SPELLING MISTAKES: HOW WELL DO CORRECTORS PERFORM?

D. G. Hendry & T. R. G. Green
MRC Applied Psychology Unit
15 Chaucer Rd., Cambridge, England CB2 2EF
david.hendry@mrc-apu.cam.ac.uk

ABSTRACT
Commercial spelling correctors were tested on *mistypings* and *misspellings*. Mistypings were 'corrected' more successfully. Success rates for misspellings covered a fair range, but it is hard to quantify comparisons between correctors, and an accepted evaluation procedure is urgently needed. Improved correction techniques would benefit foreign speakers and poor spellers.

KEYWORDS: spelling correction, word processors

INTRODUCTION

In this short paper we ask two questions. First, how well do spell checkers correct spelling mistakes? This is almost completely ignored in product reviews, which focus on superficiae and on feature lists. Although spelling correction is a prime example of successful practical interaction between people and computers, we know of no published HCI-style evaluations of recent spelling correctors, nor of a standard evaluation procedure, nor even of any serious discussion of the problems of evaluation. Second, do spell checkers need to conform better than they do at present to the human cognitive processes involved in misspelling?

BACKGROUND

What's Important in a Spell Checker?

Of course it's important that a spell checker be fast, that the dictionary be large enough (but note that the larger the dictionary the more false-positives, and that will not be appropriate for all users), and that technical words can be added to a supplementary dictionary. However, for many people (e.g., poor adult spellers, children learning to spell, and those writing in a second language) by far the most important feature is correction performance.

Correcting Spelling Mistakes

There are two fundamental types of spelling mistake: *misspellings* and *mistypings*. Misspellings occur because the writer does not know the correct spelling (e.g., *sureptitous* for 'surreptitious' is a standard mistake [2], *philpant* for 'flippant' is a mistake made by poor adult spellers [1], and *sanderstom* for 'thunderstorm' is often made by Japanese speakers of English [4]). Mistypings, however, are "hitting the keys wrong" (e.g., *flippatn*).

The 'words' produced by mistypings are often very different from the 'words' produced by misspellings. A spelling corrector has to work backwards from the error and infer the intended word. Different algorithms will be needed for different types of error.

Mistypings are the simplest form of error to correct. An influential paper by Damerau [2] reported that a very simple algorithm would 'correct' 80% of spelling mistakes from a small corpus. In this algorithm the error was transformed by inserting, deleting, or changing a letter or by transposing two adjacent letters: if the result was in the dictionary it was offered as a correction. Subsequently, many spelling correctors appear to have been extensions of Damerau's algorithm. The algorithm can be seen as a simple model of mistyping.

Misspellings are harder to correct. The simplest theory of spelling mistakes is 'spelling by ear': in lieu of the conventional spelling, the writer types a string that will achieve the correct sound. Thus we see *mearly* for 'merely', *crepped* for 'crept'. But observation of real misspellings also shows "spelling by eye" [3], where some letters of the correct spelling are recalled, possibly inappropriately (e.g., *yatch* for 'yacht', pronounced 'yot').

Some commercial correctors make claims such as "offers alternate spelling using phonetic rules which look up similar sounding words" (quotation from WordFinder publicity): they appear to rely on phonetic equivalence of letters, so that *jeraf* could be matched with 'giraffe'. But examples of 'spelling by eye', we anticipated, would not be correctable by this type of algorithm, and we therefore predicted a low success rate on them.

CORRECTION PERFORMANCE

Method of Testing

We carried out performance tests of eight spelling correctors [5]. Relying partly on published work and partly on our own experience we prepared six word lists of spelling errors (see Figure 1) and tested the spelling correctors on the lists. Here, though, only a summary of correction performance is given. (We also examined the plausibility of the corrections, the extent to which similar spelling errors generate similar lists of corrections, and other issues [5]).

Caveat. To test the thesis that better cognitive models were needed, the words in lists C-D-E were selected to illustrate interesting forms of misspelling. Hence, although these are real mistakes made by real people, the

Table 1. Proportion of errors corrected for each of the word lists of Figure 1. (These figures do NOT represent absolute success rates: see caveat in text.)

