Journal of the Simplified Spelling Society

No.20—1996/1

Contents of this Issue

The Society

Founded 1908, the SSS works to modernize English spelling for the benefit of learners and users world-wide. It currently has members and associates on four continents, focuses research from many relevant disciplines, and campaigns to educate and influence public and political opinion.

Inquiries to Membership Secretary Jean Hutchins,

Subscription £10/US$20 or equivalent per year.

The Journal

The *Journal* appears twice yearly and is edited by Christopher Upward,
Dept. of Languages & European Studies, Aston University, Birmingham
American contributing editor Kenneth H Ives,
To submit material for publication, see guidelines inside back cover.

[Journal of the Simplified Spelling Society, 20, 1996/1 p2 in printed version]
[Chris Upward: see **Journals**, **Newsletters**, **Pamflet**, **Leaflets**, **Media**, **Book and Papers**.]

1. Editorial

Chris Upward

Spanning oceans and continents

With this issue, production of the *JSSS* returns to the UK, after four successful years under Ken Ives' editor-ship in the USA. Small the SSS may be, but it truly spans the oceans and continents, as indeed a movement for reforming the spelling of world English must.

The new information technologies, now developing with breathtaking speed, suggest new opportunities for internationalizing our operations, as Bob Brown (see our Tribute to him) hinted in his last ***Newsletter*** (Items 3&6). At his last AGM, he looked round at the essentially southeast English members of the Committee, and reflected that in the future our international Society might no longer need such a geographically tight nucleus. Could our policy decisions before long be taken via the Internet by committee members spanning the world? The technology is already there, only waiting to be applied.

Since its **foundation in 1908** the SSS has been conscious of the international implications of its work (hence its collaboration with the American Simplified Spelling Board in the first half of this century). And in recent years SSS members have travelled between America, Australia, Europe, India, Japan and New Zealand, networking with reformers in other countries and continents. Building on this, the Internet may soon permit the international co-ordination of our lobbying, so that an essentially united message can be presented to whatever governmental or other bodies we may wish to influence around the world.

American spellings as world standard?

In the past, the SSS's credibility long suffered from the natural and necessary concentration by members on their own reform proposals, without offering the public clear guidance as to which might in practice deserve priority. The role the Society has more recently adopted, as a forum for airing ideas and research findings, transcends that earlier individualism and earns wider respect; but while it can now more effectively educate public opinion on spelling reform questions, it currently offers no simple answer to the burning question "What, precisely, is to be done?"

A possible and promisingly practical answer was unintentionally suggested this summer by the head of the English School Curriculum and Assessment Authority. He expressed concern at British children being exposed to American spellings in educational software. The Society's Committee felt thereby provoked into sending him an analysis of the advantages of American spellings (to appear in **JSSS 97/1** Item 12). Not merely are they nearly all orthographically and psychologically preferable, but their adoption in Britain would entail none of the complications of worldwide implementation of other first-step proposals. No international agreement would be needed; other countries would have every incentive (but no compulsion) to follow suit (Australia and Canada are perhaps halfway there already); there would be little uncertainty about which forms were involved; the British end of the SSS could concentrate on lobbying the British authorities, leaving members in other countries to lobby theirs independently; and a clear sign would be given, to make the world realize that improvements to English spelling are perfectly feasible. If the world agreed on American spellings, we could target real reforms from then on.

Features of this issue
JSSS 96/1 marks the start of the new (second British) series in several ways. The cover has been updated so that the title matches our present housestyle, and the analysis of spelling anomalies is more explicit and comprehensive than before. This analysis follows British Received Pronunciation, and readers are encouraged to send in their observations on words listed that are at variance with their own pronunciation (eg, does geyser rhyme with *freezer* or *wiser*?).

This issue also celebrates a decade of SSS *Newsletters* and *Journals* with two special features. One is a full index for the period 1985–95. The other records nearly 90 years of publishing by the SSS with a joint bibliography and sales catalogue.

We further offer the usual wide variety of articles and other contributions, including Roger Mitton's illuminating account of how spellcheckers work, two contrasting papers on research into psychological and educational aspects of English spelling, two on the spelling of other languages, and some speculations on the marketability of spelling reform.

Planned for 97/1
The special features in this issue unfortunately crowded out what we hope will become a regular item: readers' letters. Readers are therefore encouraged to write in (by email, fax, or snailmail) with their reactions to articles or their views on spelling reform questions generally.

Among the contributions expected for 97/1 is a report on the much-heralded and now at last confirmed reform of German spelling, which many readers have been wanting to find out more about. This event offers a lesson to the English-speaking world on the normality and feasibility of spelling reform, and it should form a major plank in our propaganda. A propos of which: a recent German government press release was entitled: "A solid education—the primary school system". How many governments of English-speaking countries could claim as much?

[Journal of the Simplified Spelling Society, 20, 1996/1 p3 in printed version]
[Bob Brown: see **Journals**, **Newsletters**, **Pamphlet13**, **PV1**]

2. Bob Brown: a Tribute

It was with profound shock that the Simplified Spelling Society learnt of the tragic, sudden death in Zürich at the age of 51 of its Secretary, Bob Brown, on 13 June 1996. Seven committee members and SSS President, Professor Donald Scragg, paid their last respects at Bob's funeral on 28 June. If the SSS has enjoyed something of a revival in the 1990s, much of the credit must go to Bob, whose contribution to our effectiveness was major in several important respects.

Born and educated in Cheshire in the northwest of England, Bob was involved during his career with television, telecommunications, information technology, management, and marketing, and he wrote several books for computer users. His involvement with the SSS antedated that of any other member of the present committee. He first joined in the early 1970s (we have his signature on the attendance lists for the AGMs of 1973, 1974 and 1975), in the aftermath of the dispute as to whether the Initial Teaching Alphabet or New Spelling should be the SSS's flagship system. However, Bob's work often took him overseas, with periods in Africa, Denmark and from 1987 Japan, and it was not until 1990 that, on returning to the UK, he was able to resume his active connection with the SSS. With Laurie Fennelly's retirement from the

post of Secretary imminent, Bob's offer to step into the breach was gladly accepted, as from April 1991. His willingness to combine the demanding roles of General Secretary and Membership Secretary was particularly appreciated.

Bob's efficiency, energy and enthusiasm were at once apparent. Meetings were meticulously planned and minuted. With his friendly, patient and relaxed manner, he encouraged and gave support to many members in presenting their contributions to best advantage. His production of the **'Personal View'** series of papers gave an outlet for members to publish their ideas, a process further refined by his assistance to Paul Fletcher in developing a systematic Spelling Evaluation procedure. The ***Newsletters*** Bob produced were models of typographical professionalism, and full of lively, practical and forward-looking ideas that complemented the more academic *Journal*, as well as showing his own talent as a researcher. His urging expedited the revision of the ***Cut Spelling Handbook*** and publication of its second edition. His experience of legal and financial affairs enabled him to push the Society's case for charitable status and, in conjunction with SSS Research Director Gwenllian Thorstad, to advise on grant applications. His expertise in information technology and current business practice, as well as his international experience, helped the SSS raise its publishing standards and establish a growing network of email contacts around the world and a presence on the Internet. He would have had much more to give in that sphere if he had lived. All in all, Bob personified the interdisciplinary nature of the spelling reform project.

Bob had an original mind, but his success in giving coherent guidance to the disparate independent spirits that make up the SSS was no doubt partly due to his reticence in putting forward his own ideas. Unlike many members, he was not wedded to a single scheme, indeed his own views of what might be the most desirable or practical kinds of spelling reform were rarely if ever articulated. Yet this very detachment fitted him perfectly for the SSS's role today, which is to debate and research contrasting concepts of spelling reform and to propagate a raft of possibilities (the SSS's insistence on what it saw as the only possible ideal scheme in the past had proved a recipe for fruitless dissension). It was this same detachment, as well as his wider linguistic interests (which included Basic English and E-prime, two systems for ensuring clear, simple expression) that enabled him to produce one of the Society's most valuable publications, his survey ***Spelling Reform in context****: a typology, list & bibliography of English spelling reforms*.

As it happened, Bob was on the brink of resignation as Secretary when he died, as his work was again about to send him abroad, this time to Switzerland, and some steps had already been taken to ensure the SSS's work would continue smoothly after his departure (Jean Hutchins takes over as Membership Secretary and Ron Footer as Meetings Secretary). But Bob had hoped to maintain his links with the SSS (via email, at the very least), and no one dreamt that he had made his last contribution to its cause. Bob was in the prime of life, and the SSS feels his loss not only for its own functioning, but also for the frustration of his own hopes and ambitions.

Our deepest sympathy goes out to Bob's widow Atsuko, with whom he was developing an Anglo-Japanese translation business. Those of us who worked most closely with him were well aware of her supportive presence, and indeed she was collaborating on an Anglo-Japanese literacy project run by SSS Chairman Chris Jolly. English spelling reform is a world enterprise, and it would be good to think of her as part of its Japanese dimension.

We reflect that many generations have passed in pursuit of English spelling reform, and that more may yet pass before a significant breakthrough is achieved. Bob gave much to the process during the years he was with us, and our greatest tribute to him would be to try and emulate the skills and wisdom he displayed.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp4–11 in printed version]
[Roger Mitton: see **Newsletter N9**]

3. Spellchecking by Computer

Roger Mitton

Dr Mitton is lecturer in Computer Science at Birkbeck College, University of London. His research interest is the development of a computer spellchecker for people whose spelling is poor. This article, which was the basis of an address given to the Simplified Spelling Society in April 1995, reviews the methods that have been applied in computer spellchecking. It derives from a chapter in his recent book *English Spelling and the Computer,* published by Longman, 1996.

Abstract

By the standards of the computer industry, spelling correction has a long history; people have been writing programs to detect and correct spelling errors for over thirty years. Reviews of the literature are provided by Peterson (1980a, 1980b) and Kukich (1992a). In this article I sketch the main methods and some of the unsolved problems. I will simplify the descriptions so as not to get bogged down in the detail. First I discuss methods for detecting errors; then I discuss methods for suggesting corrections.

1 Detecting errors

The most popular method of detecting errors in a text is simply to look up every word in a dictionary; any words that are not there are taken to be errors. But before I describe variations on this method, I will mention two that do not use a dictionary in this way.

1.1 Spellcheckers without dictionaries

The first uses a dictionary indirectly (Riseman & Hanson 1974). It begins by going right through the dictionary and tabulating all the trigrams (three-letter sequences) that occur; *abs*, for instance, will occur quite often (*absent*, *crabs*) whereas *pkx* won't occur at all. Armed with this table, the spelling checker divides the text into trigrams and looks them up in the table; if it comes across a trigram that never occurred in the dictionary, the word that contains it must be a misspelling. It would detect *pkxie*, for example, which might have been mistyped for *pixie*. **[1]** For detecting people's misspellings, this technique is of limited value since a high proportion of errors do not contain any impossible trigrams, but it is of some use in detecting errors in the output of an optical character reader (a machine that scans a page of text and 'reads' the letters).

The second does not use a dictionary at all (Morris & Cherry 1975). Like the previous method, it divides the text into trigrams, but it creates a table of these, noting how often each one occurs in this particular piece of text. It then goes through the text again calculating an index of peculiarity for each word on the basis of the trigrams it contains. Given *pkxie*, for instance, it would probably find that this was the only word in the text containing *pkx* and *kxi* (and possibly *xie* too), so it would rate it highly peculiar. The word *fairy*, by contrast, would get a low rating since *fai*, *air* and *iry* probably all occur elsewhere, perhaps quite often, in the passage being analysed. Having completed its analysis, it draws the user's attention to any words with a high peculiarity index. Like the previous method, it would fail to spot a high proportion of ordinary spelling errors, but it is quite good at spotting typing errors, which it was designed for. An advantage that it has

over all dictionary-based methods is that it is not tied to English; it will work on passages of, say, French, German or Greek.

1.2 Saving space

The majority of spelling checkers, however, use a dictionary in some way. I say 'in some way' because they do not necessarily hold a complete dictionary with all the words spelt out in full, though some do. Some economize on storage space by holding only the stems of words (McIlroy 1982). For example, instead of holding *doubt, doubts, doubted* and *doubting*, they hold just *doubt* and use a set of rules to remove suffixes before looking words up; given *doubting*, the checker would remove the *ing* and look up the *doubt*. They may remove prefixes also (*undoubted*) and they may carry on removing suffixes (or prefixes) until they reach a stem (*undoubtedly*). The process is known as 'affix-stripping'.

The rules have to be a bit more complicated than this in order to cope with such forms as *cutting* (to get *cut* rather than *cutt*), and *denied* (to get *deny* rather than *deni*). The rules have to have some ordering, so as to accept *undoubtedly* but not *undoubtlyed*, and they need to have some way of coping with words that look like inflected forms but aren't, such as *farthing*. The strength of this system is that the checker can accept freshly minted words that are acceptable but are possibly not in any dictionary, such as *unplaceable*. The weakness is that it will accept some words that don't exist, such as *undoubt*.

A second way to save storage space is to hold the dictionary as a bit map (McIlroy 1982, Nix 1981). Imagine the memory of a computer as a long row of lightbulbs, initially all switched off. You go through the dictionary and convert each word, somehow, into a number. For example, you might start by converting a to 1, b to 2, c to 3, and so on; the word *ace*, for example, would become 1,3,5. Then multiply the first number by 1, the second by 2 and so on, and add them up; 1,3,5 gives (1x1)+(3x2)+(5x3) = 22. Finally, multiply by 10 and add the number of letters in the word: (22x10) + 3 = 223. Now you go to the 223rd lightbulb and switch it on. After you've done this for every word in the dictionary, some of the lightbulbs are on and the rest are still off.

Now you are ready to do a spelling check. You take each word of the text and convert it to a number by the same process you used before; if you came across the word *ace*, you'd convert it to 223. You look at the appropriate lightbulb. If it's on, the word is acceptable; if it's off, it isn't. So, *ace* (lightbulb 223) is accepted. *Ade*, by contrast, would be converted to 243; the 243rd lightbulb would be off, so *ade* would be rejected.

The long line of lightbulbs is the 'bit map', an array of thousands of binary digits (0's and 1's). Converting a word to a number is known as hashing, and the method you use is a hashing function. The hashing function described above is too simple to do the job properly — *dcd, hdb* and various other non-words would all hash to 223 and be accepted — but it's possible to devise more complicated hashing functions so that hardly any non-words will be accepted. You may use more than one hashing function; you could derive, say, six numbers from the same word and check them all in the bit map (or in six separate bit maps), accepting the word only if all six bits were set.

If there is no need to save storage space, the dictionary can be held as a straightforward list of words, inflected forms included. The computer looks a word up in much the same way as a person looks one up in a printed dictionary. The words can be stored in alphabetical order, so the computer can go straight to the right place and check if it's there or not.

1.3 Real-word errors

There are two ways in which a spelling checker can fail: it may flag a word as an error when in fact it's correct, or it may let an error slip through. Obscure words and proper names are the cause of the first problem. The frequency of these false alarms can be reduced by having a larger dictionary or a specialized one for particular kinds of text, such as a dictionary with lots of medical terms, but there is no real solution to the problem of proper names — there are just too many of them. Many checkers have a facility whereby the user can build up a private supplement to the dictionary, to prevent the checker from constantly flagging names that crop up often in the user's documents. False alarms, though irritating, may be acceptable in moderation since the user can always ignore the checker's output; but the second problem — letting errors slip through — is more worrying since the user cannot be sure that a passage is error-free even when the checker has gone over it. The problem arises because some misspellings match words in the dictionary, as in '*Their* she goes', 'The *wether* was glorious', or 'The Continental restaurant company is developing a chain of French-style *brassieres*'. [2] I call these 'real-word errors'.

Unfortunately the problem gets worse as the dictionary gets larger; including more obscure words in the dictionary, to reduce the number of false alarms, increases the risk of missing real-word errors. The word *wether* illustrates this. The word is, arguably, so obscure that any occurrence of *wether* in a passage is more likely to be a misspelling of *weather* or *whether* than a genuine occurrence of *wether*, so a checker that did not have the word in its dictionary would do better than one that did.

Drastic pruning of the dictionary, however, is not a solution; a checker with a small dictionary raises too many false alarms. A recent study has shown that, when an uncommon word occurs, it is far more likely to be a correct spelling of a rare word than a misspelling of some other word (Damerau & Mays 1989). This may not be true of some highly obscure words that resemble common words, such as *yor* and *stong*, so perhaps some judicious pruning is advisable. Nor is it true of certain medium-rare words that occur commonly as misspellings of other words, such as *cant* and *wont* which are often misspellings of *can't* and *won't*; these seem to require special treatment. But, with these provisos, big is beautiful for a checker's dictionary.

How serious is the problem of real-word errors? At first sight, it appears to be only marginal; the proportion of words that can be changed into another word by a small typing slip, such as *whether* into *wether*, is only about half of one per cent. However, the proportion is far higher among short, common words than among long, rare ones. Mistyping *sat*, for instance, is quite likely to produce another word (*set*, *sit*, *sad* and so on), whereas mistyping *antirrhinum* is not. Taking this into account, the proportion of all typing errors that produce other words may be as high as sixteen per cent (Peterson 1986).

When spelling errors, as well as typing errors, are included, the problem becomes much more alarming. In a corpus of about four thousand errors taken from the free writing of secondary-school pupils of school-leaving age, forty per cent were real-word errors (Mitton 1987). In some cases the misspelling was based on pronunciation, and it was only by chance that it matched a dictionary word, such as *tort* for *taught*, but, more often, the misspelling was some other familiar word, as if the person writing it had two known spellings in mind and chose the wrong one. The wrong one was often a homophone of the right one, but not always. Errors of this kind were particularly likely to occur on function words (words like *of*, *and*, *be* and so on); in eighty per cent of the misspelt function words, the error consisted in writing some other function word in place of the one intended, such as '*He* name was Mrs Williams', and 'You *we* treated like babies'.

1.4 Detecting grammatical anomalies

Very few spelling checkers make any attempt to detect real-word errors, but at least three research projects have tried to tackle the problem. The first is a system called critique (previously called epistle) developed by IBM (Heidorn et al. 1982). This is a piece of software that will check the spelling, grammar and style of business correspondence. Armed with a complicated set of grammar rules for English, it attempts to parse each sentence of the text, ie, to analyse a sentence into its syntactic parts — Noun (Subject of sentence), Adjective (qualifying Subject), Main Verb, Prepositional phrase, and so on. If it fails, because the sentence is grammatically incorrect, it tries again, this time relaxing some of its grammar rules, and it carries on doing this until it achieves a successful parse. Since it knows which rule or rules it had to relax, it can work out what was grammatically wrong with the sentence. A real-word error quite often produces a grammatically incorrect sentence (such as 'I might of done'), so critique can detect such errors and can sometimes suggest a correction, since the syntactic context gives a lot of clues to what the word should have been.**[3]**

The second project also depends on syntax as a way of spotting real-word errors. It is a modification of a system, developed at Lancaster University, for tagging words in a text with their parts-of-speech (Marshall 1983, Garside et al. 1987). Given a sentence such as 'The fly bit the goat', it first consults a dictionary to find out which tags (parts-of-speech) each of the words can have; it will find that *the* is a definite article, and that *fly* (likewise *bit*) can be a noun or a verb. It also has a table, derived from a large corpus of English text, showing the probability of a given tag being followed by another in a sentence; the table will show, for example, that a definite article is very likely to be followed by a noun, but not likely to be followed by a verb. It then works out, purely on probability, that *fly bit* in this sentence is likely to be Noun-Verb, rather than Verb-Noun (or Noun-Noun or Verb-Verb).

The system can be applied to looking for real-word errors by modifying it to report when it finds a sequence of tags that is very unlikely. For example, it would query 'Please complete the *from* in capitals', since the sequence *the from in* (Definite article, Preposition, Preposition) has only a low probability (Atwell 1983, Garside et al. 1987).

Both systems have some success in spotting real-word errors, but both tend to give too many false alarms because of sentences which are grammatically unusual but not ungrammatical (Richardson 1985, Leech et al. 1986). Neither can do anything about real-word errors that are not syntactically anomalous, such as 'We had thirty *minuets* for lunch', 'We used to *pant* on Thursdays and hang up the *pitchers* on the walls', 'There was a *fate* every summer'.

