[Later designated Journal 2]

Simplified Spelling Society Newsletter
1986/1 (Spring)

Contents

The Society
Founded in 1908, the Simplified Spelling Society has included among its officers: Daniel Jones, Horace King, Gilbert Murray, William Temple, H.G.Wells, A.C.Gimson. Its stated aim is to "bring about a reform of the spelling of English in the interests of ease of learning and economy of writing".

The Society's present officers are:

President: Professor John Downing.	Public Relations Officer: Mona Cross
Chairman: Chris Jolly	Enquiries to the Secretary
Secretary: Stanley Gibbs	Subscriptions (£5 p.a.) to the Treasurer
Treasurer: Laurence Fennelly	

The Newsletter
The Simplified Spelling Society Newsletter appears 3 times p.a. Editor: Chris Upward, Material for the 1986 No.2 (Summer) issue should reach the editor by 31 July 1986.

[SSS Newsletter Spring 1986/1. Later designated Journal 2. p2 in the printed version.]
[Chris Upward: see **Journals**, **Newsletters**, **Pamphlet**, **Leaflets**, **Media**, **Book and Papers**.]

1. Editorial. Chris Upward

THIS ISSUE
This Newsletter is half Conference number and half more regular format. The remaining conference papers are now included, but we also resume publication of non-conference material, including correspondence, occasional articles and David Stark's series.

However this number also contains an important innovation: a feature article. The idea is that feature articles shall represent a substantial contribution from an acknowledged specialist, and the Society was fortunate in being addressed last autumn by Professor Frank Knowles on the subject of Information Theory. Readers of the edited transcript of his talk (Item 3) will be impressed by this synthesis of such diverse fields as cryptography, computing, the orthography of languages that use

non-Roman alphabets, and the psychology of reading. Interestingly, a spelling-reform concept for which Information Theory appears to have some rather direct relevance is Cut Spelling. But in more general terms, this feature article is also a reminder that spelling is a field with wide horizons that we all too easily lose sight of, as we wrestle with the minutiae of sound-symbol correspondences or the tricks of teaching the English writing system to struggling learners.

Edgar Gregersen's paper, as readers present at the Conference will remember, confronts spelling reformers with an objection to reform which is widespread in the academic community, both in the United States and in Britain: namely that a phonographically regular system would reduce the visual connections between the many linked words like sane/ sanity, author/ authority if it respelt them perhaps as saen/ sanity, authr/ authority. Since these views are influential, reformers need to take them seriously and consider their response.

READERS' REACTIONS TO NEW STYLE
Numerous readers have written to say how much they like the new style of Newsletter, and there has been some curiosity about how it has been produced. Articles are firstly entered by Macwrite word-processor on an Apple Macintosh personal computer, then proofs are produced, and corrections and typographical ajustments made. Printing of the final master is by laser printer on A4 sheets, from which the reprographics unit at Aston University runs off the requisite number of copies on A3 sheets for saddle-stitching into the covers. The total cost has been about a third more expensive than before, but it may be brought down by more comprehensive computerized typesetting in future.

In a natural anxiety to see visible progress, some readers have urged that the Newsletter should adopt the 5 SRs agreed at the 1984 AGM. Others have felt that though mostly those forms are undoubted improvements on t.o., they neverthless leave some loose ends (the <-igh> words for example), define DUE inadequately, and include such dubious forms as SRI *et* for *ate*, and that in view of their shortcomings it is premature to trumpet them too publicly. Furthermore, the Society's present working party hopes to refine and develop them within a coherent overall system, and the time to consider a house-style will be when that task is complete. Nevertheless contributors are welcome to use whatever spellings they prefer, and (barring editorial oversight!) these will be used in publication.

PUBLICITY LFAFLET
Enclosed with this *Newsletter* is a leaflet designed to spread publicity for the Society as a vehicle for spelling reform and hopefully attract many new subscribers. Readers are encouraged to photocopy the leaflet and circulate it — libraries, newspapers and education advisers would be useful targets, but it would be interesting to see the response if each reader posted a copy on a notice-board in a public place, such as a public library.

