Not the word I wanted? How online English learners' dictionaries deal with misspelled words

Conference Presentation

Not the Word I Wanted?
How Online English Learners' Dictionaries Deal with Misspelled Words

Robert Lew
Adam Mickiewicz University
E-mail: rlew@amu.edu.pl

Roger Mitton
Birkbeck College, University of London
E-mail: roger@dcs.bbk.ac.uk

Abstract

This study looks at how well the leading monolingual English learners' dictionaries in their online versions cope with misspelled words as search terms. Six such dictionaries are tested on a corpus of misspellings produced by Polish, Japanese, and Finnish learners of English. The performance of the dictionaries varies widely, but is in general poor. For a large proportion of cases, dictionaries fail to supply the intended word, and when they do, they do not place it at the top of the list of suggested alternatives. We attempt to identify some of the mechanisms behind the failures and make further suggestions that might improve the success rate of dictionary interfaces when identifying and correcting misspellings. To see whether it is possible to do better than the dictionaries tested, we compare the success rates of the dictionaries with that of an experimental context-free spellchecker developed by the second author, and find the latter to be markedly superior.

Keywords: online dictionaries; English; spelling; spelling correction; spelling errors; misspelling; access

1. The role of spelling in dictionary consultation

One painful limitation of traditional paper dictionaries is that the primary access route — at least for the most popular semasiological (form-to-meaning) reference works and for languages with alphabetic writing systems — requires that the user (1) is familiar with the access alphabet (Nielsen, 1995) of the dictionary, and (2) knows how the target item is spelled. With reference to the first point, users of modern electronic dictionaries are indeed (if only up to a point) 'liberated from the straitjacket of ... alphabetical order' (Atkins, 1996: 516), thus making alphabetical ordering less of a critical factor in the success of the access process. However, point (2) remains a valid concern: the dictionary engine still needs to match the search term entered by the user against the available list of keywords covered in the dictionary, which include, but need not be limited to, the headwords.

Of course, dictionary users cannot always be expected to replicate standard English spelling. Further, cases of misspelling can result from a mechanical typo (performance errors) or erroneous lexical-graphemic representation (competence errors). In the second case in particular, misspelling patterns (Mitton & Okada, 2007) typical of native speakers of English may be different from those of learners of English.

Further, online dictionaries are increasingly used in conjunction with online work and entertainment. This includes the need for lexicographic assistance in the context of listening, such as when learners of English attempt to look up a word which they hear being spoken while watching a TV show on their computer. Such a lookup situation is bound to generate queries where the search term, rather than representing a specific vocabulary item from the learner's lexical repertoire, is a 'creative spelling', a shot-in-the-dark: a transcription of what the user imagines he has heard. This is a little similar to what some call *phonetic spelling* (cf. Proctor, 2002), but more complex, as here not one but at least two phonological systems are involved, with their own phonotactic regularities and spelling-to-sound correspondences. We would expect the best electronic dictionaries to be able to offer useful assistance in all of the above cases; but do they actually provide such assistance?

2. Spelling correction in e-dictionaries

No matter how rich and sophisticated the lexicographic content a dictionary, it will be completely lost on the users if they do not succeed in finding their way to the appropriate dictionary entry.

In a common type of e-dictionary interface where lookup consists in the user typing in a search term into a search box, the string entered needs to be matched against a list of keywords held in the dictionary itself, to see if it corresponds to a lemma present in the dictionary, or possibly (in more sophisticated dictionaries) is part of a multi-word unit treated under a different lemma. An exact-match algorithm would assume that dictionary users are perfect spellers, which is obviously not a realistic assumption. Ideally, a good dictionary interface should be able to guess the user's intention even if they misrepresent the orthographic form of the word. However, in a recent analysis of three online German dictionaries (Bank, 2010), only one dictionary has been found to be at all 'rechtschreibtolerant' — that is, able to deal with misspellings in any useful fashion.

A good dictionary interface — when presented with an unknown string — should make reasonable guesses as to the possible alternative forms the user may have meant. Furthermore, if the guesses are presented as an ordered list, then the best guesses should be close to the top of the list. In an ideal case, the one word actually intended by the user should be presented at the very top of the list of suggestions, but this ideal is not always achievable —

even in the best possible system — due to the inherently erratic nature of misspellings. The demands on the dictionary are here similar to those on a state-of-the-art spellchecking system in a word processor, though not identical.

First, the dictionary needs to recognize that the search term entered is not a standard spelling. Then, it needs to home in on a compact set of the most likely alternatives and rank them, so that they can be presented back to the user as an ordered list. Or, less commonly, it might just take the user to the entry for the top-ranking alternative (much as the Google search engine currently does). In broad outline, the procedure is similar to that involved in checking texts; however, there are differences, such as the opportunity to use context to refine the list (typically absent in dictionary lookups), or the need to handle proper nouns.

2.1 Types of spelling errors

Many of the spelling errors in running text are single-error departures from the target word. Taking the target word *trepidation* as an example, these are usually understood as being one of the following four subcategories: a single letter is omitted in a word (*tepidation*); a single letter is wrong (*trepitation*); one extra letter is inserted (*treppidation*); two adjacent letters are transposed (*trepidaiton*). According to some studies (Damerau, 1964; Pollock & Zamora, 1984), such simple errors may account for over eighty percent of misspellings. However, this percentage is likely to be lower with a more realistic representation of poor spellers in the corpus: sixty-nine percent in Mitton (1996: 46). Many (though not all) of these simple errors tend to be the result of mistyping words. As such, they are mechanical errors of performance, rather than errors of competence, and some authors use the term *misspelling* in a narrower sense which excludes mistypings (e.g. Deorowicz & Ciura, 2005). Though it is not always possible to categorize an error as one type or the other (e.g. **accomodation* for *accommodation*, or **consistant* for *consistent*), their underlying causes are different. It is misspellings of the competence type that are our primary focus here.