1.5 Detecting unlikely word combinations

The third assault on real-word errors, again by researchers at IBM, resembles the Lancaster work somewhat in using probabilities derived from a very large corpus of text, but the probabilities are not of the co-occurrence of tags but of the co-occurrence of actual words (Mays et al. 1991). Given any two words from their 20,000-word dictionary, they can say what the probability is of any other of their dictionary words occurring next. Given, for instance, 'I think' as the first two words, they could say what the probability was of the word *that* occurring next. Or of the word *slowly* or *big* or *therefore* or *teapot*. (Presumably the probability of *that* after 'I think' is relatively high, whereas the probability of *teapot* after 'I think' must be close to zero.)

In an experiment, they took sentences containing a single real-word error, such as 'The thief licked the lock' (for *picked*). The misspellings were all of the simple typing-slip kind, ie, differing by just one mistype from the correct spelling. (I explain below what I mean by that.) When

considering a word as a possible error, the system first generated all the words that might have been changed into this word through a typing slip. For example, from *licked* it would have generated *kicked*, *ticked*, *locked*, *liked* and so on, including *picked*. For each of these alternatives, it calculated the probability of the whole sentence from its table of three-word probabilities, ie, one value for 'The thief *kicked* the lock', another for 'The thief *ticked* the lock', and so on. It also calculated the probability of the original sentence, 'The thief *licked* the lock'. If 'The thief *picked* the lock' came out as more probable than 'The thief *licked* the lock', it would conclude that *licked* was a real-word error that should be corrected to *picked*.

It could be wrong either by leaving the original error uncorrected or by preferring the wrong alternative or by 'correcting' some other word in the sentence. It had no way of knowing in advance that *licked* was the misspelling here. It would go through the same procedure with all the other words. It would generate *dock*, *rock*, *sock*, *look* and so on for *lock* and might possibly prefer 'The thief licked the *rock*'.

There was a further factor in its calculations, namely a general level of expectation of errors in the text. This was set by the experimenters at levels between 0.1 and 0.0001. Essentially, if it was told to expect a lot of errors, it tended to make a lot of corrections, ie to rate the alternatives as more probable than the original, though many of its 'corrections' were wrong. If it was told that errors were rare, it was more respectful of the original text; when it did make a correction, it was nearly always right, but it left a lot of the misspellings uncorrected.

It is not clear what application this method could have to ordinary spelling checkers in the near future because of its considerable demands on memory and computing power, but it is the only method I know of that has been capable of detecting (and correcting) syntactically acceptable real-word errors in unrestricted text.

2 Spelling correction

Many people find that a spelling checker is all they need; they know how to spell and they just want their occasional slips to be pointed out to them. People who have trouble with spelling, however, need something more. Suppose you have written *neumonia* and the checker has told you this is wrong. If you don't know how to spell *pneumonia*, you're stuck. The dictionary is no help. You want the computer to tell you the correct spelling.

To correct someone's spelling errors, you have to be able to guess what words the person meant and you have to be able to spell them correctly. People generally find the first part easy but the second part harder; most people would understand 'She was excused swimming because of her verouka', but they would not be able to correct it. For computers, it's the other way round. Producing a correct spelling is easy — they can store a complete dictionary and retrieve any word as required; the hard part is deciding which word was intended.

It is for this reason, incidentally, that one cannot say in general whether computers are better or worse than people at spelling correction. Given a minor misspelling of a long word, such as *innoculation*, a computer will detect it and correct it better than most people would, because it is easy to guess what word was intended but not easy to spell it. By contrast, with a misspelling of a common word, such as *cort* ('We got cort in the rain'), a computer might have difficulty deciding that *caught* was the word intended, whereas most people would correct it easily.

Given a dictionary of realistic size — say 30,000 to 80,000 words — it is not practical to go through the entire dictionary for each misspelling, considering every word as a possible

candidate; a corrector has to select a section of the dictionary, of some tens or hundreds of words, and search through these in the hope of finding the correct word.

2.1 Simple errors

Analyses of errors — mainly typing errors — in very large text files (Damerau 1964, Pollock & Zamora 1984) have found that the great majority of wrong spellings (eighty per cent to ninety-five per cent) differ from the correct spellings in just one of the following four ways:

one letter wrong (*peaple*)
one letter omitted (*peple*)
one letter inserted (*peopple*)
two adjacent letters transposed (*pepole*)

It has also been found (Yannakoudakis & Fawthrop 1983a) that the first letter is usually correct. Given a mistyped word, therefore, there is a good chance the correct spelling will begin with the same letter and will be either the same length or just one letter longer or shorter. If the words are held in order of first letter and length, the corrector can easily restrict its search to the appropriate section of the dictionary (Turba 1982).

Words that are misspelt, as opposed to mistyped, tend to differ from the correct spellings in more than just the simple ways listed above (Mitton 1987). For example, *disapont* — a misspelling of *disappoint* — is two letters shorter than the correct word; looking through the dictionary at words beginning with d and of seven to nine letters long would fail to find *disappoint*. You could simply increase the number of words to be considered, perhaps taking in words that are two letters longer or shorter than the misspelling, but this would increase substantially the number of words the corrector had to look at, so it would take longer to produce its correction. It would also be inefficient since a large proportion of the words it looked at would be nothing like the misspelling; for *disapont*, it would take in *donkey* and *diabolical*, which are obviously not what *disapont* was meant to be. What is needed is some way of retrieving those words that have some resemblance to the misspelling.

2.2 Soundex

This problem has been around for a long time in the context of retrieving names from a list of names. Suppose you are working at an enquiry desk of a large organization, with a terminal connecting your office to the central computer. A customer comes in with a query about her account. She says her name is *Zbygniewski*. You don't want to ask her to spell it — perhaps her English is poor and other customers are waiting. To make matters worse, the name may be misspelt in the computer file. You want to be able to key in something that sounds like what she just said and have the system find a name that resembles it.

The Soundex system was devised to help with this problem (Knuth 1973, Davidson 1962). It dates, in fact, from the days of card-indexes — the name stands for 'Indexing on sound' — but has been transferred to computer systems. A Soundex code is created for every name in the file. The idea of the code is to preserve, in a rough-and-ready way, the salient features of the pronunciation. Vowel letters are discarded and consonant letters are grouped if they are likely to be substituted for each other — an s may be written for a c, for instance, but an x for an m is unlikely.

The details are presented in Figure 1, with some examples.

1) Keep the first letter (in upper case).
2) Replace these letters with hyphens:
   A, E, I, O, U, Y, H, W.
3) Replace other letters by numbers as follows:
   B, F, P, V : 1
   C, G, J, K, Q, S, X, Z : 2
   D, T : 3
   L : 4
   M, N : 5
   R : 6
4) Delete adjacent repeats of a number.
5) Delete the hyphens.
6) Keep first three numbers or pad out with zeros.

For example:

Birkbeck	Zbygniewski	toy	car	lorry	bicycle
B-621-22	Z1-25---22-	T--	C-6	L-66-	B-2-24-
B-621-2	Z1-25---2-	T--	C-6	L-6-	B-2-24-
B621	Z125	T000	C600	L600	B224

Figure 1: The Soundex code, with some examples

So, every name in the file has one of these codes associated with it. The name *Zbygniewski* has code Z125. Let's say you key in *Zpignyefsky*. The computer works out the Soundex code for this and retrieves the account details of a customer with the same code — *Zbygniewski* — or perhaps the accounts of several customers with somewhat similar names.

It is fairly obvious how this system can be applied to spelling correction. Every word in the dictionary is given a Soundex code. A Soundex code is computed from the misspelling, and those words that have the same code are retrieved from the dictionary. Take as an example the misspelling *disapont*. A corrector would compute the code D215 from *disapont* and then retrieve all the words with code
D215: *disband*, *disbands*, *disbanded*, *disbanding*, *disbandment*, *disbandments*, *dispense*, *dispenses*, *dispensed*, *dispensing*, *dispenser*, *dispensers*, *dispensary*, *dispensaries*, *dispensable*, *dispensation*, *dispensations*, *deceiving*, *deceivingly*, *despondent*, *despondency, despondently, disobeying, disappoint, disappoints, disappointed, disappointing, disappointedly, disappointingly, disappointment, disappointments, disavowing.*

2.3 The SPEEDCOP system

The purpose of the SPEEDCOP project was to devise a way of automatically correcting spelling errors — predominantly typing errors — in a very large database of scientific abstracts (Pollock & Zamora 1984). A key was computed for each word in the dictionary. This consisted of the first letter, followed by the consonant letters of the word, in the order of their occurrence in the word, followed by the vowel letters, also in the order of their occurrence, with each letter recorded only once; for example, the word *xenon* would produce the key *XNEO* and *inoculation* would produce *INCLTOUA*. The words in the dictionary were held in key order, as illustrated by the small section shown in Figure 2.

PLTDOE	plotted	PLTNGAI	plating
PLTE	pellet	PLTNSUO	plutons
PLTEI	pelite	PLTNUO	pluton
PLTIO	pilot	PLTOU	poult

Figure 2: A section of the SPEEDCOP dictionary

When the system was given a misspelling, such as *platin*, it computed the key of the misspelling and found its place in the dictionary. In this example, the key of *platin* would be *PLTNAI*, which would come between *PLTIO* and *PLTNGAI*. Moving alternately forwards and backwards from that point, it compared the misspelling with each of the words to see if the misspelling could be a single-error variation on that word, until either it had found a possible correction or had moved more than fifty words away from its starting point. The SPEEDCOP researchers found that, if the required word was in the dictionary, it was generally within a few words of the starting point. In the example, the corrector would quickly find the word *plating* as a possible correction (*platin* being an omission-error variant of *plating*).

The Soundex code and the SPEEDCOP key are ways of reducing to a manageable size the portion of the dictionary that has to be considered. Confining the search to words of the same length (plus or minus one) restricts the search even further. The price to be paid is that, if the required word is outside the set of those considered, the corrector is not going to find it. **[4]**

2.4 String-matching

The next task facing the corrector is to make a selection from the words it looks at — a best guess, or at least a shortlist. If the corrector is intended mainly to handle typing errors, this task is not difficult. Given that the great majority of mistyped words fall into one of the four classes listed above, the corrector compares the misspelling with each candidate word from the dictionary to see if they differ in one of these four ways. If they do, then that candidate joins the shortlist. Given the misspelling *brun*, for instance, the corrector would produce the list *brunt* (omitting one letter gives *brun*), *bran* (changing one letter), *bun* (inserting one letter) and *burn* (transposing adjacent letters).

Another way of selecting candidates is to calculate, in some way, how closely each word resembles the misspelling and to shortlist those that have the best scores. This process is called 'string-matching', and there are many ways of doing it. One way is to see how many chunks of the shorter string are present in the longer string (Joseph & Wong 1979). For instance, given *medsin* and *medicine*, you could say that *medsin* has the *med* and the *in* of *medicine*, a total of five letters out of the eight in *medicine*, a score of sixty-three per cent. Another method considers the number of trigrams (three-letter sequences) that the two strings have in common (Angell et al. 1983). *Medicine* and *medsin* would be divided up as follows (the # symbol marks the beginning or end of a word):

medicine	#me med edi dic ici cin ine ne#
medsin	#me med eds dsi sin in#

The more trigrams the two have in common, the better match they are considered to be. Some methods give more weight to letters near the front; others rate letters near the end more highly than those in the middle; some rate certain letters more highly than others, such as consonants over vowels. **[5]** Some hand-held spellcheckers make use of a special-purpose chip which implements string comparisons at high speed (Yianilos 1983).

2.5 Feature vectors

A project at Bellcore is investigating the use of spelling correction in an unusual setting, namely to assist deaf or speech-impaired people to use the telephone (Kukich 1992b). Deaf people can communicate with each other over a telephone line by using a screen and keyboard. When they want to converse with a user of a voice telephone, they go via a relay centre. The voice user speaks to the relay person who types the message to the deaf person; the deaf person types back and the relay person speaks it. Bellcore would like to automate this process, and part of this involves the generation of computer speech from the keyed text. But this text typically contains typing errors which upset the speech generator, hence the need for spelling correction. The corrector is allowed to make only one correction for each misspelling, not a list of possible corrections such as a spellchecker would produce.

Experiments have found that one of the simpler methods is the most effective. A 'feature vector' of about five hundred bits (think of a line of lightbulbs again) is computed for each word in the dictionary. If the word contains an a, the first bit is set (the first lightbulb is turned on); if it contains a b, the second is set, and so on. If it contains aa, the 27th is set; if it contains ab, the 28th is set. (There is no place in the line for letter-pairs that don't occur in English, such as yy.) A corresponding feature vector is computed for the misspelling and this is compared with the vectors of the dictionary words. The word whose vector is most like the misspelling's vector (most nearly has its lightbulbs on and off in the same places) is chosen as the correction.

2.6 Error probabilities

Some methods of string-matching make use of tables showing the likelihood of this or that letter being involved in an error. One of these methods (Wagner & Fischer 1974) is incorporated in the spellchecker that I have developed, and I describe it in more detail in my book. It was developed originally for correcting the output of an optical character reader. These machines are prone to make certain errors more than others; for example, they are likely to read a lower case e as an o, but not likely to read lower case t as m. The corrector has a table showing the probability of one letter being mistaken for another, and it uses these figures in deciding what the word ought to be. Given *gom*, it would guess that the word was *gem* rather than *got*.

Probability is also the basis of an approach developed at Bell Labs for correcting typing errors (Kernighan et al. 1990, Church & Gale 1991). This system has tables of error probabilities derived from a corpus of millions of words of typewritten text. The tables give the probability of an a being substituted for a b, a p being inserted after an m, and so on. It also has an estimate of the probability of any particular word occurring in the text.

When it detects a misspelling (which it does by dictionary look-up), it first retrieves from the dictionary all the words that could have given rise to this misspelling by a single mistype. (It doesn't handle more complicated errors.) For example, from the misspelling *acress*, it retrieves *actress*, *cress*, *caress*, *access*, *across* and *acres*. Taking *actress*, it consults its table for the probability of having a t omitted after a c and combines this with the probability of meeting the word *actress*. In this way it produces a probability estimate for each of the candidates and it then puts the candidates in order of probability for presentation to the user.

The errors that poor spellers make are more complicated than those of an optical character reader or a typist, but a similar approach can still be used. One system (Yannakoudakis & Fawthrop 1983b) has a table of error patterns, derived from the analysis of a corpus of spelling errors; the table might show, for instance, that au is sometimes written as or, or ch as tch. It

compares the misspelling with each of the words in the section of the dictionary that it's looking at to see if the difference follows the patterns in its table. For example, given *lorntch*, it would find that *launch* differs from it in two of these ways. The table also contains information about the frequency with which each of these error patterns occurs, so the corrector can put the shortlisted candidates into order. When trying to correct *lorntch*, it would also find *lounge*, but it would rate this as less likely than *launch* because the table contains the information that or for ou and ge for ch are less likely than or for au and tch for ch.

2.7 Phonetic similarity

Some of the more advanced commercial correctors also retrieve candidates on a 'phonetic' basis. Their dictionaries presumably contain information about pronunciation, and the correctors use this to offer words that might sound like the misspelling, even though they don't look much like it; for *newmoanya*, for example, their list would include *pneumonia*.

Commercial companies tend not to publish details of how their spellcheckers work, but there is one pronunciation-based spellchecker described in the research literature; it was developed in the Netherlands for the correction of Dutch, though the principles would apply to English also (van Berkel & De Smedt 1988). It uses a variation on the trigram system mentioned earlier, but with pronunciations rather than spellings. Given the misspelling *indissceat*, for example, it would begin by making a guess at the pronunciation — perhaps /indiski:t/ — then break this up into 'triphones', and then compare this with the pronunciations of various words in its dictionary, also broken up into triphones. The comparison with *indiscreet* would look like this:

indissceat #in ind ndi dis isk ski: ki:t i:t#
indiscreet #in ind ndi dis isk skr kri: ri:t i:t#

The more triphones a dictionary word has in common with the misspelling, the better match it is considered to be. **[6]** Homophones, of course, match perfectly.

2.8 Ordering the list

Most correctors simply offer a small selection of possible corrections, generally about six, for the user to choose from, though some correctors offer dozens of suggestions if the user wants them. This shortlist, however, is often a curious rag-bag of words. When asked to make suggestions for *perpose*, Microsoft Word Version 6.0 produced the list (in this order) *preppies*, *propose*, *papoose*, *prepuce*, *preps* and *props*, but not *purpose*. The lists often contain obscure words with no indication of their level of obscurity; many of the offerings are wildly inappropriate for the context and perhaps not even syntactically possible. When asked for suggestions for cort in 'I've cort a cold', Wordperfect 5.1 produced — take a deep breath — *cart, cert, coat, colt, cont, coot, copt, cor, cord, core, corf, cork, corm, corn, corp, corr, cors, corti, cost, cot, court, crt, curt, carat, carate, card, cared, caret, carried, carrot, carte, cerate, cered, ceroid, chaired, charade, chard, chariot, charred, chart, cheered, cheroot, chert, chirred, chord, choreoid, chorioid, choroid, cirrate, cored, corrade, corrode, corrupt, coward, cowered, curate, curd, cured, curet, curette, curried, karate, kart, keyword, scared, scarred, scirrhoid, sciuroid, scored, scoured* and *scurried*, but not, alas, *caught* (perhaps because *caught* and *cort* are not homophones in American speech). One can't help feeling that the corrector ought to be able to do better — to restrict its list to plausible suggestions and to order them so that its best guess is generally the one required. Given 'You shud know', it ought to offer *should* ahead of *shad* and *shed*.

Word frequency can help; *shad* could be removed from the list for *should*, or at least relegated to the bottom, purely because of its rarity. But it doesn't help much; candidates in the shortlist are often of similar frequency, such as *there* and *their* for *ther*, and a rare word will occasionally be the one required.

Syntax can also help. I described earlier how some correctors do a syntactic analysis in order to spot real-word errors; they can use the same analysis to rule out some of the candidates. Quite often, as in *shad*, *shed*, *should*, there will be only one candidate left.

2.9 Using semantics

A semantic analysis is much more difficult for a computer to attempt, but it may be possible when the subject matter is restricted (Morgan 1970, Teitelman 1972). For example, a corrector that checked the commands that people typed into an electronic mail system would be able to correct *Snd* to *Send* (rather than *Sand* or *Sound* or *Sud* or *Sod* or *And* or *End*) in 'Snd message to Jim', because *Send* is one of the few words that could occur at that point in this sentence (Durham et al. 1983). Similarly, a system that handled enquiries on rail travel could use its interpretation of the meaning to correct 'Is there an erlier conexson?' (Hendrix et al. 1978) A system of this kind might be able to detect some real-word errors. A computerized tourist guide might detect that a query about gold courses was really about golf courses. More ambitiously, a system that conducted a dialogue with a user might be able to build up a representation of what the user had in mind and use this for spellchecking (Ramshaw 1994). If a user of the computerized tourist guide had asked about holidaying in the west country and then asked 'Are there trains to Swinton?' the system might guess that he meant Swindon, since Swindon is on the main line from London to the west whereas places called Swinton are all in the north. In general, however, spellcheckers that handle unrestricted text have insufficient information about the words in their dictionaries or about the topics people write about for them to make use of the semantic context.

2.10 The state of the art

At present, then, checkers and correctors play a small but useful role in helping people to remove minor errors from their written work. Some systems are just checkers — they flag errors but make no attempt to offer suggestions — and this is often all that is required; if you've typed *sciecne* for *science*, you can correct it easily. Most systems, however, do both checking and correcting, so that the word *spellchecker* usually means a piece of software that both checks the text and offers suggestions for misspelt words. A list of suggestions can occasionally be helpful, especially for people whose spelling is a little weak; not everyone would know, if a checker queried *occurence*, that it ought to be *occurrence*. But spellcheckers are still some way short of offering the help that a poor speller wants — the kind of job that a good typist would do.

They miss a fairly high proportion of errors; real-word errors form a substantial minority of spelling errors and most spellcheckers ignore them completely. Their suggestions are often irritatingly inappropriate, frequently including words that are obscure or syntactically out of place. If the misspelling differs from the correct word in certain ways, such as having a different first letter (*nowledge*, *wrankle*, *eny*), or being more than one letter longer or shorter (*probly*, *cort*, *cigrets*, *pollitishion*), or having several letters different (*payshents*, *powertree*, *highdrawlick*), the required word may not be in the list of suggestions at all.

References

Alberga, Cyril N, 'String similarity and misspellings', *Communications of the A.C.M.*, vol.10, no.5, pp302–313, May 1967.

Angell, Richard C, Freund, George E, & Willett, Peter, 'Automatic spelling correction using a trigram similarity measure', *Information Processing and Management*, vol.19, no.4, pp255–261, 1983.

Atwell, Eric, 'Constituent-likelihood grammar', ICAME News: newsletter of the International Computer Archive of Modern English, no.7, pp34–67, May 1983.