PUBLICATIONS OF INTEREST
The editor has 2 copies of Harry Lindgren's Spelling Reform A New Approach (Sydney: Alpha Books, 1969) which he will gladly post to the first two readers who send a stamped-addressed envelope big enough for this slim paperback. The book is strongly recommended: though readers may not agree with it all, they cannot fail to enjoy its lively style, brilliantly apposite cartoons (never has opposition to spelling reform been so devastatingly lampooned), and above all the sheer creative imagination shown in the exposition of Phonetic A and B. And of course it is the bible of die advocates of SRI.

The editor also receives copies of other relevant publications which he will be glad to lend for a week or two to readers who send an A4-size stamped-addressed envelope. Such are the Australian Spelling Action (April–June 1985) and papers from Roman Lipi Parishad in India (see **Madhukar Gogate, p.18**).

Readers may have seen the attractive new quarterly English Today, published by CUP (E9 p.a. in UK) and edited by Tom McArthur, who is accumulating so many contributions on spelling reform that he is planning a cover theme on the subject in 1987 (SSS Conference Year!). This is a boat the Society clearly must not miss.

NEXT ISSUE
The **1986 No.2 (Summer) issue** will feature an edited version of Dr John Wells' key address to the Society last January on the implications of English accents for spelling reform. By coincidence it leads on from and expands some of Edgar Gregersen's remarks in this issue.

2. Correspondence

From Mona Cross, Public Relations Officer & C'ttee Member, England.
[Mona Cross: see **Journal 28**, **Newsletters**]

Tho I am no longer editor of the *Newsletter,* I should be glad to retain a link with members of the SSS and the interested people who used to rite to me. So I hav decided to rite to tell you what we ar doing or ar particularly concerned about. Besides that I want to thank you for your own individual effort in sending advice on the Nue Spelling scheme which we have nearly completed and to suggest what els, as members, you mite, wish to do.

If you are in London when we have our meetings, we should be delighted to see you. A plesanter group of people would be hard to find. Discussions are among people whose dedication is strong, but who listen with patience to other people's point of view. The next meeting is at the Y.W.C.A., Great Russell Street, London, on April 26. It will be the Annual General Meeting — which to me sounds discouragingly formal, but which will be much more interesting than it sounds.

I have been an SSS member for a long time — since before the first conference (1974). It is disappointing to me*,* and to many members frustrating, that spelling reform has not taken place. Inspired individuals took matters into their own hands by making a scheme and spending endless time and money in promoting it. There were Vic Paulson, Herbert Wilkinson, Dr Gassner, Axel Wijk, Reg Dean and Ken Tillema and more people than I can name working on their own. Their influence rests only with those whom they have contacted. But altho that included publishers and politicians, it was not and is not a strong influence. Neither does it come into the orbit of 'the masses' who will hav to use it.

Somebody has to grant permission for the different spelling to be used in official documents and the publication of books. I myself do not rite to the bank manager in a 'reformed' spelling and neither, I guess, do you. What would you do to make it possible for all the teachers to use a reformed script, or to allow the American spelling, which children see in their reading books, to be used in their own writings?

The SSS Chairman, Chris Jolly, supported by all the Committee and helped by Alun Bye, is trying to bild on a 'reformed' spelling which is actually used. It is used for advertisements only such as *foto* for *photo* and *Kwiksave* for *Quicksave.* Letters have been written to the firms concerned, asking them to extend the use of their reformed word. That's fine. It's a start in public involvement. Can you, and will you, help in fostering this idea, and have you any notions of your own?

We should be grateful for your help. We need you, for a committee alone is not powerful enough to change the stubborn or uninterested mind of most of our friends, neighbours and countrymen. Keep supporting us!

P.S. We hoped you'd write letters to the local newspapers and send copies to Mr John Ogden, but as far as I know, you didn't. Would it be a good thing for the Committee to produce a letter which could be sent by any member to any paper? What do you think?