At the other end of the mechanical-conceptual cline, there are non-standard formations at the lexical-morphological level, such as when a speaker actually has the word **unpolite* in their mental lexicon and uses it in place of (or as a variant of) the standard *impolite*. Though sometimes the source of genuine problems, especially for non-native users of a language, it is doubtful if such errors of lexical competence should be classified as strictly *spelling-related* (pace Deorowicz & Ciura, 2005).

2.2 Rare versus common words

Low-frequency words are, by definition, words that are used infrequently in running text. Therefore, it is reasonable to assume in a spelling correction system that an instance of a rare word (especially a very infrequent one) may be a misspelling if there is a common word to which it bears some similarity. For example, as pointed out by Mitton (1996: 96), the orthographic string *wether* when found in a running text is more likely to be a misspelling of either *whether* or *weather* than the rare word meaning 'a castrated ram'. Spelling correctors working with text can use this information to detect and flag such potential real-word errors. However, in a corpus of strings being looked up in an online dictionary, the frequency distribution of word forms is less skewed than in running text (De Schryver et al., 2006), so that even quite rare words have a fair chance of being looked up. This makes perfect sense: when someone reads a text, they will not usually be troubled by all the familiar common words, but the occasional rare word is likely to be looked up. So, although De Schryver et al.'s study of log files presents only a single piece of evidence, it is reasonable to assume that native speakers, and to a lesser extent advanced learners, often consult their dictionaries for less frequent words.

2.3 The role of context

Most work in spellchecking and spelling correction so far has been done with reference to forms embedded in textual context, and some of the more advanced systems attempt to utilize contextual information to improve the accuracy of guessing at the form intended. However, when online dictionaries are consulted, it is at present most usually by typing a word into the search window. In such a setup, no contextual information would normally be available to the dictionary application. Still, most spellcheckers designed for the correction of texts do not use context either, and yet achieve good success rates nevertheless (Kukich, 1992; Deorowicz & Ciura, 2005; Mitton, 2009).

3. The study

3.1 Aim

The aim of the study is to assess the performance of the leading monolingual learners' dictionaries of English in their online versions at guessing the intended headword when presented with their misspelled versions produced by foreign learners. By *performance* we here mean the particular ability to recover the intended lemma and suggest it back to the user as a plausible alternative to what the user has actually typed in the search box. Ideally, the intended word should be offered as the only suggestion, but usually several alternatives will remain plausible, so dictionaries will customarily provide not just one suggestion but a short list. In such a case, the nearer the top of the list the intended headword appears, the better the performance of the spelling correction mechanism.

More specifically, we would like to find out whether the level of performance of the most prestigious dictionaries is in general satisfactory, to what extent the different dictionaries perform similarly or differently, and how specific dictionaries compare with the others.

Our corpus (see 3.2 below) includes misspellings by

learners of varying linguistic backgrounds (Polish, Japanese, and Finnish), and it might also be interesting to see if some dictionaries are perhaps better equipped to cope with misspellings typical of learners speaking a given native language.

In view of the preliminary results indicating that the tested dictionaries performed below expectation, a further aim was added during the course of the study, and for this, the original author was joined by the second author. This further aim was to see if an experimental context-free spelling corrector designed by the second author (Mitton, 1996) would be able to perform better than the dictionaries tested.

3.2 Corpora of misspellings used in the study

The corpus of spelling errors used in the present study is made up of 202 misspellings broken down into three subcorpora, each representing attempts at spelling English words by native speakers of three different languages that are typologically very distant, as they all represent different language families: Polish (100 items), Japanese (52), and Finnish (50). A brief description of the three sets of misspellings follows, and a sample of ten items from each is given in the Appendix.

3.2.1 Polish misspellings

The most substantial part of the corpus of misspellings used in this study came from a Polish subcorpus, collected in 2010 by the first author, with the help of two student assistants as experimenters.

The data were collected by way of oral elicitation. A set of English words known to be frequently misspelled was taken from *The 200 Most Commonly Misspelled Words in English*[1] reported by Richard Nordquist, and these were used as elicitation triggers (target words). One by one, the words from the list were played back in audio form to one of two Polish learners of English at first year of college (one female from Szczecin University, one male from Gdańsk University), using the built-in audio pronunciation capability of the popular bilingual English-Polish dictionary Diki.pl, known for its decent audio quality. Thus, a target word would be played back to the participant without disclosing its orthography, and the participant would respond by typing the word into the computer. The experimenter would wait until the participant indicated that they were done, and then proceed to play back the next target word. Participants had been instructed in the warm-up sessions to proceed as if they were looking up words just heard in an online dictionary.

All the typed wordlike strings were logged. Correctly spelled words as well as obvious mistypings, which in all likelihood would not have challenged the spellchecking algorithms of the dictionaries, were subsequently removed, with the remaining strings yielding the Polish subcorpus of 100 misspellings. This elicitation technique is believed to mimic dictionary lookup behaviour for stimuli perceived aurally (i.e., while listening).

[1] http://grammar.about.com/od/words/a/misspelled200.htm

3.2.2 Japanese misspellings

The 52 Japanese misspellings were taken from the SAMANTHA Error Corpus created by Takeshi Okada at Tohoku University, Japan. In order to collect the misspellings, Japanese students had been asked to write down an English word based on its definition in Japanese and an approximate representation of English pronunciation in the Japanese moraic (or, more loosely, syllabic) script katakana. For this study, the most common misspelling was selected from the corpus which was not a single-error type (and thus not challenging enough for spellcheckers). Up to a point, though perhaps not as much as for the Polish sample, the elicitation technique used would be likely to produce misspellings influenced by Japanese orthotactic and phonotactic rules (i.e., the typical sequencing of letters and sounds, respectively), as well as the native spelling-to-sound correspondences.