Blair, Charles R, 'A program for correcting spelling errors', *Information and Control*, vol.3, pp60–67,1960.

Church, Kenneth W & Gale, William A, 'Probability scoring for spelling correction', *Statistics and Computing*, vol.1, pp93–103, 1991.

Cornew, Ronald W, 'A statistical method of spelling correction', *Information and Control*, vol.12, pp79–93, 1968.

Damerau, Fred J & Mays, Eric, 'An examination of undetected typing errors', *Information Processing and Management*, vol.25, no.6, pp659–664, 1989.

Damerau, Fred J, 'A technique for computer detection and correction of spelling errors', *Communications of the A.C.M.*, vol.7, pp171–176, March 1964.

Davidson, Leon, 'Retrieval of misspelled names in an airlines passenger record system', *Communications of the A.C.M.*, vol.5, pp169–171, March 1962.

Durham, I, Lamb, D.A, & Saxe, J.B, 'Spelling correction in user interfaces', *Communications of the A.C.M.*, vol.26, no.10, pp764–773, October 1983.

Garside, Roger, Leech, Geoffrey, & Sampson, Geoffrey, *The Computational Analysis of English: a corpus-based approach*, Longman, 1987.

Hall, Patrick A.V. & Dowling, Geoff R, 'Approximate string matching', *Computing Surveys*, vol.12, no.4, pp381–402, December 1980.

Heidorn, G.E, Jensen, K, Miller, L.A, Byrd, R.J, & Chodorow, M.S, 'The EPISTLE text-critiquing system', *IBM Systems Journal*, vol.21, no.3, pp305–326, 1982.

Hendrix, Gary G, Sacerdoti, Earl D, Sagalowicz, Daniel, & Slocum, Jonathan, 'Developing a natural language interface to complex data', *A.C.M. Transactions on Database Systems*, vol.3, no.2, pp105–147, 6.1978.

Joseph, D.M. & Wong, Ruth L, 'Correction of misspellings and typographical errors in a free-text medical English information storage and retrieval system', *Methods of Information in Medicine*, vol.18, no.4, pp228–234, 1979.

Kempen, Gerard & Vosse, Theo, 'A language-sensitive text editor for Dutch', in *Computers and Writing*, ed. Noel Williams, pp68–77, Intellect Books, 1992.

Kernighan, Mark D, Church, Kenneth W, & Gale, William A, 'A spelling correction program based on a noisy channel model', in *COLING–90 13th International Conference on Computational Linguistics*, ed. Hans Karlgren, pp205–210, Helsinki Unive., 1990. Vol. 2

Knuth, Donald E, The Art of Computer Programming: Vol. 3 Sorting and Searching, Addison-Wesley, 1973.

Kukich, Karen, 'Techniques for automatically correcting words in text', *Computing Surveys*, vol.24, no.4, pp377–439, 1992. (1992a)

Kukich, Karen, 'Spelling correction for the Telecommunications Network for the Deaf', *Communications of the A.C.M.*, vol.35, no.5, pp80–90, May 1992. (1992b)

Leech, G.N, Garside, R.G, & Elliott, S.J, Development of a Context-sensitive Textual Error Detector and Corrector: Final project report submitted to International Computers Limited, Unit for Computer Research on the English Language, Lancaster University, November 1986.

Marshall, Ian, 'Choice of grammatical word-class without global syntactic analysis: tagging words in the LOB Corpus', *Computers and the Humanities*, vol.17, pp139–50, 1983.

Mays, Eric, Damerau, Fred J, & Mercer, Robert L, 'Context-based spelling correction', *Information Processing and Management*, vol.27, no.5, pp517–522, 1991.

McIlroy, M. Douglas, 'Development of a spelling list', *IEEE Transactions on Communications*, vol. COM–30, no.1, pp91–99, January 1982.

Mitton, Roger, 'Spelling checkers, spelling correctors and the misspellings of poor spellers', *Information Processing and Management*, vol.23, no.5, pp495–505, 1987.

Morgan, Howard L, 'Spelling correction in systems programs', *Communications of the A.C.M.*, vol.13, no.2, pp90–94, February 1970.

Morris, Robert & Cherry, Lorinda L, 'Computer detection of typographical errors', *IEEE Trans Professional Communication*, vol. PC–18, no.1, pp54–64, March 1975.

Nix, Robert, 'Experience with a space efficient way to store a dictionary', *Communications of the A.C.M.*, vol.24, no.5, pp297–298, May 1981.

Peterson, James L, 'Computer programs for detecting and correcting spelling errors', *Communications of the A.C.M.*, vol.23, no.12, pp676–687, December 1980. (1980a)

Peterson, James L, *Computer Programs for Spelling Correction: an experiment in program design*, Springer-Verlag, 1980. Lecture Notes in Computer Science: 96 (1980b)

Peterson, James L, 'A note on undetected typing errors', *Communications of the A.C.M.*, vol.29, no.7, pp633–637, July 1986.

Pollock, Joseph J. & Zamora, Antonio, 'Automatic spelling correction in scientific and scholarly text', *Communications of the A.C.M.*, vol.27, no.4, pp358–368, April 1984.

Ramshaw, Lance A, 'Correcting real-word spelling errors using a model of the problem-solving context', *Computational Intelligence*, vol.10, no.2, pp185–211, 1994.

Richardson, Stephen D, *Enhanced Text Critiquing using a Natural Language Parser*, IBM, 1985. Research Report RC 11332 (#51041)

Riseman, Edward M. & Hanson, Allen R, 'A contextual post-processing system for error correction using binary n-grams', *IEEE Trans Computers*, vol. C–23, no.5, pp480–493, May 1974.

Teitelman, Warren, 'Do What I Mean: the programmer's assistant', *Computers and Automation*, pp8–11, April 1972.

Turba, Thomas N, 'Length-segmented lists', *Communications of the A.C.M.*, vol.25, no.8, pp522–526, August 1982.

van Berkel, Brigitte & De Smedt, Koenraad, 'Triphone analysis: a combined method for the correction of orthographical and typographical errors', in *Second Conference in Applied Natural Language Processing, Austin, Texas*, pp77–83, American Association of Computational Linguistics, 1988.

Wagner, Robert A. & Fischer, Michael J, 'The string-to-string correction problem', *Journal of the A.C.M.*, vol.21, no.1, pp168–173, January 1974.

Yannakoudakis, E.J. & Fawthrop, D, 'The rules of spelling errors', *Information Processing and Management*, vol.19, no.2, pp87–99, 1983. (1983a)

Yannakoudakis, E.J. & Fawthrop, D, 'An intelligent spelling error corrector', *Information Processing and Management*, vol.19, no.2, pp101–108, 1983. (1983b)

Yianilos, Peter N, 'A dedicated comparator matches symbol strings fast and intelligently', *Electronics*, pp113–117, 1 December 1983.

Footnotes

[1] Riseman and Hanson's trigrams are actually more complicated than my short description suggests. The letters in a Riseman and Hanson trigram are not necessarily consecutive, and the position in the word is important. For example, from the word *trunk* they would store the information that all of the following trigrams were possible: *t-r-u* at positions 1-2-3, *t-r-n* at positions 1-2-4, *t-r-k* at positions 1-2-5, *t-u-n* at positions 1-3-4, *t-u-k* at positions 1-3-5, *t-n-k* at positions 1-4-5, *r-u-n* at positions 2-3-4, *r-u-k* at positions 2-3-5, *r-n-k* at positions 2-4-5, *u-n-k* at positions 3-4-5.

[2] The *brassieres* (for *brasseries*) error was in a report quoted in a short piece about spellcheckers in *The Times* of 16 February 1995.

[3] A similar system has been implemented in a language-sensitive text editor for Dutch (Kempen and Vosse 1992). It can, for example, detect the misspelling *word* in *Peter word bedankt* (English *Peter am thanked*) and correcting it to *Peter wordt bedankt* (*Peter is thanked*).

[4] The SPEEDCOP researchers found that the most frequent cause of failure with their system was the omission of consonant letters near the beginning of a word (Pollock and Zamora 1984). For example, the misspelling *pating* would produce the key ptngai, which might be some distance away from pltngai, the key of *plating*. They therefore computed a second key, called the 'omission key'. They knew from their analysis of a large corpus of spelling errors that consonant letters were omitted in the following order of increasing frequency — the letter j was omitted the least and the letter r the most: j k q x z v w y b f m g p d h c l n t s r.

The omission key consisted of the consonant letters of the word sorted in this order, followed by the vowel letters in their order of occurrence in the word. The omission key for *pating* would be gpntai, probably close to the omission key for *plating* — gplntai.

[5] Some of these variations are described in Hall and Dowling (1980), Alberga (1967), Blair (1960) and Cornew (1968).

[6] For some misspellings it is possible that more than one variant pronunciation might be generated, though many details of the pronunciation, such as stress pattern, can be ignored since this application is less demanding than, say, speech generation. The dictionary also stores the frequency with which each triphone occurs and these frequency values are taken into account; if two pronunciations share an unusual triphone in common, this will be considered more significant than if they share a run-of-the-mill one

[Journal of the Simplified Spelling Society, 20, 1996/1 pp12–13 in printed version]
[Valerie Yule: see **Bulletins**, **Anthology**, **Quarterly**, **Journals**, **Newsletters**, **Personal Views** 10 & 16, **Media**, **Books**.]

4. Spelling Needs Reserch and Reserch Needs Replication

Valerie Yule

Valerie Yule is a reserch psycologist, formerly clinicl child psycologist in hospitals and scools, lecturer, and reserchr in literacy and in the development of imagination, also author of Orthografy and Reading: Spelling and Society, an investigation of writing systems and of the feasibility of spelling chanjes in English.

Abstract
Spelling reforms need reserch into how they meet the needs and abilities of users and lernrs. Without this assurance the most theoreticly impressiv scemes may prove useless. Argument is not enuf.

Reserch must also be replicated and extended to ensure that findings apply widely, and not only to small sampls. The problems and importance of these requirements ar ilustrated by the experience of the riter's own reserch into readrs' responses to surplus-cut spelling modifications.

1 The need to repeat experiments to ensure findings ar reliabl
Statistics ar almost idolised in reserch in the behavioral siences. This is because findings from small sampls of peple ar then jeneralised to whole populations by calculating the probability that they are not simply chance results. Such inferential statistics ar cheapr than repeating experiments to make sure the findings stand up to repetition. Howevr, in human afairs the variabls ar oftn so complex and uncertain that replication is far mor convincing evidence than any singl study. This is the issue I woud like to call atention to now. I hav done what I can in arguing for reserch rather than argument in spelling reform, and in piloting experiments in reformd spellings and in testing responses to modified spellings. We now need replication and extension of existing reserch.

2 Problems of repeating experiments
The problem is not just expense. A greatr problem is that publishing new reserch that makes new findings is mor important for reserchrs, to hav mor chance of publication, and indeed, of academic employment and promotion. There is litl kudos or reward for those who test the reserch of others by replication—unless it explodes some great myth of our time, and oftn not even then. For exampl, Milgram found that his sampls of American students woud obey experimentrs' ordrs to giv electric shoks to human victims. Australian reserchrs found that Australian student subjects refused do what their experimentrs askd, if it seemd to involv hurting others. But Milgram's gloomy conclusions ar stil reportd as the truth about human nature.

3 Recomendations for reserch jurnals
I woud like to see reserch jurnals include, as regular features, notes about replications of publishd reserch — whether it has replicated wel, or not replicated at all. This woud hav four purposes:

i) to publicise the present status of erlier publishd findings
ii) to recognise and encuraj those necessary peple, replicaters—and so encuraj mor replicaters.
iii) It woud also be fruitful for theory and practice, since it coud point out the criticl conditions undr which a finding woud apply or not.
iv) It woud enable the orijinl experimentrs to say if they had been taken up rongly or misrepresentd, particulrly, as is not unknown, when those misrepresentations hav been made in the wider media and no public rebutl has been allowd.

4 Replication of reserch into surplus-cut spelling
I woud like to support this call with ilustrations from my own long-term work on readrs' responses to spelling chanjes.

As far as I kno, no one has atemptd to replicate any of my pilot reserch on readrs' response to modified spellings, altho the materials for them hav been ofrd for some years. Some of my experiments hav been of necessity no mor than explorations, so that only speculations coud be made as to their implications. Others, eg, experiments in responses to tachistoscopic presentations of singl words, hav replicated with so many difrent word lists and typs of subjects as wel as having statisticl support, that I am fairly confident of the conclusions drawn from them—that in those conditions readrs ar not significantly disadvantajd on first meeting chanjed spellings of words.

5 Adaptation to spelling chanje with practice — criticly important reserch
Howevr, the typ of reserch that it is most necessary to replicate concerns readers' adaptation to modified spellings in text with practice over time. Initial modifications in spelling reform need to be those that present readrs can acsept most esily, becaus they wil normaly be th arbitrs of chanje in a society alredy literat. John Beech (1983) made a study of how 13 subjects adaptd to reading a 'Regular Spelling' of his own design. In the 1980s a coleag and I carried out an unpublishd study of 94 subjects' adaptation to reading in surplus-cut spelling over a series of up to 20 sessions. In the event, this reserch was fataly handicapd by a combination of misfortunes that preventd its propr completion and pland follow-up. There had also been dificulties during the experiment in obtaining reliabl and completed responses from some of the 'poor readr' subjects in some of the hour-long sessions of reading practice when co-supervisors had not been availabl. Howevr, the experiment was able to sho no significant difrences between group post-test scors reading in the modified spelling and initial scors in conventionl spelling, and there was the intriguing finding that is worth pursuing, that poor readrs in the experimentl condition appeard to hav improved in reading in conventionl spelling, as wel as making subjectiv reports of greatr enjoyment of reading. Some aspects stil need further investigation.

Since then I hav carried out a further series of experiments and pilot experiments in Australia (see reports in the JSSS). As I now no longr hav access to facilities to continue these, I am returning to look at individual difrences in the data, testing ocasionl subjects individualy on some of the paper and pencil tests, and looking at raw data in the longitudinl study that I did not previusly analyse because too much was missing. In the final fully supervised sessions, for which complete data wer availabl, experimentl subjects' reading rates in 'surplus-cut' spelling wer not significantly difrent from those of control subjects and their own initial rates in conventionl spelling. However, the data for all eighteen practice sessions ar not complete. Some protocols hav missing sections, or ar scrappy because some subjects showd exessiv variability in response, while others did not fill in complete responses for every text they read. But group findings for the subjects reading in modified spelling during many of the practice sessions wer that quite oftn they read mor slowly than the control group. Was this contrast of reading rate with the post-practice test a factor of the experimentl situation? Woud the missing data hav chanjed these apparently inconsistent findings?

The few anecdotal reports availabl 18 months later indicated that some poor readrs continued their improvement in reading skil and intrest aftr their extended practice in reading in surplus-cut spelling—what of those who coud not be contacted? Mor longitudinl studies ar needed.

6 Difrent findings for difrent groups
Another study (Yule & Greentree, 1986) compared readrs' initial responses to difrent typs of modified spelling with reading in conventionl spelling. Strong confirmation of the relativ dificulty of difrent typs of spelling modification when first encountrd is provided by ocasionl tests I hav givn other subjects individualy. I expectd this since the statisticl findings wer repeatd over three

groups of subjects and with difrent texts. Howevr, in this ocasionl testing, I hav found that while a good many individuals replicate the finding of rapid adjustment to norml reading speed, most subjects pikd up mor sloly than the Scottish subjects. Why hav Australian subjects been slower in adaptation to text reading in surplus-cut spelling? Wer difrences in conditions important? Ar there difrences in the subject sampls or populations? Woud students from Scottish scools 10 years later, following the pedagojicl chanjes ther, now respond difrently?

One speculation is whether difrences in how peple wer taut to read afect their adjustment to improved fonemicity in spellings. That is, does it afect adaptation to surplus-deleted spellings if readrs wer orijinly taut—or lernt—by fonic methods, See-and-Say, or Languaj-Experience? There is evidence that for realy fluent readers, their latent fonics stratejies improve their skil, whatevr visual and orthografic stratejies they hav also developd (Ch.5, Yule 1991). Cutting surplus letrs might therfor help readrs with a fonics base among their skils mor than those whose basic stratejies wer formd by predominantly visual methods of teaching such as Look-and-Say or the Languaj Experience Aproach.

Most or all of the subjects in the Scottish experiments woud probably hav been taut by fonics or eclectic methods, as the Whole-word and Languaj-Experience tecniques had only recently begun to suplant them for the yungr jeneration then stil in primary scools; but in Australia many of the subjects almost certnly startd with look-and-say typ teaching. They woud therefor be less likely to hav an undrlying fonics stratejy that coud benefit as much and as quikly from surplus-cut spellings.

7 Typs of experiments now needed

Many experiments for testing responses to surplus-cut spellings ar simpl and cheap—they require only pencil and paper. Others require computers, which can program tachistoscopic presentations and carry out the statisticl analyses. Sofisticated experiments with tecniqes such as maskd priming ar also posibl and desirabl.

The most valuabl experiments woud use subtitles on video and television. This has many advantajes. Mass testing woud be posibl, and chanjes coud esily be made in the modified spellings presentd as the experiment progressd. Vewers' subjectiv responses on the relativ ese of chanjes and objectiv mesures of facilitated reading woud then help to identify the spelling principls and exampls that realy help lernrs and users to read texts. With continued exposure, vewrs woud becom familiar both with spelling chanjes and the idea of spelling chanjes. The informd opinions of the great English-speaking public and internationl EFL readrs coud then swing in support of spelling improvement against the elitist reargards that woud deny them mor efectiv means of languaj comunication in print and on screen.

Refrences

Beech, J. 1983. The effects of spelling change on the adult reader. **Spelling Progress Bulletin, 23 .1.**Item 5.

Upward, C., Fletcher, P., Hutchins, J. & Jolly, C. 1990. **A Handbook to Cut Spelling**. Birmingham, U.K.: Simplified Spelling Society.

Yule, V. 1986a. The design of spelling to match needs and abilities. Harvard Educational Review, 56: 278–297.Chapters 1, 10.

Yule, V. & Greentree , S. 1986b. Readers' adaptation to spelling change. Human Learning 5: 229–241. Ch.9.

Yule, V. Orthography and Reading: Spelling and Society. Doctoral thesis, Monash University, 1991.

Note: this articl is ritn in a miniml surplus-cut spelling which is based on studies of 'what th market wil bear' when new readrs encountr cut spellings for th first time. It is therefor pragmatic rather than completely consistent. As it is, even a 4.3% cut in surplus letrs greatly reduces th clutr than handicaps lernrs and confuses spelrs.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp14–18 in printed version]
[Patrick Groff: see **Bulletin**, **Journals**, **Newsletters**]

5. Recent Spelling Research and Simplified Spelling

Patrick Groff

Patrick Groff is professor of education emeritus San Diego State University, USA

Recent empirical research on traditional spelling has implications for the spelling reform movement. Most deals with how schoolchildren spell.

1 A Massive Study of Spelling Errors

One of the most massive studies of spelling errors ever is by Cramer and Cipielewski (1995). They analyzed what they decided were 55 types of spelling errors in 18,599 unedited children's compositions written on topics of the children's choice. These children were enrolled in grades 1 thru 8, in 256 classrooms in all 50 states of the USA. A total of 1,584,758 written words were examined.

These investigators contend that "the English language is not the chaotic beast of mythology it is often made out to be. On the contrary, it is systematic and reasonably predictable" in the conventional way it is spelt (p15). However, in a doubtless unintended acknowledgement of the guiding principle of simplified spelling, the authors agree that conventional "spelling knowledge has been shown to be much more than the ability to match letter to sound".

The authors present four "features" in conventional English spelling that they feel make it "reasonably predictable" (p16). These are:

1 the predictable way affixes are spelt;
2 the fact that two words related in meaning may have similar spellings altho they are pronounced differently, eg, *signal/sign*;
3 regular consonant letter-sound matches";
4 spelling patterns within words.

While not so stated in their report, feature number 4 presumably refers to the fact that there also are some "regular" vowel letter-speech sound matches in English spelling.

The simplified spelling movement has made a strong case that too many spellings of words are not controlled by these four influences. Hence its insistence of the need for a highly systematic procedure for spelling all speech sounds.

2 Boosting Case for Simplified Spelling

The Cramer and Cipielewski (C&C) study does reiterate key information on which spelling reform is based. Thus they found there were over three times as many categories of misspellings of vowel sounds as of consonant sounds. Misspellings of vowel sounds also constituted 38% of the total spelling errors in the study. For consonant sounds the figure was 17%. The 10.5% of spelling errors the study found appearing exclusively in affixes and inflections also involved vowel sounds. Therefore probably close to half the misspellings involved defective transcription of vowel sounds.