From Doug Everingham of Australia, concerning the SR proposals accepted at the Simplified Spelling Society's AGM in 1984.
[Doug Everingham: see **Bulletins**]

Here are some objections to your proposals. They could all be by-passed by adopting Lindgren's approach:

SR:ph is based on the assumption that the user knows existing <ph> spellings for /f/ sounds. It is a rule separated from SR:ough which similarly changes <gh> and assumes a knowledge of <ough> spellings with /f/ sound. It does nothing about the confusion of *<f, ff>* e.g. in 'ruf'. The logical way to fix this is what for now we can call SR:f (to label it by sound, not by spelling), which could read: the consonant common to *off, if, cough, sulphur* and *sapphire* may be written F. But it is not appropriate to do this for *off* till SR:v is attended   to and widely used in the form of *ov*. This SR:v would   read: the consonant sound common to *of, have, navy, navvy and nephew* may be written V.

This in turn is not appropriate till we get rid of doubling of consonants as a device for distinguishing the different vowel sounds in *navy/ navvy.* This will come when we tidy up the difference between the vowel components in *table/ babble, hater/ hatter, cape/ cap* and thousands of others, so why have 2 rules for partial rationalization of /f/ sounds when one rule will do for all of them, if you will be patient enough to follow a logical progression? Besides reform of the *navy/ navvy* distinction we should also reform the *rifle/ riffle* distinction before contemplating SR:f. Lindgren and most reformers have made clear that spelling should follow sound. Spelling reform should do so too. It should not correct some deviations from one sound at a time. Let our children first have one sound, then another, of their speech on which they can rely when they write! *NONE* of SR:ph, SR:augh, SR:ough or SR:DUE offers this. Incidentally, how would you write the intellection *phew*? It may be the bilabial <ph> of Latin which begins the word. The Romans presumably invented the <ph> to get as near to a Greek bilabial fricative which the Greeks used to write (and Russians still write) as B, sounded rather like our /v/ but with lips touching, a sound absent from Latin.

SR:augh does not have the consonantal complications of SR:ough, but agen why not wait till you can use *everywhere* a single *spelling* for the vowel in *all*, *aught, ought, law, awe,* and decide if it is to be distinct from that in *oar, ore, o'er, four, Boer, boor?* And then whether <au/aw> is not the most rational way to show the component sounds of the diphthong in *sauerkraut, mauser*, *landau, Macau,* rather than the single vowel sound of *awe*? Let's decide those   consistently before tinkering with changes that should be consequential. Your SR:au produces a final <au> spelling that is not only inconsistent with the diphthong structure but is also inconsistent with the existing final spellings *landau; caw daw haw law thaw gnaw raw saw*...

SR:ough1. Here agen you put a final U where existing usage has W: *bow cow how now row sow vow.* 'I'he existing English spellings ending with <ou> as far as I recollect use those letters for their sound in *bijou caribou route roulette*. Pronouncing dictionaries often use <ou, ow> for the diphthong sound in *coal dole low mould poultry soul toe yoyo.* They use <au,aw> for the sound as in *sauerkraut, bough*. Let's wait till we decide on a logical spelling for the *awe* vowel before we go down the <ou> road and so perhaps have to backtrack with a later SR.

SR:ough2. See above on SR:augh and ough1.

SR:ough3. See above on SR:ph.

SR:ough4. Here you have 4 rules, not one of them offering beginners a *consistent* spelling for a consistent sound. The <o> or <oe> if consistent, would extend to replace your *doh*. So why is it not *doe*? There is a conflict with an existing spelling in either case. *Please avoid rules which are limited by or definable by* existing spellings! Make rules limited only by the sounds concerned and you will not need exceptions. Adopt one of the 40 approx. sounds of English *one* at a time *completely* and in the right *sequence* and you will avoid confusion. Your 'thurra' is a triple rule:
1 double R to make preceding vowel short
2 sound U before double letter as in *up*
3 Use A for final indistinct vowel.
Three rules for rationalizing one word is poor rationality.