3.2.3 Finnish misspellings

The set of Finnish misspellings was obtained from the Birkbeck spelling error corpus (Mitton, 1985) via the Oxford Text Archive. The Finnish data themselves were collected by Suomi (1984) as part of her MA research. Her corpus also included data from native speakers of Swedish, but for this study, only the data from native speakers of Finnish were used. We also discarded most obvious mistypings, as for the Polish corpus. This resulted in a list of 50 misspellings.

3.3 Dictionaries tested

Each of the misspelled words in the corpus was looked up manually in each of the following seven dictionaries, all except the Google Dictionary being dictionaries for advanced learners of English, and all but one freely available online. The seven dictionaries tested were (URL's are given in the **References** section):

1. *Longman Dictionary of Contemporary English*, free online version (henceforth, LDOCE Free);
2. *Longman Dictionary of Contemporary English*, premium subscription version (LDOCE Premium);
3. *Merriam-Webster's English Learner's Online Dictionary* (MWALED);
4. *Macmillan English Dictionary Online* (MEDO);
5. *Cambridge Advanced Learner's Dictionary* (CALD);
6. *Oxford Advanced Learner's Dictionary* (ALD), and
7. *Google English Dictionary* (GoogleED).

The general idea was to test English monolingual dictionaries for learners of English available freely online. The set of leading English monolingual learners' dictionaries is actually well defined, and is frequently referred in the lexicographic literature as the Big Five, and includes: ALD, LDOCE, COBUILD, CALD, and MEDO. Of these, COBUILD has not been tested as it does not currently offer a free online version. For LDOCE, two versions were tested: the free online version, and also a Premium version. This version is available by

subscription, with time-limited access granted to buyers of paper and DVD-Rom copies. It was included in order to see if paying users were being served better than users of the free version (in fact, quite the reverse turned out to be the case, as we shall see below).

In addition to these four British learners' dictionaries, we also included MWALED. Even though in terms of lexicographic content this American-made learner's dictionary may still not compare very favourably with the Big Five (Hanks, 2009; Bogaards, 2010), its web interface does offer some commendable features (Lew, 2011).

Finally, GoogleED was also included in the study. GoogleED used to be a learners' dictionary of sorts, with the core lexicographic content apparently based on COBUILD. In August 2010, GoogleED switched over to the *Oxford American College Dictionary* (Lindberg, 2006), which is not a dictionary targeted at language learners, but primarily at American college students speaking English as their native tongue. However, four factors spoke in favour of including GoogleED in the sample.

First, being associated with Google, the unquestioned leader in search engines, it was reasonable to expect it to become a very significant player also as an online dictionary of English for non-English-speaking netizens.

Second, its history as an online version of COBUILD, one of the Big Five, is in itself significant, and may have attracted a number of learner users who remained regular users even after the switch.

Third, although the *Oxford American College Dictionary* is a native-speaker dictionary, it is largely based on the *New Oxford American Dictionary* (McKean, 2005), which, in turn, grew out of the *New Oxford Dictionary of English* (Hanks & Pearsall, 1998). This latter dictionary benefited from Patrick Hanks' prominent involvement with the COBUILD project, and so in many ways it is closer to the learner dictionary model than a traditional dictionary for native speakers of English.

Finally, Google has become a sort of a synonym for data search and access. We therefore wanted to challenge the experts, as it were, and see if GoogleED would perform better than the 'regular' dictionaries.

Somewhat surprisingly, GoogleED is no longer officially available online as of this writing (2 Sept 2011). Apparently, it was discontinued without warning in August 2011. However, much of the functionality can still be accessed by using the *define: term* syntax in a general Google search, and then clicking on *more* within the top item on the results list, which selects the *Dictionary* tab from the sidebar currently appearing to the left of the Google search user interface. Alternatively, the same effect can be achieved more directly by appending a parameter value of *tbs=dfn:1* to a Google search. For example, to get directly to the Google dictionary entry for the word *bay*, one would at this time use the following URL: http://www.google.com/search?q=bay&tbs=dfn:1. In some browsers (Opera, for example), it is possible to define customized search shortcuts of this type, so that lookups in the Google English Dictionary can still be performed conveniently from the address bar.

3.4 Procedure

All lookups were performed manually online by the first author, between January 16 and 19, 2011. For each misspelled word, the misspelling was pasted into the search box of each of the dictionaries. In every case, it was noted whether the dictionary was able to identify the correct target word, and, if the dictionary provided a list of alternatives, what was the position of the target word relative to other, irrelevant, hits. The word (or non-word string, as was sometimes the case) presented at the top of the suggestions list was also noted, as well as any other striking suggestions further down the list.

As an illustration of the procedure, consider Figure 1 below, taken from a test lookup in CALD. The intended word was *temporary*, and it was misspelled as **tempori*. The dictionary returned a list of ten suggestions. The top suggestion (number 1 on the list) was *temporise*, which was not the intended word. However, the correct target word *temporary* was found further down the list: in this case it was listed ninth. So, position 9 was noted for this misspelling in CALD.