This finding supports the heavy concentration by advocates of simplified spelling on reformation of vowel spellings. With reform in this area of spelling, a large percent of present spelling errors

would decline, consistent with C&C data. Similarly, Treiman (1993) found that 22% of first-grade children's misspellings of vowel sounds in words were "legal substitution errors". That is, these spelling mistakes were not correct for the particular word, but were possible conventional spellings of the vowel sound. There are 22 different possible spellings of the vowel /i/ (Groff & Seymour, 1987). Reducing the number of legal substitutions undoubtedly would facilitate children's learning to spell /i/ and other vowel sounds.

The C&C study calculated high coefficients of correlation (r's) between words children misspell from grade to grade. (An r of +1.00 indicates a perfect positive relationship between two variables.) For example, the r found between grades 3 and 5 was .85; between grades 4 and 7 was .83; between 6 and 8 was .91; and between 5 and 8 was .83. As the study correctly noted, "the words primary grade children misspell are, in many instances, the words intermediate and junior high school children continue to misspell" (p28). The investigators then inadvertently repeat the simplified spelling solution to this problem: "Clearly if one could reduce the errors children make on a relatively small subset of troublesome words, substantial progress in spelling proficiency would be made" (p28).

Simplified spelling is the most rational way to cope with this subset of "troublesome" words, its promoters maintain. Acceptance of this relatively small gain in the direction of simplified spelling also would indicate that advocates of spelling reform are willing to heed the advice that the future of reformed spelling depends on avoiding "the radical and wide-sweeping proposals that have doomed previous simplification movements" (Venezky, 1983, p26).

Taking this limited step toward simplification of the spelling of vowels therefore might do better to overcome the natural conservative attitude of the populace toward change in any forms of the language, including its spelling. The probability of popular acceptance of regularization of vowel spellings would likely increase if the proposed changes did not eliminate the morpheme identity of words, i.e. did not obscure their shared semantic bases (eg, the spellings *signal/sign* would be retained).

Other findings of the C&C study unintentionally buttress a main argument for spelling reform. For instance, I applied a simplified number of phonics rules (common generalizations that predict how words are spelt conventionally) to the 100 most frequently misspelled by children in each of the grades 1 thru 8 in the study. The percent of predictably spelt words that are misspelled decreased consistently from grades 1 thru 8. Thus, as children progress in their spelling ability, they tend to master the spelling of predictably spelt words. By grades 7–8 in the study, 85% of the words that were misspelled frequently by students were unpredictably spelt ones. If the simplified spelling solution — reduction of the conventional ways to spell vowel sounds (the present "legal substitutions") — was applied to the spelling of words, the 85% of frequently misspelled words by children in grades 7–8 theoretically could be reduced almost to zero.

3 The Study's Implications for Teachers

Cramer and Cipielewski unfortunately offer some controversial opinions as to why children make spelling mistakes, and especially the same ones year after year. These researchers make the dubious assumption that one can look at a child's misspelling of a word and tell whether it was caused (a) "by misunderstanding how to spell words correctly", or (b) by "inattention" on the speller's part (p30). The fact that misspellings of certain homophones persist as the most common misspellings made by children, grades 1 thru 8, is presented as "proof" of this "inattention" to the spelling task.

The homophone *too* was found to be the most frequently misspelled word across eight grade levels. The fact that *their* was found to be the fifth most frequently misspelled word, *there* the sixth, and *they're* the 15th, are viewed as signs of "carelessness or indifference by children as they spell homophones." The omission of a letter in a word also is seen as a prime example of "inattention".

"Omitted letters proved to be the single greatest *cause* of spelling errors" across all grade levels, the study deduced (emphasis added) (p30). Children in the study misspelled the word *because* in 175 unique ways, most of which involved the omitting of a letter (Marine, 1995). Since such omitted letters "are due to inattention to the spelling task", the study rationalized, this psychological factor has overwhelming influence on the incidence of spelling errors.

However, the claim that inattention is a major *cause* of spelling errors is a hypothesis open to question. The authors of the study in effect admit so when they properly note that "most people value the ability to spell correctly very highly" (p36), and therefore do not take learning to spell lightly. A person's "educational qualifications and even intelligence" may be assumed from observations of his or her spelling performance, the study's investigators concede (p36). This judgment evidently acts as a stimulus for students to be attentive when spelling words. Thus only 1.5% (!) of the words handwritten by accomplished students (applicants to Cambridge University, the United Kingdom) were misspelled (Wing & Baddeley, 1980).

4 Is Inattention the Villain?

Children's thoughts are ahead of their hand and finger movements during handwritten spelling. Thus, spelling errors are produced that their writers later are able to correct, provided they were pointed out to them. However, there was no indication from the C&C study that the children were examined on their ability to subsequently correct the misspellings they made. Therefore citing children's purported apathy toward correct spelling as a principal cause of their misspellings appears much like blaming a victim for the offense committed against him or her.

This "offense", spelling reformers maintain, is the unpredictability of conventional spelling. This handicap to spelling utility cannot be remedied satisfactorily by trying to make conventional spelling tasks easier for children to master. Spelling reformers would consider, as largely a diversion from the essential issue, C&C's advice (p38) that words given to children to learn to spell be based on factors such as the frequency of their appearance in oral language and in school subjects, the frequency with which certain words are misspelled (the prime contribution of their study), the four "features" that govern conventional spelling (noted above), and information on "developmental spelling stages". Spelling reformers contend that if words were spelled predictably these considerations would become minor.

5 Developmental Spelling Stages

The "developmental spelling stages" that children are said to pass thru are of much current interest to educators. These are supposedly important for teachers to consult when deciding what words children are given to learn to spell, and how instruction for them is to be provided. When a child is encouraged to "invent" the spellings of words (instead of writing them according to direct and systematic instruction), over time this pupil will spell a word differently, depending on the particular "natural" stage of spelling development in which he or she happens to be. That a peculiar form of invented spelling is used is held to be proof that a student is at a certain one of these various stages.

This information is considered useful to teachers who stress the use of invented spellings by their pupils. It "helps those teachers make sense of misspellings", Beers (1995, p54) contends.

This teacher "is likely to feel less overwhelmed by the number of invented spellings if the misspellings can be systematically identified and organized for instruction" (p54). Treiman (1993), among others, agrees.

An immediate flaw in such advice (Groff, 1986) is that these teachers appear to be given an unmanageable task. They have the overwhelming job of (a) identifying accurately which of their students is at each of the various developmental stages, and then (b) devising uniquely different instruction for each developmental stage. The developmental spelling experts (eg,. Beers, 1995) so far have failed to provide a practical plan (the valid and reliable criteria to be fulfilled) for the successful completion of this first task.

To meet the second task, Beers (1995) simply advises dividing up the customary sequence of direct and systematic teaching of spelling skills into successive parts, and then implementing these separate parts at each of the developmental stages. As it turns out, there appears to be little essential change in content and sequence of instruction given in invented spelling/developmental spelling classrooms from that provided in classrooms that base spelling instruction purely on how predictably words are spelt.

6 The 'Whole Language' Connection

Confusing this is Cramer's (1995) mistaken assumption that there are great similarities between the 'Whole Language' (WL) approach to spelling development (which promotes invented spelling), and direct and systematic teaching of a sequence of spelling skills, carefully arranged into the order that students previously have demonstrated difficulty in learning. Cramer correctly notes that "the principles of Whole Language ... point the way to an integrated reading-writing" approach to literacy development (p78). However, the WL recommendation that reading, writing and spelling instruction be integrated, ie, be taught so that learning one reinforces learning the other, was widely advanced long before the advent of WL. This tenet of WL therefore is not a unique, guiding principle of WL.

WL advocates argue that children best learn to spell in school in the same natural, individualized, informal way they learned to speak at home, as preschoolers. A leading supporter of WL (Gentry, 1987) explains the implications for teachers of this WL presumption. To be a *bona fide* WL teacher, he relates, one must inform students that "weakness in spelling is okay" (p8). Teachers thus should expect that "not all children will learn to spell well" (p25). "Good spelling is merely a convenience", Gentry argues, not a necessity (p8). Therefore, the quality of spelling in a student's written composition very seldom should be considered when assigning a grade to it.

Including quality of spelling as a criteria of students' writing proficiency is unfair, Gentry (1987, p10) argues, since "expert spellers are born, and cannot be developed in school". Thus, he submits, there is no significant relationship between correct spelling and intelligence. That is, the "visual memory" necessary for proficient spelling is "not a skill one can consciously acquire". Neither is learning the rules for correct spelling effective as a way to become a good speller, he cautions teachers. Children learn more from "free writing", Gentry continues, than from engaging in spelling "exercises" or other formal instruction. Moreover, "doing well on spelling tests" does not mean "competency in spelling" is being developed.

The preponderance of relevant experimental evidence *contradicts* each of these admittedly novel WL notions (Groff, 1986, Petty, 1982). Spelling reformers unfamiliar with the current debate over spelling instruction need to understand furthermore that this controversy centers on which is more valid, anecdotal evidence about spelling development (which WL advocates readily supply), or on pertinent empirical data (which they summarily reject). The major question

therefore becomes: which kind of evidence about spelling development should rule when the two types come into direct conflict?

Using anecdotal findings, Gentry (1987) contends that unless children "invent" the spellings of words they will not progress satisfactorily in learning to spell. He thus urges teachers to adopt this WL practice, and to defend it when talking to parents. Honig (1996, p91) adds that invented spelling is "a powerful tool in developing skills and knowledge about reading". Thus, neither spelling nor reading skills supposedly will develop as well as possible unless children are urged to invent spellings of words.

In contrast to the developmental stages of spelling theory, the experimental evidence indicates (Groff, 1986), skills for conventional spelling are best developed by direct and systematic instruction of words carefully arranged as to how predictably they are spelt. Gentry (1987) flatly rejects such evidence. He contends that the "best strategy for a formal spelling lesson" has not been discovered. "I don't think there is one!" he exclaims (p29).

He then assures teachers there is no danger in encouraging children to invent spellings for as long as they see fit. "There is no evidence that invented spellings become habitual." The act of inventing personalized spellings of words also is sufficient to cause a child to "refine those spellings and progress developmentally toward correctness", he argues. He allows only one reason for holding children accountable for correct spelling in their writing. This would be if their manuscripts are "being readied for publication", which would only happen rarely. Again, it must be emphasized that the claims that invented spelling develops more spelling ability than otherwise has not been verified empirically.

7 The Treiman Study

From her study of the spellings of first-grade children, Treiman (1993, p291) also concludes that invented spelling by a child will help his or her teacher "make a good guess about why the child [mis]spelled the word in the way that he or she did". This guess purportedly will satisfactorily "determine what help the child needs in order to spell better". In fact, she continues, teachers who are not aware of "the logic behind children's [spelling] errors" instruct in ways that "may indeed be worse than no instruction at all" (pp 150–151).

One of the prime examples offered by Treiman of instruction based on interpretations of "the intricacies of children's spelling" (p277) is a first-grader's misspelling of *her* as *hr*. Because *er* represents the "syllabic /r/", the child has "used reasonable graphemes *Hr* to represent" /hur/, Treiman notes (p292). How would a teacher's knowledge about this help a child spell /hur/ as *her*? Or, how would a teacher, ignorant of this information, proceed differently to help a child learn to spell *her* correctly, from a teacher having this information?

Treiman criticizes what she calls the "orthographic classification" of spelling errors (p277). Here, misspelling of *her* as *hr* is classified as the omission of the letter *E*. This classification supposedly is too simplistic because it does not "determine how children derive spellings from sounds" (p278). However, Treiman then recommends formal instruction that impresses upon children that they "must memorize the *E* of *her*" (p201). Children who reverse letters in words when spelling them "especially need drill on the correct order of the letters" (p258). Teachers' knowledge about the "syllabic *R*", and why children reverse or omit letters, thus appears to be relatively unused in deciding how to instruct children to spell words conventionally.

Therefore, Treiman's (1993, p94) insistence that it is necessary to "get a good idea of what instruction the students need" appears to be wishful thinking that her intricate research findings

will have some practical usefulness. Her seemingly desperate search for utility for her findings is illustrated further by her assumption that it would "do little good" for the child who spells *neck* as *nak* to hear how *neck* sounds different from *knack*. Treiman violates here her recommendation to conduct formal instruction on children's phonological awareness. Her advice (p123), "rather the child [who misspelled *neck*] might be assured that /e/ does sound similar to /a/, but that /e/ is usually spelt with *E*" is far too abstruse instruction for first-grade children to understand. Obviously, not all that can be discovered from intricate linguistic studies is applicable when teaching children to spell.

8 Phonological Awareness and Spelling

Another area of research on children's spelling of which orthographic reformers should be aware is the extent to which students' phonological awareness affects their acquisition of spelling skills. Unfortunately, "there has been little research on the relation between [children's] phonemic [phonological] awareness and spelling" (Treiman, 1993, p32). Much more study has been conducted on the effects of children's phonological awareness on their development of reading than of spelling proficiency.

Thus "it is now well-established that there is a strong connection between children's ability to detect and manipulate the sounds making up spoken words, and their reading development" (Goswami, 1994, p.32). Likewise, "phonological sensitivity, coupled with letter knowledge, is sufficient for comprehending the alphabetic principle" (i.e., understanding that written language is a graphic representation of its oral version) (Bowey, 1995, p67).

This sensitivity to speech sounds is critical for children's learning to decode words. Thus "phonological processing skills should be considered to be important human abilities in their own right, similar to the intellectual abilities assessed on measures of general intelligence" (Torgerson, Wagner, & Rashotte, 1994, p282). It is predicted confidently that a "7-minute phonological awareness test will predict ease of initial reading acquisition [by children] better than a 2-hour intelligence test!" (Stanovich, 1994, p284). Thus phonological awareness appears to have a crucial influence on children's spelling development, more than for their reading acquisition, since to spell a word correctly the child must be more aware of its speech sounds than to read the word (Tangel & Blachman, 1955).

Phonological awareness by children refers to their ability to answer successfully questions such as these about spoken monosyllabic words: (1) Do *run* and *sun* rhyme? Say a word that rhymes with *cat*. (2) How many sounds are there in *at*? In *cat*? (3) Do *run* and *sun* begin the same? (4) Does *run* begin with an /f/? (5) Does *sun* end with an *N*? (6) What is the first sound in *big*? The last sound? The second sound? (7) What word does /r/-/a/-/n/ say? (8) Do *sit* and *meat* have the same middle sound? (9) Say *meat* without the /m/ sound. Say *meat* without the /t/ sound. (10) Say *os* with the first sound last. Say *os* with the last sound first. (11) Change the middle sound of *beat* to /a/.

This sequence of phonological tests is said to represent the approximate order of difficulty of the items for young children. But there is only limited evidence as to the precise degree that improving students' phonological awareness affects their ability to spell conventionally. This data will likely be forthcoming, due to the current high interest among educators in the subject.

9 Summary

This discussion indicates that recent research on children's spelling is conducted under the assumption that conventional spelling is 'maligned' by those who claim that it "is complex, illogical, and irregular" and therefore difficult for children to learn (Treiman, 1993, p21). Spelling reformers who promote these supposed harmful mistruths about conventional spelling (e.g. Rondthaler & Lias, 1986) are viewed as narrow minded and misinformed about how words should be spelt.

According to Treiman (1993, p21) spelling reformers are convinced that there is "one reason and one reason alone" (unpredictable spelling) why children have difficulty in learning to spell. The movement to simplify spelling thus is unaware of or rejects the influence of ineffective teaching, and of learning disabilities of students. The movement also "unreasonably" discredits the evidence that conventional spelling contains letter-speech sound correspondence rules that have only a few exceptions. These views of spelling reformers explain why they believe "children's only hope of success [in spelling] is to memorize the spelling of each word" while exercising "little intelligence or thought" (p21). Spelling reform movement members would rush to exclaim that these are inaccurate and misleading characterizations of the movement.

A major point that spelling reformers fortunately will concede, Treiman (1993) correctly goes on, is that young children may not share their contention that the appearance of *T* in *tape, sty*, and *city*, for example, represents regular or predictable spelling. First-grade children "may consider *sty* to contain /d/ rather than /t/" and accordingly spell it with *D* (p286). Thus "making the spelling system more regular for adults would not make it more regular for children" (p200).

Spelling reformers doubtless are bemused, however, at the serious suggestion that our system of spelling should be contingent on the uninstructed, transitory, and natural (if not instinctual) impressions that first-grade children have about the ways the phonemes should be spelt. This elevates the quality of child reasoning to an undeserved high. It illustrates the romantic child-centeredness, child empowerment fad that presently envelops public education. We do not establish systems of metaphysics, epistemology, logic, ethics, or aesthetics on the inexperienced and mentally immature perceptions of children. There is substantial empirical evidence that children's reasonings are remarkably inadequate in these fields. There is thus little basis to hope that young children's notions about spelling are any more reliable or valid.

Interpretations of recent research on children's spelling also would put heavy new burdens on teachers. These interpretations indicate that teachers not only should be expected to notice carefully whether children misspell words. Teachers must also master, and recall when needed, all of the intricate and expansive interpretations from research as to why young children misspell words as they do. Thus, no longer may a *bona fide* teacher simply instruct children to substitute, omit, add, or rearrange letters in words they misspell. Teachers now also must be prepared to reveal to individual young children the unconscious thoughts they exercised when they misspelled a word. It is highly doubtful, however, whether these all-encompassing new demands on already over-burdened teachers are either practical or expedient.

In sum, it will not be surprising to find that spelling reformers find most of the new evidence on children's spelling (excepting that on phonological awareness) to be frivolous, irrelevant, and/or useless. The reaction from the simplified spelling movement to this research may be much like that given by the inventors of incandescent lamps to investigators of the efficiency of candles.

References

Beers, J. (1995). Spelling development: stages and strategies. In: Scott Foresman Publisher (Ed.), *Spelling research and information: An overview of research and practices* (pp53–66). Glenview IL: Scott Foresman.

Bowey, J. A. (1995). On the contribution of phonological sensitivity to phonological recoding. *Issues in Education*, 1, 65–69.

Cramer, R. L. (1995). Making better spellers: Integrating spelling, reading, and writing. In: Scott Foresman Publisher, op. cit. (pp77–92).

Cramer, R. L. & Cipielewski, J. F. (1995). Research in action: A study of spelling errors in 18,599 written compositions of children in grades 1–8. In: Scott Foresman Publishers, op. cit. (pp11–40).

Gentry, J. R. (1987). *Spel ... is a four-letter word.* Portsmouth NH: Heinemann.

Goswami, U. (1994). Phonological skills, analogies, and reading development. *Reading,* 28 (2), 32–37.

Groff, P. (1986). The implications of developmental spelling research: A dissenting view. *Elementary School Journal*, 86, 317–323.

Groff, P. & Seymour, D. Z. (1987). *Word recognition: The why and the how.* Springfield IL: Charles C. Thomas.

Honig, B. (1996). *Teaching our children to read.* Thousand Oaks CA: Corwin.

Marine, K. (1995). Using the research: Developing a spelling curriculum. In: Scott Foresman, op. cit., pp41–52.

Petty, W. T. (1982). Handwriting and spelling: Their current status in the language arts curriculum. In: W. B.

Barbe, A. S. Francis, and L. A. Braun (Eds.), *Spelling: Basic skills for effective communication.* (pp53–60). Columbus OH: Zaner-Bloser.

Rondthaler, E. & Lias, E. J. (1986). *A dictionary of fonetic spelling.* New York NY: American Literacy Council.

Stanovich, K. E. (1994). Romance and reality. *Reading Teacher,* 47, 280–291.

Tangel, D. M. & Blachman, B. A. (1995). Effect of phoneme awareness instruction on the invented spelling of first-grade children: A one-year follow up. *Journal of Reading Behavior,* 27, 153–185.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1994). Longitudinal studies of phonological processes and reading. *Journal of Learning Disabilities,* 27, 276–286.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children.* New York NY: Oxford University.

Venezky, R. (1980). From Webster to Rice to Roosevelt: The formative years for spelling instruction and spelling reform in the USA. In: U. Frith (Ed.), *Cognitive processes in spelling* (pp9–30). New York NY: Academic.

Wing, A. M. & Baddeley, A. D. (1980). Spelling errors and handwriting: A corpus and a distributional analysis. In: U. Firth (Ed.), Ibid, (pp251–285).