SR:ough5. This is inconsistent with the sound of U in SR:ough3/4. Few existing spellings give final U this value. The only ones I can recall are *tutu, caribou*, and the Chicago Tribune's *thru* which has become well recognised but with no analogous change in eny other word. Not a promising start for a rational spelling code.

SR:DUE. This is the most confusing and delaying proposal you have made. Its applicability is not clear in hundreds of cases like *awe aye ey borne bourne come catalogue done foe gone judge more none programme statue rue toe*. The simple solution is to abandon rules based on bad spellings and go for rules based on clear speaking.

Where the US (Webster) usage differs from UK (Oxford) dictionaries on pronunciation, go for the commonest usage. There is no such problem with SR1 and that is one of the reasons for Lindgren's choosing it. A similar simplicity would apply with an SR for most short clear vowels with provision for a variant. Thus the vowel in the *lad plaid add plait sate have salve* (US) would have provision for the variant /a:/ sound in some words. Many Australians say *aunt* like *aren't*, while most US speakers say it like *ant*. We don't mind alternatives like *reinforce/reënforce, manoeuvre/maneuver, programme/program, catalogue/catalog.*

Similarly for the vowels *of in on urn all ski one put rue*. Once these are consistent, the diphthongs *was may my boy mow now* (*hey ay oil low kraut*) can be chosen in the order that will avoid clashes and conform with the components of the diphthongs. Then it will be unnecessary to have doubling of consonants, one of the sources of confusion in Italian, French and English (but not Spanish).

My closing appeal: promote SR1 and existing, normal controversial alternatives like the *realize* for *realise*, *thru* for *through* perhaps and nothing else still SR1 is widely known. Then let's agree on SR2.

From Professor Ayb Citron, Better Education thru Simplified Spelling, USA.
[See **Bulletins**, **Anthology**, **Journals**, **Newsletters**]

Some time bak we were discussing *ow* versus *ou*. Chris Jolly suggested we obtain some indication of public preference on this question. A frend of mine, a counsellor at Oakland Community College, opened some doors there for me. Between December 4 and 10 I was in English or psychology classes for from seven to eight minutes each. In each class I distributed ballots, as here.

THIRTY SECOND RESEARCH IN SPELLING

If you used only one spelling for the vowel sound in these words, which spelling would you prefer'!

These are all spelled with *ow*.
cow	owt
sow	showt
how	trowt
now	clowd

These are all spelled with *ou*.
out	cou
shout	sou
trout	hou
cloud	now

I like *ow*. [………….]

I like ou. [………….]

Ow starts and ou starts were distributed alternately to the students. Then I opened with three points:

1   Our traditional spelling is often quite irregular.
2   One of its irregularities consists in using different spellings for the same vowel sound, for example, <ou> (*cloud*) and <ow> (*cow*).
3   The ballot they mark is *not* a vote for changing our spelling. It is completely hypothetical. It asks: If we were sometime in the future to use only one spelling for the above sound, which spelling, <ow> or <ou>, would they prefer? (If a male, place a small M in the upper right corner of the ballot.)

I obtained a total of 150 ballots, 113 female students and 37 male student. They voted 100 for *ow* and 50 for *ou.* The 113 woman voted in the same proportion, 76 *ow* and 37 *ou.* The men voted in the same proportion, 24 *ow* and 13 *ou.*
Using the same or similar ballots and procedure, perhaps you would wish to obtain a sampling. Let me know if you want to approach this.