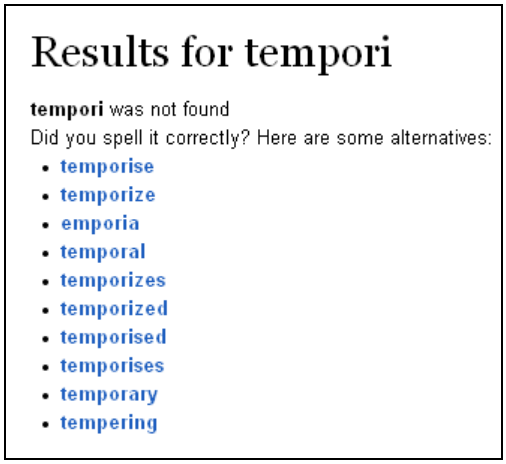

Figure 1: Example suggestions list in CALD for the target word *temporary* misspelled as **tempori*

This example is quite representative of six of the seven dictionaries tested; the exception was GoogleED, which did not provide a longer list of suggestions, but only a single alternative (if any).

Data for all dictionaries and misspellings were keyed into a database and analyzed so as to evaluate the relative performance of the seven dictionaries.

3.5 How well the dictionaries performed

Aggregated results for the complete corpus (i.e. Polish, Japanese, and Finnish) are presented in Table 1 and Figure 2 below. Percentage figures in the table cells indicate what proportion of the 202 target words were found in the respective positions within the individual suggestions lists returned by the dictionaries.

The figures under the heading *First* cover those cases where the target word was presented at the very top of the

suggestions list. *Top 3* means that the target was listed as first, second, or third, and so on. These figures are cumulative, so if a target was listed at the top of the list, it was automatically counted under all four categories (i.e. *First*, *Top 3*, *Top 6*, and *Top 10*). Figure 2 conveys the results in a more visually appealing form.

	Target word listed in position:			
Dictionary	First	Top 3	Top 6	Top 10
LDOCE Free	51%	65%	75%	79%
LDOCE Premium	50%	59%	60%	62%
MWALED	47%	57%	63%	65%
MEDO	25%	44%	51%	54%
CALD	25%	44%	51%	55%
ALD	22%	42%	47%	52%
GoogleED	44%	(44%)	(44%)	(44%)

Table 1: Success rates for the seven dictionaries across all data. Figures indicate the proportion of target words found in the respective positions in the suggestions list.

Two things are immediately obvious in the results: the relatively wide variation between the different dictionaries, and the generally disappointing performance of most of the dictionaries tested. To get some perspective on these figures, it is worth remembering that our corpus of misspellings was designed to be challenging. Unlike some other studies, we did not focus on typos, most of which are simple errors that can be corrected with unsophisticated algorithms. Still, the very wide disparities between the success rates do indicate that at least some dictionaries are not doing the best job possible, to put it mildly.

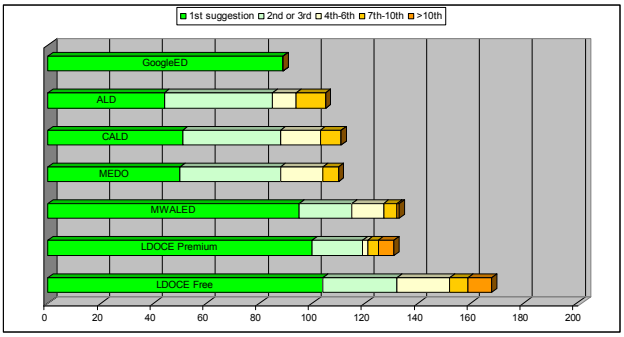

Figure 2: Performance of the seven dictionaries for all data (N=202). Colour bars indicate the number of target words ranked in the respective positions in the suggestions list.

There is a very clear gap between ALD, CALD and MEDO on the one hand and the two versions of LDOCE and MWALED on the other. The first three dictionaries only get between one-fifth and one-fourth of the target words right in the sense of placing the target at the very top of the suggestions list. In contrast, LDOCE and MWALED succeed in guessing the target word about half of the time, with LDOCE being marginally better than MWALED. GoogleED does only slightly worse than LDOCE and MWALED in this respect.

If we now lower the standard and include all suggestions in the top ten, then ALD, CALD and MEDO catch up somewhat, largely thanks to being able to include more of the target words in second or third place (pale green bars in Figure 2). But even with the top ten items on the list included, these three dictionaries only succeed in between 52% and 55% of the cases, which is comparable to the success rate of the better dictionaries for their *first* suggestion only. On the *top ten* measure, MWALED gets slightly ahead of LDOCE Premium, but it is LDOCE Free that really surges ahead, with a lot of accurate guesses in its lists found between the ranks of 2 and 6. It clearly outperforms all the other dictionaries, including — surprisingly — its sister LDOCE Premium. GoogleED has the lowest *top ten* score, but it has effectively thrown in the towel by failing to offer anything beyond the first suggestion.

3.6 Where the dictionaries failed

Since we have access to records of top suggestions offered by the respective dictionaries, it may be interesting to look at some of the problematic cases and offer comments as to what may have caused the less-than-optimal guesses, and how these could have been avoided.

Starting with the ALD, it seems this dictionary attaches too much weight to substring matching. This might explain why it would offer *apology* for **sakology* (a phonetically-motivated misspelling of *psychology*). Apparently, the dictionary homes in on the *-ology*, and then repeats the process with what remains, finding that *ap-* and *sak-* share the letter 'a'. The remaining items on the suggestions list are as follows: *sexology*, *sinology*, *ecology*, *zoology*, *ufology*, *urology*, *geology*, *cytology*, and *tautology*, in this order, and one wonders why *apology* was listed first. In general, ALD does not seem to give much regard to the first letter, even though research has shown that people generally get the first letter right (Yannakoudakis & Fawthrop, 1983; Mitton, 1996). For instance, it offers *masons* for **laysons* (misspelling of *license*), *newbery* for **lajbery* (*library*), and *deferens* for **referens* (*reference*).