[Journal of the Simplified Spelling Society 20 1996/1 pp19–24 in printed version]

6. Italian spelling, and how it treats English loanwords

Christopher Upward and Virginia Pulcini

Virginia Pulcini researches and teaches English at the University of Turin, Italy, and is author of 'The English language in Italy' (*English Today* 40, October 1994, pp49–52). Chris Upward is Editor-in-Chief of the Simplified Spelling Society.

Abstract.
This paper first outlines the spelling principles of Italian, which is known for the regularity of its sound-symbol correspondences. It then describes how Italian has borrowed many words from English, especially in the second half of the twentieth century, and in the process of integrating and assimilating such loanwords it has sometimes modified their spelling according to its own orthographic rules. Finally, this process of modification is examined as a form of regularization of the incongruities of English spelling.

1. Introduction.

The phonetic (or better, phonemic) basis of Italian spelling, as opposed to the strong etymological element in the French or English orthographies, has been held up as an example of coherence and simplicity. The Italian-American linguist Mario Pei commented on its effect for the native-speaking learner as follows (Pei 1968):

You are taught the alphabet, then you are given sequences of spoken and written syllables.... There are a few confusing moments when you are taught to insert an H after C, G, SC, and an I after the same consonants, to show certain sounds before front or back vowels. Beyond that, your ear is a guide to your spelling if you speak standard Italian... The word *spell* does not exist in the Italian vocabulary, which is a clue to the entire situation.

Pei then recounted his dismay on meeting written English, which he described as "one of the world's most awesome messes".

There is in fact some pardonable exaggeration here. Italian does have the verb *compitare*, to 'spell out', but it is rarely used, since educated Italians would feel embarrassed to ask for the spelling of a word, as that would be perceived as a sign of ignorance. In extreme cases, eg surnames, Italians might ask for clarification of certain details, for example "Do you write *Cerutti* with one T or two?"

The contrast with English was starkly demonstrated by Gwenllian Thorstad's recent comparative study of literacy acquisition in the two languages (Thorstad 1991). This showed English children making over four times as many misreadings and nearly eight times as many misspellings in their mother tongue as equivalent Italian children in theirs.

2. Evolution of Italian orthography.

To assist readers' understanding of Italian spelling, we will briefly outline its basic rules. In fact, although Italian is certainly far simpler than English from an orthographic point of view, even this 'highly regular' language has some inconsistencies.

The codification of Italian orthography goes back to the 16th century (the *Cinquecento*). It was achieved through the joint work of grammarians and printers who produced a set of stable rules, subsequently recorded in the *Vocabolario degli Accademici della Crusca* (1612) and its following editions. The reference model was the educated Florentine dialect, which emerged as a standard in the 'Trecento' (14th century) because it was used by the outstanding writers of that period

(Dante, Petrarch [anglicized from Italian *Petrarca*], Boccaccio) and because of the political and economic prestige of Florence. This codification marked a historical break with Latin and established the precedence of the 'phonetic' principle over the 'etymological' one.

Since the 'Cinquecento' orthographic reform, the spelling system of Italian has undergone some further adjustments, but its basic phonetic principle has been maintained, with minor divergences between graphemes and phonemes. An important innovation which became established only from the end of the 17th century was the differentiation between the vowel U and the consonant V. Some Latin digraphs were adapted (MN>NN, CS>SS, CT>TT, PT>TT), as for instance with Latin *columna*, pronounced /koˈlonna/ in Italian and therefore written *colonna* (=English *column*), and similarly Latin *dīxĭt*, now pronounced /disse/ and written *disse* (=*said*); Latin *ŏttō*, now pronounced /otto/ and written *ɔtto* (=*eight*); and Latin *scrīptŭm*, now pronounced /skritto/ and written *scritto* (=*written*). The letters K, X, Y, which existed in Latin though rarely used, were gradually abandoned, as were the digraphs CH, PH, TH, which Latin had used to transliterate the Greek letters *chi, phi, theta* (χ, φ, θ). Since Italian pronounced them no differently from C, F, T, the latter became the standard modern spellings, as in *caos* (formerly *chaos*)*, filosofia* (formerly *philosophia*) and *teatro* (formerly *theatro*). Likewise with J, used in Latin merely as a variant (allograph) on I: though maintained throughout the 17th and 18th centuries for the semivowel value /j/ (=English Y), it was then gradually abandoned. Today J is still used in some words such as *Jugoslavia, Jole* (alternatively *Iugoslavia, Iole*) and in foreign words. **[1]**

3. Regularities and irregularities in Italian spelling.
The relationship between the alphabet and the phonemic system of contemporary Italian is shown in the following table (from Maraschio, 1993):

Letter	Phoneme	Letter	Phoneme
A	/a/	N	/n/
B	/b/	O	/o, ɔ/
C	/k, ʧ/	P	/p/
D	/d/	Q	/k/
E	/e, ɛ/	R	/r/
F	/f/	S	/s, z/
G	/g, ʤ/	T	/t/
H	see below	U	/u, w/
I	/ɪ, j/	V	/v/
L	/l/	z	/ts, dz/
M	/m/		

As shown above, certain pairs of phonemes (/e, ə/, /o, ɔ/, /ɪ, j/, /u, w/, /s, z/, /ts, dz/) are homographic, that is, they are identically spelt. Conversely, certain other phonemes (/k, g, ʧ, ʤ, ʃ, ɲ, ʎ/) are heterographic, that is, they may be spelt in more than one way. For instance, /k/ is spelt C in *casa* /ˈkaza/ (=*house*), but Q in *qui* /kwi/ (=*here*); /g/ is spelt G in *gatto* /gatto/ (=*cat*), but GH in *ghiro* /ˈgiro/(=*dormouse*); /ʧ/ is spelt C in *cera* /ˈʧera/ (=*wax*), but CI in *ciao* /ˈʧao/ (=*hello, goodbye*); /ʤ/ is spelt G in *gelo* /ʤəlo/ (=*cold* [n.]), but GI in *giusto* /ʤusto/ (=*right* [adj.]); /ʃ/ is spelt SC in *scemo* /ˈʃemo/ (=*stupid*), but SCI in *sciarpa* /ʃarpa/ (=*scarf*); /ɲ/ is always spelt GN as in *gnomo* /ˈɲɔmo/ (=*gnome*) but the digraph GN may exceptionally be pronounced as separate sounds as in the German loanword *gneiss* /ˈgnɛɪs/; and /ʎ/ is spelt GL in *gli* (=*the* [m.pl.]), but GLI in *glielo* (=[eg, *give*] *it to him/her*), though in a loanword like *anglicano* /angliˈkano/ the two letters may again have separate values.

These incongruities of Italian spelling are a cause of some difficulty for Italians, and errors are typical of low levels of education. The main spelling problems encountered by Italians are the following:

- the digraphs or trigraphs SC for /ʃ/, CH+I for /k/, GL for /ʎ/, and GN for /ɲ/. These are perceived as violations of the phonetic principle of 'one symbol = one sound' in Italian.
- the grapheme H, always silent, which functions as a diacritic in indicating the value /k/ for the digraphs CH, GH, but has an etymological origin in distinguishing the homophones *anno* (=*year*) and *hanno* (=*they have*), these being derived from Latin *annus, habent* respectively. However, since Latin H has generally been dropped (Italian *abitabile, eroismo, ippopotamo, onesto, umanità* parallel English words beginning with h), one may say that its retention in *hanno* also has a psychosemantic function, enabling the reader immediately to distinguish two common homophones.
- the presence of I in certain plurals (-CIE /tʃɛ/, -GIE /ʤɛ/ as in *camicie* (=*shirts*), *ciliegie* (=*cherries*), compared with its loss from singular *faccia* (=*face*), *pioggia* (=*rain*) in their identically pronounced plural endings of *facce, piogge*.
- double (geminated) consonants, which cause problems especially in the early stages of learning and for speakers of regional varieties of Italian.
- the choice of the nasal consonants M or N, as determined by assimilation to a following homorganic stop; thus M must precede B or P as in *bambino* (=*child*), *campana* (=*bell*), and N must precede D or T as in *andare* (=*to go*), *vento* (=*wind*). Typical errors are then **banbino, *canpana*.

These difficulties are of some significance, when one remembers that a census in 1981 found that 62% of Italians were 'semi-educated' (Maraschio, 1993: p142).

4. Dialect and standardization.
Finally, there is a further problem which undermines sound-symbol correspondences in Italian. The relative stability and uniformity of the written norm is not matched by a homogeneous spoken norm, which is in fact marked by strong regional variation, even among educated speakers. As the Italian linguist Canepari (1983) pointed out, the Florentine model on which Standard Italian was based failed to extend to the rest of the peninsula because of the strong presence of competing dialects. At present there is no standard pronunciation of Italian equivalent to RP in Standard British English, but instead many regional standards whose pronunciation is strongly influenced by dialects. The strongest unifying force, beside the mass media, is the homogeneous orthography, and, as linguists have repeatedly pointed out, the recommended pronunciation for foreign learners is the one based on the spelling. The allophonic contrasts between {e, ɛ}, {o, ɔ}, and {s, z} are etymologically motivated (eg, *pesca* /ˈpeska/ [=*peach*] from Latin *pĕrsicum*, contrasting with *pesca* /ˈpɛska/ [=*fishing*] from Latin *piscarī*), being used in Tuscany (the Florentine region), but nowhere else in Italy.

The tendency to unify regional pronunciations on the basis of a regular spelling system in Italian may be seen as the exact opposite of the historical development of English, where the unified pronunciation model of RP was established against a historical, non-phonetic orthography, which does not provide a pronunciation model encouraging convergence of regional accents.

At present Italian is undergoing a process of rapid transformation and innovation. Dialects are gradually losing ground to Standard Italian, owing to increasing literacy and mass communication. A strong influence is exerted by English, especially through lexical borrowing in many areas of international contact: in the language of newspapers, in advertising and in the microlanguages of science and technology. Consequently, graphemes and letter-strings have been (re-)introduced which were previously hardly used in Italian. The letters J, K, W, X, Y, which had been gradually lost over the centuries, are now used in words of foreign derivation. For

instance, the letter J has acquired the English sound-value /ʤ/ even in Latin words such as *junior* (/ˈʤu:niə / instead of /ˈjunjor/). The number of words beginning with H has increased, as well as typically English consonant clusters such as TH (*thriller*), SH (*shampoo*), RTN (*partner*), NGST (*gangster*). The grapheme Y has become fashionable in Christian names such as *Tony* and *Mery* (spelt with e [e] according to its pronunciation). The grapheme K is often used in advertising because its unfamiliarity attracts attention, and often in politics (eg, *okkupazione studentesca* [=*student occupation*], *Amerikano*) to produce an alien effect.

5. Two-way influences.
English and Italian have long exercised strong influences on each other. An early wave of influence by Italian on English can be traced back to the Renaissance (especially from the mid-15th century), when the artistic golden age in Italy (in music, poetry, the visual arts) had a huge impact on styles, techniques and fashions throughout western Europe. The accompanying vocabulary has in many cases gained a permanent place in the English language, an excellent account being given in the *Oxford Companion to the English Language* (*OCELang*).

Apart from Italian words which have reached English via French and which therefore tend to be spelt as in French (eg, *caprice*, ultimately from Italian *capriccio*), most Italian loans in English have retained their original spelling. Although Italian spelling is highly regular in its own terms, some of its characteristic sound-symbol correspondences are not otherwise native to English, and have therefore added a further layer to the irregularities of English spelling. Characteristically Italian are the values of C in *cello* /ʧ/, CH in *Chianti* /k/, SC in *crescendo* /ʃ/, SCH in *scherzo* /sk/, Z in *mezzo-soprano* /tz/(in Italian actually /ddz/), GH in *spaghetti* /g/, GI in *Giotto* /ʤ/, GL in *serraglio* /ʎ/, GN in *lasagne* /ɲ/, and final E in *minestrone* /e/ (the latter not rhyming in Italian with final I in *macaroni*). Only occasionally has English distorted the Italian spelling, as in *macaroni* from earlier Italian *maccaroni* (modern *maccheroni*) and *seraglio* from Italian *serraglio*.

6. Assimilation of English loanwords.
The influx of English words into Italian began in the 18th century, usually through the mediation of French, but only in the second half of the 20th century has this phenomenon had a massive impact on the Italian language and Italian culture (Pulcini, 1994). The attitude of Italian linguists to foreign words has been generally 'tolerant'. **[2]** In this respect, Italian has been judged a 'democratic language', open to neological borrowing from other languages, as opposed to 'introvert languages' like German, French and Spanish, which are more inclined to try and resist.

As regards the spelling of English loans, a number of patterns are evident.

6.1. Word unchanged.
Whereas in the past English loanwords tended to be adapted in pronunciation and form according to the rules of Italian (eg, *cartone animato* from *animated cartoon*), today their original form tends to be retained. Sometimes their spelling is relatively unproblematic (eg *stop, trend*), but sometimes they incorporate sound-spelling correspondences that do not conform to Italian patterns (eg *budget, deadline, show*). As mentioned above, graphemes that were not part of the Italian alphabet are coming back into use, and complex consonant clusters and syllables ending in a consonant are new to the phonotactic patterning of Italian.

6.2. Graphic assimilation.
Graphic assimilation, ie, respelling according to Italian rules, of unadapted loanwords is not particularly noticeable. It is generally motivated by the pronunciation of graphemes in a particular position in the word, or by hypercorrection: Y assumes the [i] or [ai] sound value (eg,*bike* is respelt *byke* in Italian and *nylon* is sometimes respelt *nailon*); K changes to C or CH depending on the following vowel (*go-kart* becomes *go-cart*); CK and CH are confused, both having the velar value /k/ in Italian (eg, *back* may be italianized as *bach*). J and Y are confused

(*jersey>yersey*; *New York>New Jork*); W is replaced by V (*Walter>Valter*) or, by hypercorrection, vice versa (*voodoo>woodoo*; *volt>wolt* — although the word derives from the Italian name Alessandro Volta). However, apart from isolated examples (*goal>gol*, *roastbeef>rosbif/rosbiffe* — but also *roastbeef*), graphic assimilation is rare and typical of uneducated spellings.

6.3. Morphological assimilation.
In many other instances there is morphological assimilation, with an Italian suffix added to the English form to accord with normal Italian word-structures. This process is facilitated by the equivalence of Italian/English suffixes such as -ATION/-AZIONE (eg, *standardization>standardizza-zione*), -ism/-ismo (eg, *tourism>turismo*). Otherwise, a noun may acquire an O (eg, *dollaro*), or an E (eg *alligatore*), and similarly with adjectives, as with *shakespeariano* (also Italianized as *scespiriano*), *manageriale*. Verbs on the other hand typically add -ARE, so giving *flirtare, standardizzare*, etc. Further examples listed by Klajn are: *acro* (<*acre*), *ancestrale, atollo, behaviorismo, bluffare, boxare, crossare, darwinismo* (also *darvinismo*), *dragare, filmare, flanella, gallone, ione* (<*ion*), *malto* (<*malt*), *mocassino, mormone, quizzare, romantico, scalpare, standardizzare, stressare, truismo, vaselina*. The form *toboga* is unusual, in that the Italianate ending is achieved not by adding a suffix, but by removing the final n of *toboggan*.

6.4. Transmission through French.
Another category, typically of older loans, reached Italian through French, sometimes in the Middle Ages, but continuing into the 20th century. In these cases the spelling may reflect the French rather than the modern English form. Among the oldest are *battello* <OldF *batel* <OldE *bat* (=boat), and *est, ovest, nord, sud* (=east, west, north, south). Some place names also reflect transmission through French: *Galles* (also French, =*Wales*), *Irlanda* (French *Irlande* =*Ireland*), *Londra* (French *Londres*, =*London*), *Tamigi* (French *Tamise*, =*Thames*). Others antedating the 20th century include: *frac, redingote* (old anglicisms in French [<*frock-coat, riding-coat*] which were later passed on to Italian as *frac, redingote/redingotto*), *lingotto* (French *lingot*, =*ingot*), *frammassone* (French *franc-maçon*, =*freemason*), *milordo*, also *milord* (French *milord*, =*my lord*), *nababbo* (French *nabab*, =*nabob*), *pinguino* (French *pingouin*, =*penguin*), *vagone* (French *wagon* =*wag[g]on*), *sportivo* (French *sportif* =*sporting*), *deragliare* (French *dérailler*, =*derail*), *tatuare* (French *tatouer*, =*tattoo*). In many cases it is hard to tell the provenance of words because of the historical links between these languages (eg, *photography/photographie/ fotografia*). *Cinema, hotel, premier, routine* derive from French, although they are fully naturalized in English. Traces of French mediation are evident in the spelling of some modern Italian anglicisms: *boxe, stripteaseuse*. Some loans occur in Italian with both French and English spelling: *shock/choc* (alongside the verb which may be fully Italianized as *scioccare* or semi-Italianized as *shoccare*), *comfort/confort, cashmere/cachemire, pony/poney, rally/rallye, rum/rhum*.

6.5. Italianizing English consonants.
Among consonants, adaptation is particularly seen with velars and palatals. For instance, a complication arises with the letter G. Before the front vowels E and I it has a soft, palatized value in Italian, /ʤ/, as often in English too (thus in both Italian *generale* and English *general*). If the hard velar value of G needs to be indicated before front vowels, English sometimes writes GU, which in Italian is pronounced /gw/. Italian by contrast writes GH to show the hard value before front vowels (as in *spaghetti*), and has accordingly converted the old English currency unit *guinea* to *ghinea*; on the other hand the geographical term is spelt *Guinea* in both languages, but is pronounced /gwɪˈnea/ in Italian. Similarly, since the fronted A of English *gang* is represented by Italian E, the same GH is used in *ghenga*.

The fact that H is always silent in Italian has led to uncertainty over *(h)andicappare*. English *pariah* loses final Hin Italian. The letter J only appears in modern loans in Italian (eg *jazz, jeep, jeans* from English, where it is pronounced as in English, and *jodel, Jugoslavia* where it is

pronounced, as in German, like English Y); older English J loans are by contrast Italianized to GI, as in *giungla<jungle*,*giuria<jury*, *pigiama<pyjamas/ pajamas*, or to I in the case of *iuta<jute*. The letter K is not natively used in Italian and is respelt as C in *bistecca <beefsteak, folclore < folklore, quacchero < Quaker, risciò < rickshaw* (also *ricsiò, ricsò*), *scioccare/shoccare < shock* (though this may equally be derived from French *choquer*).

Although the sounds of English CH (/ʧ/), SH (/ʃ/) are normal Italian phonemes, they are spelt (as explained in §2 above) with just C or SC before the front vowels E, I, while before back vowels A, O, U an I is inserted to give CI, SCI. English loans are seen thus respelt in *cip<chip* (poker), *linciare<lynch, ponce<punch* (alcoholic),*scellino<shilling, sceriffo<sheriff, scialle<shawl, scioccare<shock*.

Consonant doubling before the verbal suffix -ARE is seen in *stoppare*, but there is uncertainty as to whether to double B in *drib(b)lare<dribble* (soccer). English W is Italianized as V in *tranvia<tram(way)*. Greco-Latin PH is regularly rendered as F in Italian (eg *filosofia*), and the pseudo-Greek English literary term *euphuism* therefore becomes Italian *eufuismo*.

6.6. Italianizing English vowels.
The Italian vowel system is far simpler than the English, there being essentially just five values for the five letters A, E, I, O, U, though (as explained in §4 above) some varieties of Italian distinguish two values of E, pronounced as [e, ɛ], and two values of o, pronounced as [o, ɔ]; these distinctions are however not functional (cf the possible distinction between two values of A in English *lass/grass*). English vowel spellings using a single vowel letter can be absorbed fairly unproblematically into Italian, if with pronunciation sometimes adapted, as when the long /ai/ value for I in English *ion* is reduced to a mere I-glide in Italian *ione*.

On the other hand, the English vowel digraphs, which have been described as representing the heart of English spelling irregularity, are more often found unacceptable, and may be Italianized as in the following words (some of the examples are archaic or rare, but are given here to illustrate the general respelling procedures): the long AI digraph of *drain* is phoneticized as E in the verb *drenare*; the long EA, EE digraphs of *leader, meeting, beefsteak/roastbeef, speech* have been phoneticized to I in *lider, bistecca/rosbif, mitingo* (though modern Italian normally has *leader, meeting*); the long vowel of *nylon* may be phoneticized as AI in *nailon*, though *nylon* is also used unchanged; the long oa of *ferryboat, goal* is phoneticized as O in Italian *ferribot, gol*; and the long OE/OO/OU of *taboo, tattoo, zoom, brougham, tourism* are phoneticized as U in *tabù, tatuare, zumare, brum, turismo*. The letter Y is not native to Italian (though it is seen in foreign loans such as *yacht, yogurt*), and its value is rendered by I as a semi-vowel in *iarda<yard*, as a full vowel in *linciare<lynch*, and as a part of a diphthong in *boicottare*.