Word-Proc.	Manufacturer	A(N=20)	B(N=22)	C(N=7)	D(N=33)	E(N=22)	F(N=23)
FullWrite (1.1)	Ashton-Tate	.90	.59	.57	.82	.73	.35
Nisus (3.06)	Paragon Concepts	.90	.64	.57	.85	.91	.35
Word (4.0b)	Microsoft Corp.	.70	.41	.43	.73	.82	.39
Word (5.0)	Microsoft Corp.	.90	.50	.43	.67	.86	.39
WriteNow (3.0)	T/Maker Company	.85	.05	0.0	.52	.14	.30
WordFinder (224)	SelecTronics	.85	.50	.29	.85	.68	.30
Average (Average excluding WriteNow)		*.85 (.85)*	*.44 (.53)*	*.38 (.46)*	*.74 (.78)*	*.67 (.80)*	*.35 (.36)*

figures in Table 1 do not represent absolute correction performance for those people.

Results

The spelling correctors performed well on mistypings (Table 1, column A), less well on misspellings (cols. C-D-E), rather poorly on Japanese misspellings (col. F). Their claims of phonetic matching were partly vindicated (col. B) but context sensitive phonology (e.g., -ci- vs. -co-) seemed to be a weak point. To our surprise, phonetic spellings taken from dictionaries (e.g., *krimzn* for 'crimson') were frequently correctable. And contrary to expectations, the correctors also managed better on examples of 'spelling by eye' (col. D); for example, *yatch* was successfully corrected by two correctors.

CONCLUSIONS

Metric Needed. A recent advertisement [6] compared the Nisus and WriteNow spell checkers. The size of the dictionaries and such features as supplementary dictionaries were the only bases of comparison. In the same issue a reader complains about a previous review: "You give WriteNow a poor rating for its spelling checker even though its dictionary is larger than ...". Table 1 shows that while Nisus 3.06 and WriteNow 3.0 are about equal for correcting mistypings, Nisus is far superior for misspellings—a critical fact for many consumers. Clearly, a metric is needed. Without one, manufacturers can't measure progress, reviewers can't adequately inform consumers on one of the most important features of a word processor, and users can't choose between fast 'typo' correction versus slower misspelling correction.

A) Mistypings. Words containing typing slips (e.g., *dsign* for 'design', *tghe* for 'the', *suvjects* for 'subjects').

B) Phonetic input. Misspellings that are phonetically plausible (e.g., *chelo* for 'cello' and *goozberi* for 'gooseberry').

C) Misspellings from teenagers (e.g., *spieces* for 'species', *disearppearing* for 'disappearing').

D) Misspellings from poor adult spellers [1], mainly 'spelling by eye' (e.g., *alcahmay* for 'alchemy', *vaccumn* for 'vacuum').

E) Common misspellings made by secretaries that Damerau's [2] original algorithm could not correct (e.g., *phamplet* for 'pamphlet', *sissers* for 'scissors').

F) Misspellings made by Japanese users of English [4] (e.g., *Octover* for 'October', *plathome* for 'platform').

Figure 1. Summary of word lists used.

Cross-language Correctors Needed. Most of the Japanese-English errors in [4] were simple phonetic confusions. Although they defeated our sample of correctors, as pointed out in [4], a specialised corrector would be easy to build and would probably meet a growing need.

Better Help for Poor Spellers. Specialised correctors could help users to disambiguate the desired word from other possible corrections both by speaking it and by providing a short definition.

Cognitive Models Good Enough? The unexpected success in columns C and D suggests that the simple-minded approach turns out to be fairly adequate. However, superior techniques will be required if the goal is to *completely* support users in spelling correction at the computer alone, without recourse to bleary-eyed search through dictionaries. Correction algorithms will need to be matched to particular types of error; designing them means that both psycholinguists and computer scientists need to get together.

Finally, *a moral for HCI*: HCI is fond of complaining that psychology lags behind technology. But, here it is the opposite: HCI is lagging behind in the evaluation of spell checkers just as it has failed to investigate the spreadsheet, another overwhelming popular invention. And, *a need from applied linguistics*: more corpora of spelling errors from user populations are needed.

REFERENCES

[1] Burden, V. (1990) An investigation of persistent spelling difficulties in "normal" adult readers. Unpublished PhD thesis: Cambridge University.

[2] Damerau, F. J. (1964) A technique for computer detection and correction of spelling errors. *Comm. ACM*, 7 (3), 171-176.

[3] Ellis, A. W. (1984) Spelling and writing. In A. W. Ellis (Ed.) *Reading, Writing and Dyslexia: a Cognitive Analysis.* London: Erlbaum.

[4] Furugori, T. (1990) Improving spelling checkers for Japanese speakers of English. *IEEE Trans. on Professional Communication*, 33(3), 138-142.

[5] Green, T. R. G and Hendry, D. G. (in preparation) Spelling correctors and spelling mistakes.

[6] *MacUser* (Dec/1992). Advertisement, p. 10.