A particular oddity of the suggestions served up by ALD, CALD, and MEDO alike is their tendency to offer words with an *–s* at the end, even though there is no indication in the misspelling that one is required. Thus, all three suggest *citizens* for **sitizen*, with the correct *citizen* only appearing in second place. Similarly, we get at the top of the list: *recommends*, *repetitions*, *disappoints*, *forwards*, and even *spaghettis* (for **spagetti*) — that despite the fact that the dictionaries mark the noun as UNCOUNTABLE, and so not usually plural. This mysterious tendency loses the three dictionaries quite a few easy points for top suggestion, at the same time inflating their *top 3* counts, as the reasonable suggestion tends to appear second in such cases. Why would all of ALD, CALD, and MEDO be affected by this overeagerness to tag on *–s*?

Perhaps this has something to do with the software for dictionary compilation and publication that all three use: the DPS Writing System, developed and maintained by the company IDM. However, as far as we know, LDOCE also uses the DPS system, and yet it does not exhibit the *–s* problem.

At times, the suggestions offered by our dictionaries can be downright bewildering. A case in point are MWALED's offerings for **das*, a misspelling of *does*. Admittedly, this is indeed a challenging item, but the suggestions are puzzling, to say the least. The dictionary's output is given in Figure 3 below, and it includes three suggestions: *cream soda*, *giant panda*, and *piña colada*. Only a closer look at the entry can reveal why MWALED should come up with such a list of suggestions. As it turns out, in the comment on form section, the plural for these compounds is given in a traditional compressed form as '~-das', and apparently it is this string that the dictionary has homed in on. Obviously, such a suggestion is a complete red herring. Another surprise from MWALED, though this time with no apparent explanation, is the suggestion *archdiocese* for **ridicyles* (a misspelling of *ridiculous*).

Figure 3: MWALED's suggestions for **das*, a misspelling of *does*

It is difficult to see why MWALED would have a problem with the misspelling **spagetti* — probably the easiest item in the whole corpus, which all the others get perfectly right (except ALD, which only lists the intended word *spaghetti* in second place, following the pluralized *spaghettis*). MWALED offers here no less than 16 alternatives (*spigot*, *spectate*, *spotted*, *spotlight*, *speculate*, *spectacle*, *septet*, *aseptic*, *sabotage*, *septic*, *sceptical*, *sceptic*, *seepage*, *sceptically*, *slippage*, *spatula*), but the obvious *spaghetti* is not among them, even though, to be sure, the entry for it is in the dictionary.

MWALED's algorithm seems to focus excessively on transpositions — it tends to rearrange the original letters: it offers *heir* for **hier* (*here*), *tire* for **trie* (*try*), but also *grade* for **gread* (great) and *crane* for **crean* (*clean*).

Life is made difficult for the spellchecker by the oddity of some of the entries in the dictionary. This is to some extent true of all our dictionaries, but especially of GoogleED. In the absence of any data on word frequency — and it does not seem to be using any — these odd words just enlarge the set of (apparently) plausible corrections, and so we find the following unhelpful suggestions among the 'best' guesses: *deferens* (probably from *vas deferens*), *etyma*, *xylem*, *inf*, *umbrae*, *commis*, as well as proper names like *Du Bois*, *Tok Pisin* and *Wat Tyler*. On top of that, GoogleED would not infrequently provide suggestions that are clearly not genuine words, and often only partially closer than the misspelling to any real English words. Thus, GoogleED offered **petryszyn* for **repetyszyn* (a Polish misspelling of *repetition*), **sejfy* for **sejfty* (*safety*), **trulli* for **truli* (truly), **sinirli* for **sinsirli* (*sincerely*), **temprecher* for **tempreczer* (*temperature*), **bicikli* for **beisikli* (*basically*), **existens* for **egzistens* (*existence*), **identiti* for **aidentiti* (*identity*).

The *–ing* ending seemed to be another cause of difficulty for these dictionaries. Of the lot, only GoogleED is able to correct **useing* to the intended *using*. Instead, LDOCE Free and MEDO offer *unseeing* (true: not entirely unlikely), LDOCE Premium suggests *suing*, MWALED *seeing*, and — strangest of all — CALD proposes the nonce form *useding* (see Figure 4), apparently as a hypothetical inflected form of *used to*, as this is the entry to which the user is taken upon clicking on *useding*.

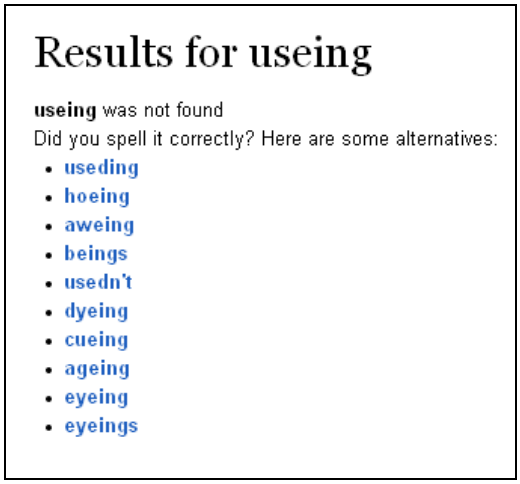

Figure 4: CALD suggestions for the target word *using* misspelled as **useing*

Another easy case is **diging*, a straightforward misspelling of *digging*. As for *useing* above, GoogleED gets it right, and so does ALD this time. LDOCE Free suggests *dining* (and, in third position, *diggings*, but never *digging*!). MWALED insists on *Diegan*, MEDO would like *dinging*, and CALD — *ziging*.