7. Loanwords as a general problem.
The spelling of loanwords is a perennial problem for all languages. One difficulty is that, insofar as the phonology of different languages differs, the borrowing language may have no obvious way to spell alien sounds, hence for example the varied attempts at spelling the Russian consonant o at the end of English *borsch, borshch, borsht, borstch, bortch*. Another difficulty is that, even when different languages do share roughly the same sounds, they may use different spelling conventions to represent them. The borrowing language may decide to keep the foreign spellings, and let people pronounce the words as best they can; or it may decide to change the foreign spellings to accord with native spelling conventions. Speakers may then have a better chance of achieving something like the foreign pronunciation, but the visible, internationally compatible forms of the words may be lost.

Languages like Swedish, Turkish and Welsh tend to adopt the latter procedure, systematically adapting foreign spellings. Modern English and French tend to leave foreign spellings

unchanged, while German sometimes adapts and sometimes does not. It is notable that in medieval times English was much more likely to adapt foreign spellings to represent English pronunciation, but, with the influx of Greek and Latin vocabulary from the 15th century onward, respect for the classical, foreign spellings has usually been paramount in more recent times. We perhaps see a similar shift of procedures in Italian, inasmuch as many of the adapted English forms (eg *banconota<banknote, contraddanza<country dance*) clearly belong to an earlier age, while contemporary (especially American) loans seem to undergo fewer changes.

How should we judge the relative wisdom of the two approaches? On the one hand, adapting foreign spellings, so preserving a coherent set of sound-symbol correspondences for domestic use, makes literacy acquisition easier for native-speaking learners. But on the other hand, international communication and learning by foreign students are made harder when spellings vary apparently arbitrarily from one language to the next. This dilemma is further aggravated when a major source of foreign loans in many languages is English, whose spelling is notorious for its unpredictable sound-symbol correspondences. Not merely are they unpredictable within English, but they are often also unpredictable vis-à-vis other languages. For example, the present variation in consonant doubling between English and French (eg English *abbreviation, apartment*, French *abréviation, appartement*) presents additional traps for learners of both languages. The consonant gemination (ie doubling consonant letters to reflect lengthened pronunciation) so characteristic of Italian and the often rather arbitrary consonant doubling of English produce a number of Anglo-Italian anomalies, as seen in *accomodare/accommodate, appartamento/apartment, comodità/ commodity, comunicazione/communication, milione/million, repubblica/republic*, but we are here typically looking at separate developments from Latin, not Anglo-Italian loanwords. We noted in §5 above how English simplifies Italian CC, RR in *macaroni, seraglio*.

8. Italian versus English.
The two sides of this dilemma are epitomized by Italian in its relations with English. It seems particularly unfortunate that a language like Italian, whose writing system is known for its general coherence, should allow its qualities to be undermined by the growing import of unadapted English spellings. At the same time, by not adapting English loanwords, Italian speakers benefit from improved access to the world language, English. Conversely, in its own terms Italian enjoys a highly predictable writing system, but in an international context the distinctive features (listed in §2 above) which make some Italian loanwords anomalous in English are idiosyncratic internationally too; indeed they create problems for the spelling of loanwards in Italian whatever language they derive from, and for Italian loanwords in all other languages.

A key example is the spelling of the phoneme conventionally rendered as SH in English. Latin had no such sound, and the Roman alphabet therefore had no letter designed unambiguously to represent it. Since they lacked any procedure for co-ordinating the evolving sound-symbol correspondences of their writing systems, the different languages that adopted the Roman alphabet developed different spelling devices to represent the sound if it occurred in their phonologies. Thus Old English, like Italian, first tried SC. In Italian SC first appeared around the 6th-8th centuries, while in English SC was first used around that time to represent /sk/. However, in English this sound subsequently developed into /ʃ/ with the spelling SC then coming to represent that value. Middle English found SC ambiguous (it could suggest either /sk/ or /ʃ/) and added H for the latter value to give SCH (so creating the SCH trigraph now used for /ʃ/ by German), until modern English finally dropped the C and standardized on its internationally unique digraph SH. Meanwhile, the sound rendered by CH in modern English and Spanish and in Old French, was deaffricated in French, so that modern French now uses the digraph CH for the sound today spelt SH in English. In one respect, today's Italian SC digraph is in practice the most complex of all, because it requires a following I if the next vowel is A, O or U — at least English SH, German SCH and French CH are immutable, regardless of the following vowel. But

in another respect English may be regarded as having the most complex sound-symbol correspondences, in that they vary according to the derivation of the words concerned, as between *shell, chef, schist, sugar, crescendo* (not to mention *ration, passion, fashion*, etc, etc).

9. Ideas for a long-term solution.

What solutions to this dilemma of orthographic cross-infection between languages can we envisage? Clearly, a most helpful single step would be the regularization of English spelling, so that its words could be borrowed by other languages with less damage to their own orthographies. Most proposed spelling reforms for English have not considered such a criterion in their design, but we may note that, although the Cut Spelling (CS, Upward, 1996) proposal was not initially designed for compatibility with other languages, it has subsequently been found to display such compatibility to a surprising degree **[3]**. Thus the CS simplification of doubled consonants generally aligns with Spanish and Portuguese, and its replacement of 'Greek' PH by F and removal of redundant H, as from *chaos, honest*, accords with Italian *caos, onesto* (and often with Spanish and the languages of Scandinavia).

Footnotes.

[1] A short but clear introduction to the historical development and structures of the Italian language is given by A L Lepsky & G Lepsky (1977).

[2] Klajn (1977) found 2150 anglicisms in various sources (dictionaries, old and new written and spoken texts), whereas the more recent study by De Mauro (1993) on word frequency in contemporary spoken Italian gives 1049 English words. The actual proportion of anglicisms in Italian is between 0.2% and 1.4% of the global lexicon, figures which are not especially meaningful. A much higher proportion of English words has been registered in the lexis of special fields (eg, 30% in the terminology of the shoe trade).

[3] as described in detail in: Upward, C (*JSSS* 1998/2) 'Overcoming Orthographic Frontiers'.

References.

Berruto, G (1989) *Sociolinguistica dell'Italia contemporanea*, Rome: NIS.

Canepari (1983) *Italiano standard e pronunce regionali*, Padova: CLEUP.

Colluccia, R (1992) 'Grafematica' in eds. A Mioni & M A Cortellazzo *La linguistica italiana degli anni 1976–1986*, Rome: Bulzoni.

Klajn, I (1972) *Influssi inglesi nella lingua italiana*, Firenze: Olschki

Lepschy, G C (1978) 'La pronuncia dell'italiano' in *Saggi di linguistica italiana*, Bologna: il Mulino, pp95–100.

Lepsky, A L & Lepsky G (1977) *The Italian Language Today*, London: Hutchinson & Co. Ltd.

Maraschio, N (1993) 'Grafia e ortografia' in L Serianni & P Trifone *Storia della lingua italiano*, Vol.1 I luoghi della codificazione, Torino: Einaudi, pp.139–227.

De Mauro, T, et al. (1993) *Lessico di frequenza dell'italiano parlato*, Milano: Etaslibri.

ed. McArthur, T (1992) *The Oxford Companion to the English Language*, Oxford University Press, p536.

Migliorini, B (1990) *La lingua italiana nel Novecento*, ed. M L Fanfani, Florence: Le Lettere.

Pei, M. (1968) Preface to Abraham Tauber 'Better English thru Simplified Spelling — a History of Spelling Reform', reprinted in ed. Newell W Tune ***Spelling Reform, A Comprehensive Survey of the Advantages, Educational Benefits, and Obstacles to Adoption***, North Hollywood, California: Spelling Progress Bulletin, July 1982 (second edition), Item 10.

Pulcini, V (1994) 'The English language in Italy' in *English Today* 40, 10, October 1994, pp49–52.

Thorstad, G (1991) 'The effect of orthography on the acquisition of literacy skills' in *British Journal of Psychology*, 82: 527–37. [Also in TESS **Books and Papers**]

Upward, C (2nd ed. 1996) ***Cut Spelling: a handbook*** *to the simplification of written English by omission of redundant letters*, Birmingham: Simplified Spelling Society.

Zingarelli, N *Lo Zingarelli 1995. Vocabolario della lingua italiana*, Bologna: Zanichelli, 1994 (XII edizione).

[Journal of the Simplified Spelling Society, 20, 1996/1 pp24–25 in printed version]

7. The Galician Spelling Problem

Susana Doval

Susana Doval is a doctoral student in the Faculty of Philology at the University of Santiago de Compostela, Galicia, Spain, where she is researching into English spelling reform.

Galicia is an autonomous region in the NW of Spain, where, alongside Castilian (=standard Spanish), the Galego (=Galician) language is spoken. This Romance language formed a linguistic unity with Portuguese ('Galego-Portugués') in the Middle Ages, when a rich literature in that language flourished, especially in the reign of Alfonso X 'El Sabio' ('the wise' 1252–84). After that period (the so-called 'Séculos Escuros'), Castilian colonization relegated Galician to a purely colloquial status, the main posts in church and government being occupied by Castilians. During this period, Galician and Portuguese moved apart, although they were still clearly connected languages. In the 19th century, Galician began to be written again and a rich literature re-emerged.

Today, after the parenthesis of Franco's dictatorship, Galician and Castilian are co-official in this part of Spain. At present, as usually happens with long neglected languages, there is a great deal of controversy concerning the status and corpus planning of Galician, especially regarding orthography. When 19th century writers decided to start writing in Galician again, they found that they lacked a written standard, and the spelling of the resulting texts is somewhat inconsistent. Today there is controversy because some people believe (against the official position) that it might be useful for Galician to move back closer to Portuguese, and write LH instead of LL (eg, *lhama* instead of *llama*), NH instead of Ñ (eg, *canhon* instead of *cañon*), and G, J instead of X (eg, *geología* instead of *xeoloxîa*, and *hoje* instead of *hoxe*); that is, they replace Castilian with Portuguese graphemes in order to 'purify' the Galician language of Castilian influence.

The problem is that, on the one hand, this proposal also substitutes a rather etymological orthography for the more or less phonemic one represented by the official position, which uses Castilian graphemes to represent Galician phonemes. In fact, moving back closer to Portuguese implies using forms which had been long ago abandoned by Galician. In the 20th century Portuguese and Galician differ substantially in their phonological systems. On the other hand, the official orthography fails to represent the difference between open and close vowels, which exists in Galician but not in Castilian, while this difference is correctly represented by the Portuguese spelling system.

From a purely linguistic point of view, the solution seems to be half way between the Castilian and the Portuguese spelling systems. From a sociolinguistic point of view, planners have to cope with the problem of acceptance: due to geographic and cultural rivalry, Galician users are not generally in favour of adopting a system with 'Portuguese connotations', as this is felt as a threat to their national identity.

As can be seen, the question of phonemic spelling gives rise to an interesting debate in this small part of the world. Many years may still have to pass before an appropriate solution for the Galician problem can be found.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp25–27 in printed version]

8. Selling Spelling: a marketing approach to orthographic change

Matthew Thommen

The author is a subeditor with the *Saudi Gazette*, Jeddah. He was educated in South India, taking an MA in English from Kerala University in 1982. He worked as an advertising copywriter before turning to journalism. From 1984 to 1991 he was on the staff of the *Indian Express* newspaper.

1 Spelling Reform: a selling problem

The basic reason why spelling reform has not worked is this: the people who must buy it are not the ones who will immediately benefit from it. This is also what makes it one of the biggest marketing challenges of today.

The main problem with English spelling is that it is — apparently unnecessarily — difficult to teach. And the people who will benefit the most from the introduction of simpler spelling are the illiterate and the learners, not those who are already literate.

But the decision makers are invariably literate. If they are not fully literate, they probably have a secretary who is, or at least a word processor with spell check. Not only do most of them fail to see any benefit in spelling reform, they have good reason to consider it a nuisance. Having made their painful way up the literacy ladder they can hardly be expected to take kindly to the idea that it was the wrong ladder after all. They have had no motive to allow what they have learnt to be replaced with something which will require a further learning effort from them.

What the reform movement has been trying to do all these years, not surprisingly with little success, is to sell something which even if no buyers need it, is guaranteed to give them trouble!

However, this does not necessarily mean that from a marketing point of view spelling reform is impossible, or even that the traditional approach to it is doomed to failure.

2 The traditional approach

Since the days when the energies of the reform movement were mostly directed at getting one proposal — New Spelling — accepted, there have been two significant developments. First, it was realised that reform will not work as long as it looks threatening to those who must accept it, and second, the emphasis shifted from campaigning to research.

New Spelling lost its ‘official’ position largely because it was felt that other proposals looked less frighteningly different. These now include Cut Spelling, which does not aim at total consistency but primarily at streamlining the current orthography, and the step-by-step schemes, which probably gave rise to the concept of spelling management, rather than reform.

But the movement has to do more than eliminate the threatening nature of reform; it has to identify or generate a positive need for change. And this the second development can help to do.

Despite the weight of scholarship behind New Spelling and the work of many early reformers, we now have much more evidence — evidence which cannot easily be ignored — of the need for reform.

Research data, some of it published by the Simplified Spelling Society, has shown that inconsistent orthography puts learners of English at a distinct disadvantage. And difficult spelling has clearly been linked to illiteracy and the poor quality of literacy in English-speaking countries.

This kind of data can be used more aggressively. The beneficiaries of reform are often people personally important to the decision makers, their own children for instance. Strong feelings can therefore be generated by conclusively establishing that lack of reform is a handicap for English-speaking students.

Also, by linking orthography with literacy and literacy with economic progress, it is possible to play on people's fears of poverty and unemployment.

However, these links are far from being established in the public mind, and it doesn't look as if a demand for spelling reform is about to emerge. Perhaps the time is ripe for an alternative approach to be tried.

3 An alternative

Though the traditional or mainstream approach has evolved, there is something in it that has remained quite constant: the role it envisages for officialdom.

Many influential reformers have felt the government or some kind of official authority should approve of and be involved in the implementation of their proposals. Pitman and Follick, for instance, took the issue to the British parliament. This is not really surprising, because being pro-reform was often seen as slightly subversive. (The verbal class distinction was by no means antique and was probably considered vital to the foundations of social order.)

The alternative approach to be described here is based on the premise that if reform is to work, it must bypass official authority. The threatening element does not lie in the look of this or that proposal, or even in the amount of change that is planned, but in the possibility of official involvement, and what has to be eliminated is precisely that.

Popular governments cannot accept reform as long as people don't want it, and people cannot easily be made to want it as long as they think the government has anything to do with it.

A scheme that proposes the slightest of changes with official implementation is likely to upset them much more than a fully revised alphabet sold by a private company. This is because the latter can create the impression that it can more easily be ignored. Once people are reassured that they need not have anything to do with it at all, they become softer targets for a campaign promoting reform.

Before thinking of what such a campaign can be like, this question has to be considered: What kind of reform will be best for it?

4 The parallel orthography

What is sold will have to be a Parallel Orthography (PO), coexisting with the Traditional Orthography for an indefinite period of time. It may eventually grow and replace the other, or fail and wither away, but to begin with, it must minimise confusion. Ideally, no word in the PO should look exactly as a different word does in TO. The more such words there are, the more the chance for confusion.

If an ideal and entirely new alphabet is used, no word in the PO will look like any word in TO. There wouldn't be any confusion then, but there wouldn't be any sales either, because the learning effort required will be much more than the literate buyers can possibly be induced to come up with.

This does not mean, however, that the PO should have no new letters at all. Spelling systems using only current letters and symbols need not be the easiest to learn, a rather extreme example being Starfon created by Philip Starmer, whose 1989 letter to *The Economist* was reprinted on the back page of the SSS Newsletter of January 1993.

The saleable PO may have to be immediately readable, or at least immediately learnable. Schemes that propose learning in stages are by their very nature unsuitable for this approach. They do not try to get more and more of the market for their product, they must first secure the entire market before introducing more and more of the product. Only a power that controls the market can do this. Where there is freedom of choice, it is quite impossible.

The marketing approach is not incompatible with the establishment of a mechanism to manage spelling, though with a consistent phonemic system it may not really be necessary. The problem was not entirely caused by an erosion of the alphabetic principle over centuries, the erosion was largely caused by the lack of a consistent system to begin with. Spoken language may continue to change, but a new phoneme rarely emerges, and with a consistent phonemic alphabet, the written language should naturally follow the spoken one, or at least greatly reduce the divergence.

Fully consistent spelling systems, it has often been argued, look too drastically different from TO. But that need not make them more difficult to learn. To look like TO and to be easy to learn are two different requirements altogether. The first is not a prerequisite for the second and efforts at maintaining visual conformity, far from making reform proposals simpler, usually make them more complex.

Partial reform can of course do both. It can keep changes simple and retain a visual similarity to TO. But it would not achieve much. Just cutting double consonants or replacing *PH* with *F* is hardly tackling the problem. As partial reform goes further, compromises become increasingly unavoidable and it has to either introduce more rules and exceptions, or give up the comfort of visual similarity, or both. On the other hand, total reform using an augmented alphabet can mean easier learning, despite the initial visual shock.

The mainstream approach has tried to minimise visual disruption because of fears of upsetting the literate decision-makers. It has often sought to lull them into agreeing to a new orthography which looks as much like TO as possible, and therefore presumably won't be too difficult just to read, while a coming generation is to be taught to write it as well.

But this has clearly not worked, and the marketing approach envisages no such planned switchover. It sees the buyers as users. They will have to understand and learn the whole thing.

They will have to be attracted, not just reassured. And once they have decided to buy, it assumes that they are not going to ask, ‘Will it still look like English?’ What they will then want is something that is easy to learn and easy to use. They will hardly be willing to put in any extra effort just so the new orthography can look like the old.

As much as possible of what the TO user already knows will no doubt have to be retained, but it does not help to introduce a lot more that has to be learnt, just to save a little of what is already known; nor does it help not to introduce elements which look unfamiliar, but make the learning easier.

This is exactly what happens in the case of many schemes which start out with the premise that accents or new letters have no place in English. For instance text looks more like TO in New Spelling than in any augmented/diacritical system. But the pamphlet *New Spelling 90* takes three pages to explain the indefinite vowel. A more marketable scheme would certainly be one which introduces a new letter instead, because it can cut this explaining or teaching from three pages to a single line.

A traditional objection to reform using an augmented alphabet, that it will make a lot of hardware obsolete, is no longer relevant. With computers, a new alphabet can technically be the same as a new font, and installing it need not involve any change in the hardware.

If the aim is to minimise the learning effort, nothing can be easier than the pure simplicity of one letter for one sound and one sound for one letter. Proposals that stop short of such consistency usually do so only to facilitate acceptance. But it is doubtful whether telling people ‘See how much like TO it looks’ will make them accept any proposal. That may require a whole new strategy and a different kind of campaigner.

5 A Marketing Strategy

The first requirement is not to find buyers but to find a seller. Unlike the traditional approach, which focuses on the yet-to-be-literate beneficiary and the field of education, the marketing approach should concentrate on the literate buyer and the field of publishing. The spelling reformer must first sell his product to a publisher, who can then sell it to the people.

Publishers can benefit in two ways. By using a more efficient orthography they can reduce the number of characters typed, saving time, energy and materials. For mass-circulation periodicals this can mean significant gains.

They can also aim at increasing their circulation by eventually convincing less-than-literate people that learning to use their PO is far easier than learning to use TO, or that by reading it they can learn to speak better English.

Two types of publications suggest themselves as being preferable vehicles for reform — a popular, perhaps left-wing, tabloid in a country like Britain, and a glossy magazine which many middle-class readers may not buy in a country like India unless it is offered at discounted rates. (To be avoided: all publications for children.)

The following is a possible marketing scenario. The British tabloid launches a parallel edition in the PO, priced substantially lower, and with a fabulous scheme of prizes to go with it. The advertising should highlight the prize scheme and the environmental benefits of reduced consumption of energy and raw materials. The prizes could range from a standard dictionary (with pronunciation indicated in the PO rather than in, say, the IPA alphabet) for anyone who

buys the new edition daily for a week, to holidays abroad for those who write back in the PO. Those who send in classified advertisements in the PO should perhaps get them published free.

If an advertisement is carried in English newspapers in India offering one PO issue of a magazine like *National Geographic* free to anyone who writes in, thousands will. If they are then offered a cut-price subscription for a year, many of them will take it. There could also be prizes — starting from the dictionary for all subscribers. The advertising should stress that the PO will help readers to speak better English.