Correspondence has also been received from

Harvey Barnard (Tacoma Washington),
Robert Craig, (Weston-super-Mare),
Rae Elser (New York),
Stanley Gibbs (Leicester),
Madhukar Gogate (Bombay),
Edgar Gregersen (New York),
William B Hornback
(Spelling Progress Quarterly,
Pennsylvania),

Garry Jimmieson (Queensland),
Richard Lung (Scarboro),
Roger Mitton (London),
Peh-ling Lee (Jiangxi Province, China),
Edward Rondthaler (New York),
David Stark (Cumbernauld),
Michael Stubbs (London),
Tom McArthur (Cambridge)
Valerie Yule (Aberdeen) and others**.**

3. Information Theory and its Implications for Spelling Reform
Francis Knowles

[Professor Knowles is Professor of Language in the Department of Modem Languages at Aston University, Birmingham and his interests lie particularly in the application of computers to translation and lexicography. This article arises from a talk he gave to the Society on 12 October 1985, and was transcribed and edited with his permission.]

1 ORIGINS OF INFORMATION THEORY

1.1 Shannon & Communication Theory
The idea of Information Theory was first outlined in print in 1948–49, when the eminent American mathematician Claude Shannon published the results of work he had been engaged on for several years; however he first called it Communication Theory. Shannon was in the service of the United States government in the Second World War, employed on traffic analysis and cryptological duties. Anyone who has had the remotest dealings with that type of work — and I have to plead slightly guilty myself — would see immediately why Information Theory and language are so intertwined, and how many useful things can be found out about how language works as a result of the methods used. The question Shannon was particularly concerned with was whether it is possible to measure the information that gets transferred from one person to another in daily intercourse. He decided that it was possible.

1.2 Levels of language
 Shannon was concerned with the lowest level of linguistic analysis, which we will call syntactic analysis. Following other American linguists he saw language as existing on three levels. Firstly, syntactics, which is just how symbols combine with each other, or, to use anthropomorphic terms, how symbols can be friendly or unfriendly with each other, how they muster, how they group, how they keep their distance from each other. The next level up would be the level of semantics which is the process of allocating meanings to the symbols. The highest level would be pragmatics, how the users of symbols actually handle them in practice. Other philosophers and linguists have taken things further by suggesting that there is in fact an aesthetic dimension, but we need not concern ourselves with that here.

1.3 Efficient Messages
During his military service Shannon was dealing with communications, especially intelligence communications. He would be faced with a message, and had to ask how to make it efficient so as not to waste battery power or electric energy in transmitting it, and how to protect it from unauthorized eavesdropping. He therefore tried to devise rules to help people code linguistic messages efficiently. Obviously military and diplomatic communications are a special form of communication, but what Shannon did has a wider validity, which is the basis of the communication theory we shall be discussing.

2 CONVENTIONS
2.1 Understanding Ambiguities
Most linguistic signals use a set of conventions, and effective communication requires the receiver of the signals to know the conventions and be able to interpret them when there is possible ambiguity. One may for, instance encounter the symbol ! by the roadside, which we all recognize as an exclamation mark. But if one asks in general terms what an exclamation mark is, one would

probably say that it is an orthographical sign used in text which can't be pronounced but indicates some rhetorical features of a statement. It would be typically used in, say, a letter from Mr A to Miss B, when he might be tempted to use one of these signs after the words 'I love you'. However the sign on the roadside was a symbol from the Highway Code to warn of potential danger. It is also a sign of considerable importance to mathematicians, not least to those who work in Information Theory: 3 factorial, i.e. 3 x 2 x 1, is 3! It is also a sign used in chess literature when the author wishes to indicate a particularly good move; and in like manner a question mark would suggest that the move was dubious. So conventions may entail problems, such as ambiguity. I need scarcely add that the English orthographical system is recondite with conventions whose ambiguity makes them less helpful than they might be.

2.2 Information Content

Shannon aimed to go beyond these conventions and concentrate on making the written language as efficient a vehicle for communication as possible. He started by asking how he could measure it — a very difficult question. We can get a glimmering of the sort of approach he adopted if we say 'I want to kill two birds with one...', but leave the sentence incomplete. Every native speaker of English knows that the next word should be *stone* — and would be surprised if it was *pebble* or *boulder* or *rock*. Shannon concluded that if you knew the word was going to be *stone*, then it conveyed no information whatsoever being entirely predictable; its information-content was zero. But if we say, 'Jack Smith has gone and ...', and stop there, no one would be able to say what the next words were going to be. Here Shannon would say, at that moment the atmosphere was pregnant, and whatever followed would have been rich in information-content.

