoloss	1	5	alveolars	1	albatross
alvatrose	1	1	albatross	1	albatross
alvatrus	1	1	albatross	1	albatross
alvutross	1	1	albatross	1	albatross
arbatlos	1	0	abattoirs	1	albatross
arbatoras	1	0	abattoirs	3	abattoirs
arbatrros	1	0	abattoirs	1	albatross
arbatrus	1	0	arbiters	1	albatross
arbertoros	1	0	arbiters	3	arbiters
arbertros	1	0	arbiters	1	albatross
arbertross	1	0	arbiters	1	albatross
arbertrous	1	0	abattoirs	2	abattoirs
arbtros	1	0	arbiters	1	albatross
arbutros	1	0	arbiters	1	albatross
ardatros	1	0	audacious	1	albatross
arrbatros	1	0	abattoirs	1	albatross
arrubatroce	1	0	aerobatics	3	aerobatics
arubatoras	1	0	arbiters	3	arbiters
arubatoroc	1	0	arbiters	3	arbiters
arubatorous	1	0	arbiters	1	albatross
arubatrous	1	0	abattoirs	1	albatross

arubats	1	0	Arabist	9	Arabist
arubtros	1	0	arbiters	1	albatross
arubutros	1	0	arbiters	1	albatross
arubuts	1	0	abuts	15	abuts
arudatorosu	1	0	audacious	7	audacious
aruvatoros	1	0	aviators	1	albatross
arvatoras	1	0	aviators	6	aviators
ulbatolce	1	8	oblations	7	orbital
ulbatorous	1	1	albatross	1	albatross
ulbatros	1	1	albatross	1	albatross
ulbertorous	1	3	liberators	3	liberators
ulbtlos	1	1	albatross	1	albatross
ulvatrous	1	3	olive-trees	1	albatross