Before this campaign runs out of steam, the next line of attack must be launched, on business people. This will involve selling computer software in the PO. The advertising should try to make the PO more respectable, even chic. The economic and technical benefits should also be stressed. With more and more multi-media applications, there can be more and more of these benefits. There can even be programs which convert text from one to the other, so that if documents have to be in TO, or both, those familiar with the PO keyboard can type them faster. This familiarity can soon be an additional skill, and may eventually be taken for granted.

At this stage many people should be able to read TO and PO, or even a careless mixture of both. And then the coup de grace: books.

All the reading now available in any Revised Orthography seems to be Shaw's *Androcles and the Lion* (in an alphabet which one newspaper reporter described as his idea of ancient Etruscan), some Wells in New Spelling and an O. Henry short story in Fonetic American. Hardly the kind of stuff readers will be queuing up to buy.

Imagine what it would be like if a much-awaited and widely advertised book of Gabriel Garcia Marquez were to be translated only in the PO. Or if some sensational new book were to appear in it. If the PO (even one with a few new letters) can quickly be made out, few readers will deny themselves books which they really want to read.

Unlike the earlier steps though, the last one by itself does not make business sense. There isn't likely to be any profit in book publishing alone switching to the PO, and it will have to be seen as part of a wider marketing effort. Anyone venturing into the business of selling a PO will have to think of this as part of the promotional campaign.

Any such campaign will no doubt require capital to begin with. That will be available if someone sees a business opportunity in the idea, which is how all businesses start. A businessman will have to recognise that selling spelling may have as much potential as, say, the idea of selling desktop computers to the publishing industry did.

However, no publisher will take the risks involved until there are figures to show exactly what those risks are. A lot of research needs to be done. A scientific system of evaluating reform proposals must be developed. Then the costs have to be estimated and compared with the projected benefits. And this question must be answered: Is there money in spelling reform?

The scholars and thinkers have done their bit, had their say. Now it may be time for the business people to move in. For a whole new breed of reformers to come, the rallying cry could be: There's money in this. Let's make it!

[Journal of the Simplified Spelling Society, 20, 1996/1, pp28,29 in printed version]
[Jean Hutchins: see **Journals**, **Newsletters**]

Jean Hutchins revews two booklets from the United Kingdom Reading Association
These revews ar publishd in Cut Spelng.

9. Robin Campbell Miscue Analysis in the Classroom

UKRA Minibook 3, 31pp, July 1993, ISBN 1 897638 02 7, £4.50.

As a principl, it is good to look at typs of oral readng err (= miscues) — but only if it is an expedient way of helpng lernrs to improve ther readng skils (unless of corse th analysis is intendd to provide data on spelngs that trip readrs up, for th purpos of spelng reform desyn). It takes a long time to anlyz errs; it is a dificlt task; ther ar varying lists of criteria; ther is no standrdized comparisn with othr children of th same aje or intelijnce; and th authr of this booklet aknolejs that a child may sho difrnt results on difrnt texts ("Asking a child to read a different story can change the nature of the miscues they produce", p12).

Ther is no emfasis in th booklet on being aware of and prasing wat is corect. It is mor positiv to notice th 90 out of 100 words red corectly than to concentrate on th 10 errs. One can tel a gret deal about wat a child can read from th corectly red words!

This booklet dos not mention standrdized texts. It is a good idea to combine miscu analysis with obtainng a standrdized scor for acuracy, comprehension and speed of readng. Tests such as th Neale Analysis of Reading Ability and th MacMillan Reading Analysis ofr al these facilitis and include six criteria for miscu analysis as wel. We can compare th scors with th cronlojicl aje of th child, but mor importntly, with th verbl intelijnce levl or with receptiv or expressiv vocablry scors wich giv an indication of th litracy achevemnt that one can reasnbly expect. In both th Neale and MacMillan tests (wich each hav alternativ, paralel sets of texts), ther ar six storis in asendng ordr of dificlty. It is very helpful to se how children takl esy and hardr texts. With repeatd experience of th same texts, th anlyst gets to no th typicl errs that many children make. Th comprehension scor can only relate to th storis th child can read, but givs a very valubl perspectiv — is it much hyr than th acuracy scor (a situation wich usuly confirms abov avraj intelijnce, and vice versa)? dos th child use th words of th text or parafrase? can th child deduce and infer as wel as ansr 'strait' questions? Th speed of readng is importnt — dos it take this child longr to read than his peers? Myt acuracy improve if th child red mor sloly? Is it frustrating for th child to hav to work out or self-corect a gret many words? Persistntly slo readng may qualify a pupil for extra time in examnations. Standrdized tests of singl words may tel us mor about word atak skils, as th child may make very good use of context cues and read words in a story, but be unable to read words in isolation.

Th chaptr on 'Planning Literacy Programmes' is disapointng. Apart from jenrl sujestions for one particulr child, wich evry child shud surely receve anyway in its Reception year, th only specific recmendation is "greater emphasis on the use of nursery rhymes … to encourage phonemic awareness." We ar not told how nursry ryms shud be used for this purpos.

In sevrl places, th authr sujests that 'meanngful miscues', ie, rong words that make sense, shud not always be corectd. I think it is very danjerus to alow inacuracis. How wil readrs no wen they hav got th jist corectly? One letr can make a gret difrnce to a word (eg, infection/injection) and therby to th sense. One mistake may lead to anothr — to make sense of th first mistake. We do want undrstandng and readng for meanng, but not at th expense of acuracy.

Th authr dos not explain how a teachr, with maybe 35 children in th class, can find time to repeatdly record and anlyz readng miscues (as wel as riting miscues, and scor story retelng as wel).

It is useful for teachrs to no about miscu analysis, to be aware of th difrnt typs of errs, and to keep them in mind to be cald upon wen necesry, rathr than making miscu analysis a way of life, as th authr seems to advocate.

10. Brigid Smith Teaching Spelling

UKRA Minibook 5, 32pp, July 1994, ISBN: 1 897638 05 1, £4.50.

No dout th editr of this series asumed that, as Brigid Smith "has nevr been far from th clasroom", ther wud be no need to proof-read for gramr, punctuation and spelng. Dr Smith clearly folod her own advice of seprating th authrship from th secretryship, and did not complete th latr task in producing this booklet.

It wud hav been betr to hav indicated in th title that th subject was "Teachng Spelng to pupils with Special Educationl Needs". A coleag was very surprised at th examnation stratejis sujestd — rite dificlt words in pencil and chek them later with somone or in a dictionry. It is not made clear that only a smal minority of pupils need such a stratejy and that they cud not use it in examnations.

Th first part of th booklet deals mainly with suport activitis rathr than "how to lern spelngs". Dr Smith then describes sevrl visul methods of taklng unown spelngs, but dismisses th multisensry aspect of sayng th letr-names and th word. This is a pity, as ther ar many lernrs ho remembr mor esily wat they hav herd than wat they hav seen. Using visul and auditry senses togethr is even betr. Fortunatly th kinesthetic mode using tracing and riting is included. Charles Cripps, an advocat of th 'Look, Covr, Rite and Chek' rutine, aknolejd that dyslexic pupils, with poor visul memris, may need multisensry input.

Stranjely, in vew of much reserch evidnce about th valu of fonemic awareness in litracy aquisition, ther is only a brief mention of th aplication of sound-symbl relationships (it ocurs in th 'Scool Policis for Spelng and Handriting' section). Dr Smith refers to reserchrs in conection with nolej of rymng patrns, but says that it is a 'visul analojy'. Rymng is a lisnng skil, not a visul skil, particulrly in words like stone, thrown.

Perhaps Dr Smith has a fotografic memry for words and dos not undrstand th stratejis needd by lesr mortls. Wen we want to spel a dificlt word, or a non-word, we may rite it down to se if it looks ryt. We reherse it to ourselvs silently or aloud, perhaps workng thru th sylabls. We perceve th sounds in ordr and think how to convey those sounds in letrs. Thus, if we perceve a /f/ sound, we hav to think wethr to spel it with f, ff, ph or ph. If we perceve /er/ we can try er, ir, ur, ear, (w)or, re, our, etc. We do this without realizing it. Many pupils with litracy dificltis, ho do not hav good visul memris, hav to lern these spelng choices and go thru them delibratly.

Dr Smith has no sujestions for a teachng structur for spelngs, in spite of th many sceme books on th market*. A child with lernng dificltis canot just lern evry word they get rong, or evry word they myt want to use. At th very least it wud hav been helpful to mention th Murray McNally 100 kewords. As these constitute 50% of al readng, they ar a priority for lernng, a few at a time, reinforced in many difrnt ways. It wud be even betr if they wer sortd into groups of simlr words, eg, putng could, would togethr insted of leving them in an alfabeticl list.

One wud hav expectd th Scool Policy section to include a recmendation for a clear sylabus of spelng work for each year, so that al teachrs (and parents) no wat has been taut previusly, wat wil be taut during th curent year, and wat is to com later. Othrwise, som importnt words wil be misd, and othrs may be taut sevrl times. Ther can be difrentiated spelng groups within a class, each group moving forwrd at its own suitabl pace.

I feel that inexperienced teachrs wil be very mistaken if they think that th content of this booklet is a balanced overvew of teachng spelng.

Titles on litracy teachng methodolojy wich I particulrly recmend ar:

Eds, Augur J & Briggs S (secnd edition 1991) The Hickey Multisensory Language Course, Londn: Whurr

Brand V (1984) Spelling Made Easy. Multisensory structured spelling, Baldock: Egon (Introductory Level, and Levels 1, 2, 3), with related computer software: Spelling Made Easy, Egon.

Hornsby B & Shear F (4th edition 1993) Alpha to Omega. The A–Z of Teaching Reading, Writing and Spelling, Oxford: Heinemann.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp29,30 in printed version]
[Kenneth Ives: see **Bulletins**, **Anthology**, **Journals**, **Newsletters**, **Book**]

10. Spelling Research & Information: an overview of current research and practices eds. Scott Foresman Co., Glenview,IL, 128pp, 1995, ISBN 0-673-28840-4. Kenneth Ives reviews

The research aspect of this book has many good points: the study of student spelling errors by Ronald L Cramer and James F Cipielewski is excellent, with lists of most commonly misspelt words in each grade from 1 thru 8 (pp31, 95–102); the top five types of errors in elementary/intermediate/upper grades are listed on p44; developmental stages in spelling are described by James Beers on pp55–66; spellings of related words are indicated in lists on pp72, 73, 75; 5) the material is well organized; attention is given to the problems of schwa; there is a good bibliography of 97 references.

Perhaps reflecting current views and practices, there are several assumptions and biases which make many of the conclusions open to serious question. There is much overt bias against spelling reform. This is most glaring on p67, where the 1965 (*sic*) study by Hanna, Hanna, Hodges and Rudolph which "found surprising regularity in spelling patterns" is cited. No figures are given. However, that study included a report on a computer effort at spelling from sounds, using 203 spelling rules. It was able to spell correctly only 49% of 17,000 words. This does not fit any definition I know of for "surprising regularity"!

Regularizing spellings would improve the basis for at least 10 of the 55 "error categories for grades 1 thru 8": homophones (No.2), consonant doubling (Nos. 14, 23, 29, 41), silent e (Nos. 22, 24, 28), and silent consonants (No.37). Regularizing irregular words (No.48) would no longer be an error.

On p16 an author refers to "the myth of an irrational English spelling system". That author then lists four features which "make English spelling reasonably predictable":
1) word structure and proximity principles;
2) derivational principles associated with meaning;
3) spelling patterns within words;
4) regular consonant letter-sound matches.
Again, no figures are given on how predictable this makes English words, nor on how one determines which feature to use when, nor what to do when different features point to differing spellings. These four rules must involve many more rules, and many exceptions. The authors provide no estimate of teaching time required for these features, nor a comparison with teaching times using regular or reformed spellings. The experiments with reformed spellings from the 1850s and from the Initial Teaching Alphabet (i.t.a.) in the 1960s and 1970s are not referred to, nor their effects on student performance.

There seems to be a bias towards remediation, rather than clear early teaching as prevention. Remediation can be very expensive and very complex. This in turn likely rests on the present emphasis on 'invented spelling' and the claim that each student should 'discover' phonic regularities — a case of reinventing the wheel. It encourages teachers to abdicate the teaching role when introducing reading and writing. This can be deduced from the substantial number of regularly spelt words on the lists of common misspellings — 33 in grade 1, 19 in grade 8, and from 1/3 to 1/5 of the lists of 100. If adequate teaching of the phonic regularities began early in first grade, these regularly spelt words might well become less than 10% in the 100 most commonly misspelt words.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp30–33 in printed version]

11. Wat's th Problm — Spelrs or Spelngs?

Christopher Upward revews

Writing Skills: a survey of how well people can spell and punctuate

London: The Basic Skills Agency (formrly Adult Literacy and Basic Skills Unit [ALBSU]), Decembr 1995, 17pp (unumbrd exept in Table of Contents), ISBN 1 85990 044 5.
Availbl fre from th Basic Skills Agency, London

This revew folos on from revews of erlir ALBSU reports (JSSS, 1988/2, p32; 1995/1, p37), and is ritn in Cut Spelng. Thanks ar du to Leslie Morphy of th Basic Skills Agency for comments on a draft of th revew.

1 Context of th survey

In its 95/1 issu, *JSSS* revewd th presnt surveys precursr (Carol Elkinsmyth/John Bynner *The Basic Skills of Young Adults* London: Adult Literacy and Basic Skills Unit, 1994) and noted (p38, §3) that wile th c.22-year-old subjects wer asesd for ther readng and numeracy skils, ther riting skils wer not evaluated. Th presnt 1995 survey gos som way towards filng that gap, altho th 980 subjects ar difrnt peple, constituting a representativ sampl of th population of England and Wales covrng ajes 16–60. Th riting skils examnd ar spelng (therby extendng a smal 1992 survey reportd on in ***JSSS* 95/1**, Item 11, §7), punctuation, and to a limitd extent th ability to compose gramaticl sentnces. Th National Foundation for Educational Research helpd desyn th tests, wich wer implmntd by Opinion Research Business.

JSSS has alredy discusd th importnce, aims and tecniqes of spelng-err analysis at som length in 1994 (Part I, 94/1, pp29–33; Part II, 94/2, pp21–24). That discussion was based on th asumtion that th err-anlyst has access to and can study th actul errs made. Th presnt survey of riting skils dos not provide us with such data, but merely with statistics for th percentaj of riters ho mispelt a limitd numbr of comn words. We canot therfor identify precisely wat caused th errs, tho we can sujest som probbl causes and draw som conclusions about th significnce of th survey and its relevnce to english spelng reform. One factr, as always, is th predictbility of th spelngs concernd, but, thanks to th smal numbr of test words used, we can also ask how far th frequency of each word in jenrl usaj, ie, th probbl familiarity of its orthografic form to th riters, may also hav afectd th outcom. Frequencis wer noted in terms of thousnds (designated 'k' belo) of ocurences in th COBUILD corpus of 17 milion words.

2 Dificlty machd against frequency

Th subjects of th survey wer testd on th spelng of 20 words wich, altho in no way specialized, wud be likely to ocur in an employmnt environmnt. Each word was presentd in th context of a sentnce such as *I am sorry to have missed you*, or *I am looking for accommodation*, and they wer furthr categrized in four groups of five, graded by dificlty.

Group 1: sorry, come, please, have, my
This first group consistd of very comn, mostly monosylabic words, but in no case was th spelng entirely straitforwrd (for instnce, thre of them endd in silent E), and al wer mispelt by at least one persn. Th esiest wer found to be *my* (tho confusebl with th patrns of *mine, lie*) and *have*, with its aberant E misleadngly sujestng a rym with *shave*. Trickir was *come*, with two aberant letrs (contrast *home, rum*), and about 10 subjects mispelt it. *Sorry* with th uncertnty of final Y (contrast I in *horrid*, IE in *worried*) and dubld R (contrast singl R in *story, very, bury*) trapd twice as many. *Please* with its arbitry EA digraf, its S for /z/, and its final silent E (contrast th numerus othr posbl spelngs rymng with it in *these, trapeze, cheese, sneeze, frieze, seize, pleas, bees*) tripd up thre times as many. Wile we may considr that th dificltis containd in *please* ar much gretr than in *have, my*, we may also note that COBUILD givs *have* (including *haven't*) a frequency of 86k and *my* of 46k, wile *please* (lematized to include inflectd forms and derivativs insofar as these displayd virtuly th same letr string) showd just over 3k and *sorry* only 2k; *come* (with derivativs) rated 25k, and th total frequency for this group of words amountd to c.160k. Th relativ familiarity of *have, my, come* may therfor hav reduced ther proneness to mispelng at least as much as th relativ un-predictbility of ther letrs.

Group 2: *apply, complain, would, writing, because*
Th next group containd one very comn but tricky monosylabl (*would*, 52k) and four disylabic words, of wich only *because* (23k) was particulrly comn. *Writing* rates only 2k by itself, but if lematized its frequency rises to 8k. *Apply, complain* ar both only 1k, and th total for th hole Group is c.85k. *Would* is dificlt because of its near-uniqe (ie, shared only with *could, should*) vowl spelng combined with silent L. *Because* (if pronounced to rym with *was*) sufrs from th unusul valu of AU, th /z/ valu of S, and th silent final E, and is furthrmor uniqe in its structur (it is etmlojicly od too, th jermanic prefix *be-* being atachd to th french-derived base word *cause*). *Writing* sufrs from silent initial W and th loss of final E from *write* (it dos not rym with th paralel endng of *benefiting*). *Apply* has dubld P, unlike *apologise* in Group 3. Only *complain* apears unproblmatic — until one compares it with *plane*. As in Group 1, th two comnst words produced fewst errs, altho th longr, less comn *because* surprisingly saw only 2% mispelngs, wile th shortr, very comn *would* came off over twice as badly at 5%. Th rarer but less aberant *apply, complain* both scord 7% errs, wile *writing* did worst in th group by far with 11% rong. Despite som anomlis of detail, it again apears that familiarity carrid mor weit than predictbility in facilitating acurat spelng.

Group 3: *apologise/-ize, unfortunately, allowance, receive, sincerely*
These words ar jenrly longr, with 2, 3, 4 or 5 sylabls, and of much loer frequency: *receive* (lematized) just tops 3k, *(un-)fortunate-ly* almost reachs 2k, wile *sincere-ly, apologise/-ize* and frends ar belo 500; *allowance* itself is a mere 300, but its base word *allow* and derivativs giv 5k. Th total for th Group amounts to c.11k. Th spelng of these words may look less blatantly aberant than, say, *come* or *would*, but they harbr numerus sutl traps. Th singl P of *apologise/-ize* is uncomn beside th domnnt PP of *appear, approach, appeal*, as wel as *apply* in Group 2; its singl L difrs from its greek compatriot *Apollo*; its secnd O is an indetermnat shwa-vowl; its G sujests a spelng with J; and its I contrasts with Y in *analyse/-yze*; all of wich errs togethr wud perhaps produce **appollajyze*. Th first sylabl of *fortunate* is homofnus with *four*; its U is jenrly asimlated to th preceding T, givng th sound of CH; th A is pronounced shwa, wich in *secretly* is E and in *definitely* is I; th E is silent; and th endng myt be rongly gesd at by analojy with *incidentally*; al of wich combined cud lead to th multipl mispelng **unfourchenitally*. *Allowance* ofrs th hazrd of dubl L (*allowed* is ofn mispelt *aloud*), th dubly ambiguus -ANCE endng (contrast *influence, immense, defence/defense*), and th final silent E; **alouens* myt therfor be a concevebl represntation. *Receive* is notorius for mispelngs with -IEVE (cf *achieve*, but also *eve, leave, sleeve*), but in adition th C is pronounced like S, and th final E is silent; **resiev* wud therfor not be an altogethr unintelijnt atemt. *Sincere* likewise is comnly seen with S for C, and with EER for ERE

(**sinseerly*). In th event, *allowance* incurd only 15% errs, perhaps thanks to th high frequency of *allow*, but th rest wer mor than twice as err-prone, being mispelt by at least one riter in thre. Th scors wer, in asendng ordr, *sincerely* 33% mispelt, *receive* 37%, and *apologise/-ize* and *unfortunately* equal at 40%. Th relativly hy frequency of *receive* was clearly no defence against its notoriety, but th bad shoing of *unfortunately* is less esy to explain; *sincere* and *apologise/-ize* wil hav been significntly less familir to th riters.