2.3 Statistical Knowledge of Language

So how do people know that the next word was going to be stone? It is because we are inured to that fixed idiom of English which always occurs in that context. But a foreigner might not have that knowledge and would then be unable to predict the next word, because he wouldn't have been exposed to English in daily discourse from early childhood, and would therefore not have built up his mind, or brain, the necessary statistical database, so to speak, about how the English language is used. That, I think, is proof, if proof is needed, of the importance of the statistical undercurrent to language. There is a quantitative side to language which most people are very adept at keying into and handling, though of course they couldn't describe it in extenso. Grammar-books may attempt to do so, but I doubt very much whether even the most comprehensive grammar-book could tell us the complete, exhaustive rules for the use of the definite article in English. It will lay down a whole series of rules, occupying many pages, but at best they might cover only 85 or 90% of cases. The rest is covered by that feeling for language we only have from the way we have been brought up and linguistically sensitized.

2.4 Applications of Statistics of Language

Shannon tried, and others have tried after him, to use this statistical aspect of language as a sort of crowbar to get inside the way messages are communicated, and so make telecommunications efficient. There are all sorts of ways in which this can be done. Shannon's work was in the field of cryptology and codes — the attempt to regularize and invent words to symmetrize information, so that they can be compacted into neat packets. This task was greatly facilitated of course by the advent of computers with what the computer scientists call fixed architectures. It sometimes makes the computer the boss, rather than human beings, which is always a pity, but no doubt that particular danger will recede.

3 CODES AND COMMUNICATION

3.1 Spoken and Written Codes

Let us look at some of these features that Shannon thought of as being measurable, such as codes, for instance. English exemplifies how codes are used rather well.

Writing and speech constitute different codes. If you tell someone how to use a public telephone box, you use the spoken code and probably start by saying, 'put a coin in the slot'; but printed instructions would be in the written code and probably say *'insert* the coin in the slot'. If such a code of conventions is ignored, it creates a curious impression. But that is not a code of direct interest to us from the point of view of Information Theory, though it is of very great interest to students of language in a more general sense.

3.2 Reconstituting Garbled Messages

To take a rather dated example, suppose a young man sent a telegram to his girlfriend saying 'COMING AT THREE LOVE BOB', and imagine the message actually received was something like 'COMIHG AQ THZEE LIVE ROB'. There might be confusion if there is both a Rob and a Bob in the young lady's life, but otherwise she should have no difficulty interpreting the message correctly. The original information was garbled — but communication was unaffected because the receiver of the message was able to reconstitute it using her knowledge of the statistical probabilities of language. Positive Information Theory has to do with information as sent by the transmitter and as received by the recipient. If anything goes wrong with the communication, one needs to know how the information got lost and whether it is in fact re-constitutable. Now suppose the message sent had the same spoken form, but was written 'COMING AT 3 LOVE BOB', and suppose that in the process of transmission it became garbled to 'COMIHG AQ 2 LIVE ROB'. With the symbol 3 changed to 2, there is no context in the telegram to help reconstitute the correct hour. So Bob arrives at 3, and his girl-friend is less than pleased. The point is that numbers have no contextuality, and if we are sending telegrams, we are well advised to spell out any numbers alphabetically to avoid such dangers of garbling.