A rather striking feature of LDOCE (especially the free version) is that it likes to make two correct words by sticking a space in the middle of the misspelling, thus: *of fen* for **offen* (*often*), *inter fir* for **interfir* (*interfere*), *so*

rid for **sorid* (*solid*), *back en* for **backen* (*bacon*), *be course* for **becourse* (*because*), *ail and* for **ailand* (*island*). This strategy may be occasionally successful when checking running text, but it does not work well for isolated dictionary query strings, especially if the spellchecker does not care whether the resulting pair is a likely combination.

Apart from that, LDOCE's offerings, among the dictionaries tested, tend to be the most respectful of the misspellings. The suggestions tend to retain the first letter and the general word structure.

4. Can the dictionaries do better? Mitton's experimental spellchecker

As the online dictionaries clearly performed below expectation, the first author wondered if there were context-free spellcheckers capable of outperforming, if not all, then at least some of the dictionaries. As a result of a literature search, a promising context-free experimental spelling correction system was identified (Mitton, 1996, 2009). Consequently, the second author was contacted and offered to run the same data through his spellchecker.

There is no space here to describe Mitton's spellchecker in detail, and the interested reader is invited to consult Mitton (2009) or, for even greater detail, Mitton (1996). At its heart is a dictionary primed with information about the quirks of English spelling. If faced with, say, **morgage*, it would consider *mortgage* a likely candidate because the entry for *mortgage* contains the information that the *t* is likely to be omitted. It also makes use of word frequency in ordering its list of suggestions.

Table 2 compares the success rates (as in Table 1) of Mitton's experimental spellchecker with the best-performing online dictionary (LDOCE Free), and Figure 5 compares it with all the dictionaries graphically.

	Target word listed in position:			
Dictionary	First	Top 3	Top 6	Top 10
Mitton	73%	87%	91%	93%
LDOCE Free	51%	65%	75%	79%

Table 2: Success rates of the best-performing dictionary compared with Mitton's experimental spellchecker, for all data

Mitton's spellchecker was able to place the intended target word among the top ten of its list of suggestions for 93% of the misspellings. The best dictionary in our set, LDOCE Free, performed significantly worse, achieving a success rate of 79%. The gap is even greater if we consider the spellchecker's ability to place the target word in the most valuable top portion of the list of suggestions. Here the experimental spellchecker outperforms LDOCE Free by over 20 percentage points (*First* and *Top 3*). Mitton's spellchecker manages to identify the intended word as the top candidate for 73%, as against 51% for LDOCE Free.

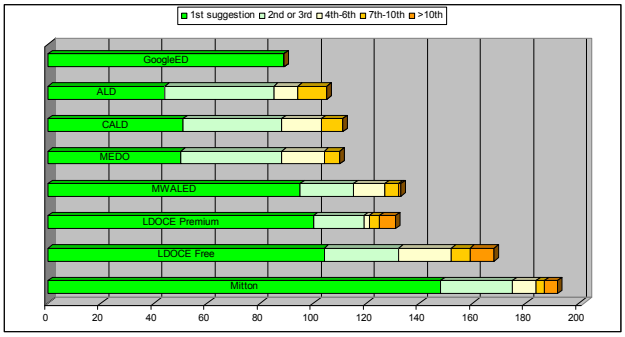

Figure 5: Performance of the seven dictionaries compared with Mitton's experimental spellchecker, for all data (N=202)

In comparison with the other dictionaries (Figure 5), the gains are still greater. The top-of-the-list success rates of ALD, CALD, and MEDO are only a third of that of Mitton's spellchecker. From another perspective, the experimental spellchecker was able to guess perfectly 23 items that none of the seven dictionaries got right.

5. Polish, Japanese, and Finnish misspellings compared

The results we have presented so far are based on aggregated data from the three subcorpora. Now we will take a closer look at the role of the native language.

Our corpus included misspellings from native speakers of three different languages — Polish, Japanese, and Finnish. Figure 6 gives the language-specific success rates in terms of the target word appearing at the top of the suggestions list, while Figure 7 includes percentages for the target word appearing in the top ten of the list.

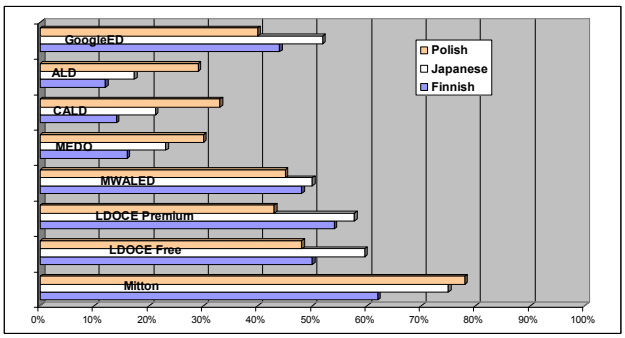

Figure 6: Percentage of Polish, Japanese, and Finnish misspellings for which the target word appeared as the first suggestion

In comparing these language-specific results with aggregated figures, we need to bear in mind that the Polish subcorpus contributed the most to the overall figures, as it represents half the data, with the Japanese and Finnish subcorpora accounting for a quarter of the corpus each. In terms of the target word being offered as the best suggestion, four systems (ALD, CALD, MEDO, and Mitton's spellchecker) seem to cope better-than-average with the Polish misspellings, while for the remaining four (GoogleED, MWALED, LDOCE

Premium, and LDOCE Free), the reverse is the case. Since the Polish data were elicited via audio stimuli, this may have to do with the inclusion or otherwise of phonological awareness, explicit or implicit, rather than with specifically L1-induced misspelling patterns. Still, it is also true that part of the Finnish data came from written responses to spoken dialogue, and the Japanese misspellings were partially inspired by their katakana representations, so it might be said that all three subcorpora had some sound-motivated misspellings.