Group 4: *maintenance, immediately, necessary, occasionally, accommodation*
This most dificlt group consistd of th longst words (3, 4 or 5 sylabls), with rathr lo frequency (lematized as apropriat, frequencis wer *necessary* 4k, *immediate-ly* 3k, *occasional-ly* 1.5k, *accommodat-ion* 800, *maintenance* undr 500), tho th total (10+k) was scarcely belo that of Group 3. Th spelng hazrds they contain ar of th kind that afects th hole english languaj, rathr than being rare anomlis like *come, would, receive*. Four of th words contain dubld consnnts, 2 x CC, 1 x LL, 2 x MM and 1 x SS. Four contain post-accentul shwa befor L, or N, or R, twice befor N in *maint**enan**c**e*, and once befor L in *occasion**al**ly*, befor N in *accommodati**on***, and befor R in *necess**ar**y.* Shwas also ocur befor othr letrs in *immedi**a**tely, nec**e**ssary*, wher they myt be difrntly spelt as in *qui**e**tly, em**i**ssary* (ie, givng **immedi**e**tly, *nec**i**ssary*). Th spelng of /s/ varis between C in *maintenance, necessary*, and Ss in *necessary*, and th sound valu of C varis between /s/ in those two words and /k/ in *occasionally*. Th adverbial endngs ar pronounced virtuly th same in *immediat**ely**, occasion**ally***, but ar difrntly spelt. Th noun *maintenance* relates to th verb *maintain*, wich cud sujest (by th chomskyan principl of morfofonemic stability) th mispelng **maint**ai**nance*. Of th 980 subjects, over 410 mispelt *maintenance, immediately, necessary*, wile over 540 mispelt *occasionally* and over 670 mispelt *accommodation*. No dout som of th mispelt words containd mor than one err, and th total numbr of errs cud even be substantialy gretr than th numbr of mispelt words. In th lyt of such figrs, it wud perhaps be surprising if any of th points of dificlty noted abov faild to atract at least one err, but th report dos not giv th evidnce. No particulr corelation between numbr of errs and frequency of words seems worth noting for Group 4.

Th politicl implications of err-rates over 50% ar worth reflectng on. They represent in efect a majority vote of a sampl of th population in favor of simplr spelngs for th words concernd. Shud not th majority principl somhow also aply in deciding how a languaj is spelt? Th implications of these percentajs shud giv dictionry-makers pause: ar they democraticly acountbl or not?.

3 Ke statistics burid

Th survey has no coments to make on th orthograficl implications of th abov results, but it dos anlyz th percentaj of mispelngs for each word by certn categris of subject: by sex, by aje-ranje, by employmnt status, by socio-ecnomic group, and by educationl qualification. It dos not, howevr, giv th results for Wales sepratly from those for England, altho ther ar grounds (initial litracy aquisition in a languaj with predictbl spelng) for thinkng welsh subjects myt hav performd betr. As one wud expect, th socialy and educationly disadvantajd performd least wel, indeed th results may even hav somwat overestmated nationl standrds, since th subjects wer intrvewd at home, so excluding th exeptionly disadvantajd homeless, as wel as servng prisnrs.

Curiusly, th survey dos not avraj out th results across al th words for th varius social categris, and th readr is therfor not givn a clear overal pictur. Thus, tables ar givn for th percentaj of mispelngs of each of th 15 words in th thre mor dificlt word groups (2, 3, 4), one table for men versus women, and anothr table for five aje-groups; but th readr is left to do th arithmetic needd to produce a composit figr for standrds by adng th percentajs for th 15 words and dividing them to produce an avraj. If we spend a few minuts doing th calculations, we ar able to say that women

outperformd men not merely for evry word exept *would, maintenance* (men betr) and *complain* (sexs equal), but that overal women (27.9% mispelngs) outperformd men (32.7% mispelngs) by nearly 5%.

This failur to avraj th statistics for th seprat words is especialy surprising wen it coms to th analysis by aje-groups, since a distinct trend is therby blurd. A table hedd 'Younger and Older' givs th percentaj of mispelngs for each of th 15 words made by five difrnt aje-groups. Th overal avraj for each aje-group, once we hav workd it out, tels us that th worst spelrs by som distnce wer th yungst: th 16–24 year-olds avrajd 35.7% words mispelt; th next worst wer th oldst (55–60 year-olds), ho avrajd 32.8% mispelngs; th thre midl groups performd noticebly betr, tho with a slyt decline thru successiv jenrations, th 45–54 year-olds avrajng 27.0%, th 35–44 year-olds 27.5%, and th 25–34 year-olds 28.6% errs.

Aje-group	16-24	25-34	35-44	45-54	55-60
% rong	35.7	28.6	27.5	27.0	32.8

It is posbl, of corse, that these results ar a quirk of th relativly smal sampl (an avraj of undr 200 subjects per aje-group), but inevitbly we find ourselvs lookng for causes of th variations. Thus we may speculate that th education of th oldst aje-group very likely sufrd from th erlir scool-leving aje and th social turmoil (wartime evacuation, etc) of th 1940s, and ther myt may also be som efect of ajing. In terms of litracy policy, howevr, th decline in standrds shown by th yungst aje-group is th most striking findng. Ther hav been repeatd alegations in recent years that th litracy standrds of yung peple hav seriusly deteriorated, yet this claim has also been contestd, and th evidnce has hithrto been tantlizingly inconclusiv, restng mor on anecdote than on relybl statistics comparing standrds of successiv jenrations. From th data givn in this report we ar able to derive such figrs — and if they ar representativ, they ar surely alarmng. They imply that th spelng-acuracy of 16–24 year-olds is over 7% worse than that of peple ten years oldr, over 8% worse than those twenty years oldr, and nearly 9% worse than those thirty years oldr. Or, to put it a difrnt way, todays 16–24 year-olds mispel over 25% mor words than do ther eldrs. It has been sujestd that spelng-acuracy improves with aje and experience, but can such a factr explain th dramaticly reduced acuracy of th yungst jenration? Mor plausbl reasns ar perhaps a jenrl weaknng of readng habits in favor of television (se *JSSS* 95/2, p7, §5, on th IEA litracy survey), and a chanje in teachng methods away from fonics toward look-and-say and of atitudes away from authoritarian prescription toward fre expression. Th Basic Skills Agency tels us (persnl comunication) that it is "wary about making firm judgements about standards declining on the basis of this research", altho they "do tend to think that there was a period in schools when rather less concentration was paid to 'secretarial' skills than was perhaps desirable".

Th survey reports mor briefly on performnce in two othr 'secretarial skils' (as th National Curriculum describes them). Regardng punctuation, over 52% of subjects wer jujd to hav a poor undrstandng especialy of th rules for use of th apostrofe, but th yungst aje-group did not perform noticebly worse than th midl groups. On th othr hand, in th riting task (form-filng), mor in th yungst aje-group performd poorly or very poorly than in th othr aje-groups.

Th survey thus dos not hylyt th crucial figrs and issus concernng spelng, wich remain burid in its vivid and colorful diagrams and colums of percentajs. It tentativly concludes that "the results suggest that more people in the youngest and the oldest age-groups had difficulties than in the other age-groups" and "more younger people, some of whom have only left school recently, seemed to have difficulties than those ten or twenty years older", but th only lesn it draws is to say that "the results reinforce the need to improve standards" because job-aplications containng

mispelngs ar ofn autmaticly rejectd. Spelng reformrs wil inevitbly feel that th most importnt questions rased by th survey hav been alowd to go begng.

4 Implications for spelng reform

Efectiv tho th survey was in drawng media atention to poor standrds of riting (especialy spelng) in England and Wales, it scarcely begins to considr th implications of its findngs. It takes a naro vew of th importnce of good riting skils, wich is seen only in terms of avoidng rejection of job-aplications, and not in terms of enhanced powrs of comunication, nor as suportng readng skils and thus enhancing litracy standrds and therby th potential for th individuls educationl advancemnt jenrly.

Ther is also ambiguity in th surveys presntation of spelng dificlty. On th one hand, certn spelngs ar described as dificlt, but on th othr hand certn peple ar described as findng them dificlt. So dos th problm lie with th spelngs, with th spelrs, or with both? Wy ar only som spelngs dificlt, and wy do only som peple find them dificlt? To crak this conundrm, we need to examn th spelngs themselvs, and find out wy som peple trip over som of them (and indeed wy most peple trip over *occasionally* and *accommodation*).

Wy som peple succeed wher othrs dont, regardless of th dificlty inherent in any particulr spelng, is a question of individuls abilitis and education. As with al skils, th levl of proficiency acheved in litracy skils depends on a combnation of aptitude and trainng. Ther now seems to be wide agreemnt that th fashnbl trainng methods of recent decades wer misconceved, and that maxmizing litracy skils depends on systmatic developmnt of fonic undrstandng. But even if deficiencis in recent trainng (wich may partly explain th jenrationl difrnces observd) ar rectifyd, we must stil expect lak of aptitude (seen in its extreme form in dyslexia) to limit th litracy levls achevebl by som lernrs. So much for th dificltis orijnating in spelrs themselvs.

Wen it coms to identifyng th dificltis inherent in som spelngs, th analyses (wich ar based on extensiv experience of err-analysis) givn in §2 abov ar intendd to explain wat it is that makes som spelngs mor dificlt than othrs. Undrlyng them al is th alfabetic principl. Th fenomnl success of th alfabet as a riting systm around th world over thousnds of years is du to th simpl device of using th letrs to represent speech-sounds. Wen a languaj uses th letrs in this way to spel its words, ther ar no dificlt spelngs (exept insofar as long words may require mor careful atention than short words), and hy standrds of litracy ar esily acheved. But wen a languaj uses letrs unpredictbly, then dificlt spelngs and consequent litracy problms ar th inevitbl result. It is to minmize this danjer that most languajs hav in th 20th century modrnized ther spelng systms to keep them as closely alynd with th alfabetic principl as is practicbl. English has not systmaticly implmntd th alfabetic principl for nearly a thousnd years now, a histry of neglect wich has produced myriads of dificlt spelngs in modrn english.

But ritn english is not bound to sufr dificlt spelngs for evr: they can be made esir. Th Cut Spelng (CS) used for this revew takes a considrbl step in that direction, mainly by removing redundnt letrs. Th 20 words on wich th 980 subjects wer testd apear as folos:

- Of th five “esiest” Group 1 words *sorry, come, please, have, my* thre ar simplifyd, producing CS *com, plese, hav*.
- Of th “mor dificlt” Group 2 words *apply, complain, would, writing, because* again thre ar simplifyd, producing CS *aply, wud, riting*.
- Of th yet “mor dificlt” Group 3 words *apologise/-ize, unfortunately, allowance, receive, sincerely* four ar simplifyd, producing CS *apolojize, unfortunatly, alownce, receve*.

- Al th “most dificlt” Group 4 words *maintenance, immediately, necessary, occasionally, accommoda-tion* ar simplifyd, producing CS *maintnnce, imediatly, necesry, ocasionly, acomodation*.

We may confidntly predict that if these CS forms had been th target spelngs testd, th litracy scors acheved wud hav been much hyr. Yet CS by no means represents th ultmat in simplicity — a mor radicl reform than CS cud acheve much mor. Of th words unafectd by CS, *sorry, because, sincerely, complain* myt perhaps becom *sori, becoz, sinsirli, komplaen*, wile th CS forms cud undrgo furthr chanjes to produce *kum, pliiz, wwd, rytng, apolojyz, unfortiunatli, alouens, resiiv maintnns, nesesri, okaezhnli, akomodaeshn*. Ther is no need for *my* to be chanjed. Clearly th abov mor radicl respelngs cause such an upheval to th visbl form of english that ther introduction cud not be contmplated in th short term, but th fact that respelngs sujest themselvs for 19 out of th 20 words testd is itself eloquent testmny to th deeply unsatisfactry natur of th presnt spelng of english.

5 Conclusion

We may conclude by drawng atention to an ambiguity in th surveys subtitle “A survey of how well people can spell and punctuate”. Th word *can* is misleadng here: experience with mor predictbl riting systms shos that peple **can** spel a gret deal betr than they **do at presnt** using th traditionl orthografy of english. Th survey reports on todays disml standrds, but says nothing about futur posbilitis. Th Basic Skills Agency shud extend its horizons.

6 Aftrword: *Difficulties with Basic Skills*

If th Basic Skills Agency has yet to aknolej th ke lesn of its riting skils survey, this may be partly because of othr targets it has been concentrating on of late. Thre months befor publishng th *Writing Skills* survey, it reportd (*Difficulties with Basic Skills* by John Bynner and Jane Steedman, 80pp) on a mor substantial pece of reserch. In th latr booklet it examns th social causes of th litracy and numeracy problms reportd in its 1994 survey *The Basic Skills of Young Adults* (se §1 abov). It concludes that th most importnt factr determnng proficiency in basic skils for yung adlts is ther childhood bakground, wich if unfavorabl tends to produce a cycl of lo achevemnt thru successiv jenrations. This unsurprising findng is bakd up by som quite sofisticated statistics, and leads to th cal for mor suport for th home and for th yungstrs concernd.

That ther cud be somthing fundmently rong with th very medium thru wich scools atemt to instil litracy skils jenrly is not mentiond. Ther is no dout that far too many children ar seriusly handicapd in ther education by varius kinds of deprivation, and that this handicap shud be adresd. At th same time, spelng reformrs wil need to go on remindng th Agency and othr oficial bodis concernd with litracy standrds that deprivation is not th only handicap afectng those standrds. Th caos of english spelng handicaps al lernrs, and to concentrate on those ho ar most severely afectd is like advocating dietry improvemnts only for th most severely malnurishd wen th hole population is sufrng from malnutrition. Giv special asistnce to th most deprived, and we may expect them to benefit. Simplify th spelng (probbly at less cost), and we help evryone.

[Journal of the Simplified Spelling Society, 20, 1996/1 pp34,35 in printed version]

12. The *Cut Spelling Handbook*, 2nd Edition, Foreword

Christopher Upward

The revised and expanded second edition of the Cut Spelling Handbook appeared in April 1996. The foreword, reprinted here, sets out the development of Cut Spelling from its first to its second edition.

Impact of the first edition

The first, limited, edition of the Handbook to Cut Spelling (CS) was published in 1992, and in just over a year all its 250 copies had been distributed. The general response was sufficiently positive for the Simplified Spelling Society to decide that the Handbook could not be allowed to remain unavailable for long. There were two possibilities: either a simple reprint could be produced to meet the immediate continuing demand, or else, with an inevitably longer delay, a revised and expanded edition could be prepared which would build on the experience gained during the intervening period. In late 1995 the opportunity arose to produce such a second edition, which now appears as this volume.

The first edition of the Handbook aroused wide interest. The publicity generated by its launch was considerable and enduring. Over the airwaves CS was covered by the BBC World Service, with further specific broadcasts going to New Zealand, Nigeria and South Africa, as well as being heard from numerous national and local radio stations in the United Kingdom. Press reports were syndicated across the United States, and appeared at least in France, Germany and the Netherlands, and in numerous newspapers in the United Kingdom. On a more academic level, CS is now recorded in general reference works on the English language as an innovative proposal for the modernization of English spelling, and has been analyzed in more specialized studies. Basic information on CS is accessible (and is being accessed) on the Internet. Publishers have proved willing to accept material in CS, with (so far) one research report in a scholarly journal and a chapter in a collection of conference papers printed in it, and other items forthcoming. Articles have appeared explaining and demonstrating CS in professional journals addressed to teachers of basic literacy skills and of English as a foreign language, as well as targeted at more general readerships. Conferences have been addressed both on the subject of CS itself, and using CS for illustrative material. CS has been regularly used in personal and professional correspondence around the world (for instance to Australia, Canada, China, India, Japan, Saudi Arabia and the USA), both in hard copy and in electronic form. And of course readers of the Simplified Spelling Society's publications have now been familiar with CS, in its evolving forms, for a decade and more.

It may not be altogether implausible to claim that, since the first edition of the Handbook appeared in 1992, CS has become more widely known than any previous proposal for reforming English spelling.

Lessons for the second edition

Since 1992 a rich body of experience has accumulated in the use, reception and operation of CS. Well over half a million words of continuous text on a range of subjects have been written in CS, almost certainly far more than in any previous reformed English orthography. This practical use has clarified some uncertainties, highlighted some problems, and reinforced the advantages of the system from the writer’s viewpoint. Readers’ reactions have been expressed through numerous comments received, ranging from the abusive to the enthusiastic. These have

described readers' initial responses to CS, their process of acclimatization, and the difficulties they may have encountered in decoding individual words. Readers have included young and old, native and non-native speakers of English, and professionals such as academic linguists and remedial literacy teachers, alongside lay persons with a general interest in alternative ways of writing English.

The second edition has benefited not only from all this practical experience, but also from improved understanding in related areas. Research on literacy teaching methods, especially in the 1980s, has become better known, confirming the phonic approach (ensuring beginners appreciate how sounds are represented by letters) as fundamental to proficiency in reading and writing. Recognition of the importance of phonics, despite its limitations in English, highlights the centrality of the alphabetic principle to a good writing system, namely that the letters should predictably represent sounds, and sounds be predictably represented by letters. Cut Spelling's claim to satisfy the demands of phonics and of the alphabetic principle far better than does the traditional orthography of English (TO) — though still not perfectly — can therefore now be advanced more forcefully than before.

Fresh support for CS also comes from three other directions, historical, geographical and technological. Historical research has revealed that a broader range of CS forms was in use in the Middle English period (eg, in the 14th century, the age of Chaucer) than had been suspected when the first edition was prepared. Geographically, it has recently become apparent that one effect of CS is to remove many arbitrary disparities between English spellings and their equivalents in other, mainly western European, languages, so making foreign language learning easier both for native and, especially, non-native speakers of English. In yet another area, that of information technology, the strides made in the past few years in developing electronic written communication (known by such terms as the Information Superhighway, the Internet, the World Wide Web, etc) open up new possibilities for implementing English spelling reform. The effect of the new technologies on the literacy practices of younger generations is emerging as a further argument for simplifying the alphabetically grotesque spellings currently promulgated as correct in TO.

Changes between 1st and 2nd edition

This second edition of the CS Handbook has thus been strengthened by the new knowledge and clearer perspectives that have emerged in all these areas. At the same time a number of specific changes have been introduced in the presentation of the CS system. The readability of the Handbook has been enhanced by more generous spacing of text and by using small capitals to indicate letters of the alphabet and spelling patterns, instead of the traditional cumbersome and unattractive angle brackets; thus what appeared as <a>, <b>, <c> in the first edition now appears as a, b, c.

Scarcely any changes have proved necessary in the proposed spelling of individual words in CS. Two minor, isolated instances may nevertheless be mentioned:

1) it is now thought better to reinforce the recommended rules for keeping SS (see Part I, Chapter 3, Rule 3, §2.4) and write CS *messaj*, rather than to harmonize this one word with the rare single s of its rhyme *presaj*.
2) it became apparent that the second E of TO elsewhere was redundant by Rule 1, E.1.1.13, and CS now recommends *elswher*.

In Part I, Chapters 1 & 2, it has been possible to introduce or expand discussion of various points which in the first edition had given rise to objections and/or misunderstandings. These include the concern that redundant letters are actually important in ensuring comprehension, the question of how far CS can suit speakers of all the world's accents of English, and the misconception that CS aims to regularize all the irregularities of TO.

The main changes to Part I, Chapter 3, which contains the detailed arguments for cutting particular letters from TO spellings, take the form of clarification. Thus the dilemma of the TO alternatives *carcase/carcass*, with first syllable stress, has now been resolved by analogy with atlas, so distinguishing CS *carcas*, and several parallel spellings, from the second-syllable stress pattern of uncut forms like *erase/morass* (Rule 1, E.1.1.13). Similarly, a new comprehensive analysis of the l, m, n, r + t sequences (Rule 2, §1.5) has shown that the advantages of the controversial long consonant strings in CS forms like *exlnt, govrnmnt, contnnt, cormrnt* are systemically far greater than originally appreciated.

But while those changes represent a strengthening of the case for CS, Chapter 6 (Part 2: Stopng short of CS, §2.3.8) now sets out in detail the various anomalies (loose ends, rough edges, warts) in CS that were not previously collected together at any one point in the Handbook, or have, in a few cases, only become apparent since the first edition was published. These anomalies range from lists of the unproblematic mergers of the *peace/piece* > *pece* type and the slightly more problematic *plaice* > *place* type, to a few gross heterophonic ambiguities of the type *err/heir* > *er*. Although none of these anomalies calls into question the integrity of the CS system as a whole, they are now more clearly recognized as blemishes, and ways of preventing them are discussed.

It goes without saying that the opportunity of a second edition has been taken to correct whatever misprints and other small errors had come to light in the first edition.