What is clear, however, is that ALD, CALD, and MEDO would have done even more poorly overall, had the Polish subcorpus not been given more weight than the others: their success with the Finnish misspellings was only half — at best — of that with the Polish data, with the Japanese figures in-between the two.

In placing the required word at the top of the list, Mitton's spellchecker did very well with the Polish and Japanese data, and not quite as well with the Finnish misspellings (though it still outperformed all the dictionaries).

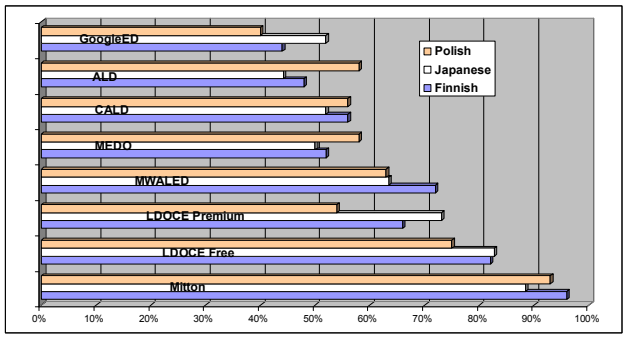

Figure 7: Percentage of Polish, Japanese, and Finnish misspellings for which the target word appeared among the top ten suggestions

Moving on now to the results for the top ten suggestions (given in Figure 7 above), we can see that there is no longer that much difference due to the native language, even for ALD, CALD, and MEDO. Apparently, the three dictionaries can still capture about half of the target words in the top ten suggested items, though somehow they find it much harder for Polish to guess the best suggestion correctly than for the other two languages.

Mitton's spellchecker performed reliably for all three languages, getting 93% of the Polish target words in the top ten, and no less than 96% of the Finnish items (actually, all of the 96% also made it into the top six suggestions).

6. Ways to improve spelling correction in e-dictionaries

6.1 Customization

While for many years the primary focus of research into spelling correction has been on native writers, recently the needs of non-native users of language speakers, particularly English, have begun to receive some attention (for an overview, see e.g. Heift & Rimrott, 2008). It is now recognized that the patterns of misspelling of non-native speakers differ both in quality and quantity from those of native users of a language. Thus, if the L1 of the user is known to the system (be it based on the Accept-Language http header, IP Geolocation, or individual user profile), the dictionary interface might use an algorithm optimized for that native language. In fact, Mitton's spellchecker used in this study has already seen a successful adaptation to better handle the typical misspellings of Japanese learners writing in English (Mitton & Okada, 2007).

However, we would not expect the influence of L1 to be uniform across a wide range of L2 proficiency levels. To account for this variation as well as for individual idiosyncrasies, customization might in the future go even further: it might be possible to design an adaptive spelling corrector, capable of tuning in to the particular areas of spelling problems exhibited by a given user.

6.2 Greater phonological awareness

A large proportion of the items at which all the dictionaries failed are recognizable as attempts at rendering the pronunciation of the English word through the spelling conventions of the native language. This is particularly evident in the case of the Polish data, no doubt partially as a result of using audio stimuli for data elicitation. Evidence for this 'phonetic access' strategy (here largely subconscious, cf. Sobkowiak, 1999) is seen in the use of L1-specific letter combinations (such as, for Polish, <sz>, <aj>, or <ej>) to approximate English pronunciation. Mitton's spellchecker handled many of these cases quite well, perhaps thanks to its level of phonological awareness, even though it has not been made aware of any Polish-specific letter-to-sound correspondences. Making provision for a few of the most common such correspondences could significantly improve a spellchecker's performance.

6.3 Dealing with real-word errors

In section 2.2 we discussed the issue of rare words. To use a specific example from the study, one of the misspellings in the corpus was **wold* for *would*. As it turns out, *wold* is also an English word, albeit very rare. Consequently, most occurrences of *wold* in text will be misspellings, and a text spellchecker would do well to flag it as a possible error. However in a dictionary look-up situation, unlike in text spellchecking, it would be risky to withhold a rare-word entry from the user and offer instead similarly-spelled frequent words. Even though the core vocabulary of a few thousand words (De Schryver et al., 2006) are looked up more commonly than the rest, it is also true that the less frequent items have a reasonable chance of being looked up (see the discussion in 2.2 above). How should a dictionary respond to such a query?

The answer need not necessarily be the same for *any* dictionary. A user of the online version of, say, the OED is much more likely to want an entry for this obscure word than a user of an intermediate learners' dictionary. The

latter dictionary might not hold the word in its wordlist at all, in which case the issue would not arise. But if it did, a happy compromise might be to take the user to the rare word entry, but at the same time alert them in a sidebar saying something like 'Did you perhaps mean *world*'?

6.4 First things first

We have suggested possible avenues to improve success in correcting misspelled dictionary search terms. However, it needs to be stated emphatically that it would be misguided to pursue any such attempts at tweaking the interface before more basic problems are addressed. This study has revealed that such fundamental problems are numerous and grave, and they affect the most authoritative of English monolingual learners' dictionaries.

7. Conclusion

Our study has shown that the leading monolingual English learners' dictionaries are inadequate when it comes to correcting misspelled input from non-native users. When challenged with a misspelling, far too often the dictionaries are unable to include the word actually intended in their list of suggestions, and if they do include it, the ordering of the alternatives is often less than optimal. While the individual dictionaries vary substantially in performance, there is much room for improvement for even the best ones, and we have shown that an experimental spellchecker achieves much greater success rates than any of the dictionaries, even though it has not been designed with non-native speakers in mind.

8. Acknowledgements

The first author wishes to thank his student assistants, Marta Dąbrowska and Aleksandra Lasko, for their help in collecting the Polish corpus of misspellings.

9. References

9.1 Online dictionaries tested

ALD. *Oxford Advanced Learner's Dictionary*. http://www.oxfordadvancedlearnersdictionary.com/

CALD. *Cambridge Advanced Learner's Dictionary*. http://dictionary.cambridge.org/

GoogleED. *Google English Dictionary*. At the time of collecting data: http://www.google.com/dictionary; at the time of writing the present version: http://www.google.com/search?q=%s&tbs=dfn:1 (where %s stands for the search term)

LDOCE Free. *Longman Dictionary of Contemporary English*. http://www.ldoceonline.com/

LDOCE Premium. *Longman Dictionary of Contemporary English*. http://www.longmandictionariesonline.com/

MEDO. *Macmillan English Dictionary Online*. http://www.macmillandictionary.com/

MWALED. *Merriam-Webster's English Learner's Online Dictionary*. http://www.learnersdictionary.com/

9.2 Other references

Atkins, B.T.S. (1996). Bilingual dictionaries - past, present and future, in Gellerstam, M., Jarborg, J., Malmgren, S.-G., Noren, K., Rogström, L. & Papmehl, C.R. (eds.), *EURALEX '96 Proceedings*. Göteborg: Department of Swedish, Göteborg University, pp. 515-46.

Bank, C. (2010). Die Usability von Online-Wörterbüchern und elektronischen Sprachportalen. M.A. Thesis, Universität Hildesheim.

Bogaards, P. (2010). The evolution of learners' dictionaries and *Merriam-Webster's Advanced Learner's English Dictionary*, in Kernerman, I., Bogaards, P. (eds.), *English Learners' Dictionaries at the DSNA 2009*. Tel Aviv: K Dictionaries, pp. 11-27.

Damerau, F.J. (1964). A technique for computer detection and correction of spelling errors, *Communications of the A.C.M.* 7, pp. 171–76.

De Schryver, G.-M., Joffe, D., Joffe, P. & Hillewaert, S. (2006). Do dictionary users really look up frequent words? – On the overestimation of the value of corpus-based lexicography, *Lexikos* 16, pp. 67-83.

Deorowicz, S., Ciura, M. (2005). Correcting spelling errors by modelling their causes, *International Journal of Applied Mathematics and Computer Science* 15(2), pp. 275-85.

Hanks, P. (2009). Review of Stephen J. Perrault (ed.). 2008. *Merriam-Webster's Advanced Learner's English Dictionary*, *International Journal of Lexicography* 22(3), pp. 301-15.

Hanks, P., Pearsall, J. (eds.). (1998), *New Oxford Dictionary of English*. Oxford: Oxford University Press.

Heift, T., Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL, *System* 36(2).

Kukich, K. (1992). Techniques for automatically correcting words in text, *Computing Surveys* 24(4), pp. 377-439.

Lew, R. (2011). Online dictionaries of English, in Fuertes-Olivera, P.A., Bergenholtz, H. (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, pp. 230-50.

Lindberg, C. (ed.), (2006), *Oxford American College Dictionary*, 2nd edition. Oxford: Oxford University Press.

McKean, E. (ed.), (2005), *New Oxford American Dictionary*, 2nd edition. Oxford: Oxford University Press.

Mitton, R. (1985). Birkbeck spelling error corpus. Accessed at: http://ota.oucs.ox.ac.uk/headers/0643.xml.

Mitton, R. (1996). *English spelling and the computer*. Harlow: Longman.

Mitton, R. (2009). Ordering the suggestions of a spellchecker without using context, *Natural Language Engineering* 15, pp. 173-92.

Mitton, R., Okada, T. (2007), The adaptation of an

English spellchecker for Japanese writers. *Symposium on Second Language Writing*. Nagoya, Japan.

Nielsen, S. (1995). Alphabetic macrostructure, in Bergenholtz, H., Tarp, S. (eds.), *Manual of Specialised Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 190-95.

Pollock, J.J., Zamora, A. (1984). Automatic spelling correction in scientific and scholarly text, *Communications of the A.C.M.* 27(4), pp. 358–68.

Proctor, E. (2002). Spelling and searching the internet: An overlooked problem, *The Journal of Academic Librarianship* 28(5), pp. 297-305.

Sobkowiak, W. (1999). *Pronunciation in EFL machine-readable dictionaries*. Poznań: Motivex.

Suomi, R. (1984). Spelling errors and interference errors in English made by Finns and Swedish-speaking Finns in the 9th form of comprehensive school. MA Thesis, Abo Akademi.

Yannakoudakis, E.J., Fawthrop, D. (1983). The rules of spelling errors, *Information Processing and Management* 19(2), pp. 87-99.

Appendix: Sample misspellings

SUBCORPUS	TARGET	MISSPELLING
PL	certain	serten
PL	easily	izli
PL	guarantee	garanti
PL	interfere	interfir
PL	interruption	interapsion
PL	library	lajbery
PL	psychology	sakology
PL	receive	reseve
PL	separate	sepret
PL	succeed	sukcid
JP	albatross	albatlos
JP	antenna	untena
JP	beautiful	butiful
JP	embarrass	enbarance
JP	enough	inaf
JP	gallery	garally
JP	graph	glaf
JP	laughter	lafter
JP	neglect	nigrect
JP	umbrella	umblera
FI	because	becourse
FI	colour	coulor
FI	delicious	delecous
FI	especially	espessially
FI	gasoline	gazolin
FI	good-bye	goodbay
FI	orchestra	orkester
FI	symphony	sinfony
FI	temperature	tempeture
FI	universities	univercitys