3.3 Chinese/Japanese Telecommunication

Let us now take a different context. Imagine a businessman sent to Japan to negotiate a business deal, with permission to spend say £100,000 for his company. When he gets to Tokyo, he finds out that the price has risen to £150,000, so he has to get permission to spend extra money. He could go to the Japanese Post Office and telegraph, 'Do I have permission to spend £150,000, Jack'. What he would actually do, if he were well-trained, is indeed to go to the Post Office and send a telegram, but it might read XPQRSZ 5683451, which is cheaper to send, and doesn't clutter up the telecommunications systems. It looks meaningless, but it isn't in fact, because the recipient will have a dictionary in which he can find that it means 'Do I have permission to spend' the sum in question. But this can only be used where this is the accepted code — it's no good just looking round the door and saying to someone XJQ £150,000. In other words there is a conventionality here too. In some countries the convention has to go much further than that. We all know about the vagaries of the Chinese language, where there is no alphabetic system, but a large number of stylized ideograms that prompt the memory about how the word should be pronounced but don't directly correlate with the pronunciation. Chinese school-leavers are supposed to know 5,000 characters, but a typewriter with 5,000 keys on it is inconceivable. However if you don't have a typewriter, you certainly can't have a teleprinter in China. To send telegrams in China, you have to use a code-convention which effectively means going to the Post Office and saying 'I want to send a telegram to someone with surname type 73, living in city 7, street something or other, and the message type is 425', which might well be 'Arriving Peking, time-type 67' — for 10 a.m. perhaps.

4 REDUNDANCY & INFORMATIVITY

4.1 Redundancy
All these forms are conventions. We have seen that when information is spoilt en route, we can sometimes reconstitute it. This leads to the notion, which Shannon initially formulated, that if this is possible, there must be some redundancy involved in the way the information was recorded in the first place. As a mathematician, he wanted to measure this redundancy. It is possible to say that written English is 50–60% redundant in this very basic aspect — quite how that is worked out we shall see a little later. Redundancy is one of the key concepts of Information Theory, and if Information Theory aims to do anything, it is to iron out that redundancy, and get rid of it where that is a sensible thing to do. It is not always a sensible thing to do, as we observed with the garbled telegrams. It has been shown that languages display a sort of constructive tension between redundancy and informativity, and if you remove redundancy, then presumably you are increasing the informativity. However there can be occasions when the information flow in writing and in speech, but particularly in speech, is just too rich, and then redundancy appears to be a necessary feature of language. All we are talking about is dispensing with it in certain circumstances, where that seems to be a sensible thing to do.

4.2 Excessive Information Content
Here are some highly contrived examples of what is meant by information being potentially too rich. Let me read out the following sentences:
The audience who just heard the person who cited the king who said my kingdom for a horse is dead is an example is a psychologist are very patient indeed.
That sentence is virtually incomprehensible. Or:

The picture that the script that the novel that the producer whom she thanked discovered became made was applauded by the critics.

That too is almost impossible to follow. If you bracket suitable word-groups, however, you could work out that it meant:

She thanked the producer who discovered the novel that became the script that made the picture that was applauded by the critics.

That example reminds one of 'the house that Jack built', which goes to show that certain structures in our language which have evolved historically have not done so by accident, but because they proved capable of conveying information. Now Shannon would presumably say that such English sentences were wayward, but their information content is exactly the same however it is arranged. That is a bit of a conundrum for linguists.

5 STATISTICS OF LANGUAGE

5.1 Distribution
If Shannon's aim was to measure total information-content in a message and leave it at that, further work has concentrated on measuring information as the text proceeds, either from left to right in written text or in time elapsed, in spoken messages. To give a silly example: supposing I happened to know that in a particular book that was 50,000 words long the word *orthography* occurred 1530 times, you would be surprised if the last 1530 words of the book were all *orthography.* So there is a distributional side that is exceedingly important: it is important at the level of syntactics, and at the level of word-fragments, such as graphemes, that people like ourselves are interested in.

5.2 Statistics and Cryptology

Shannon realized that if so much depends on statistics, then the statistics must be established, so that we know whether or not we are on safe ground. This led to an amazing upsurge, both during the war and since, in the counting of linguistic elements, ranging from letters of the alphabet and phonemes upwards, and some very extensive results are available, relating to most of the world's languages. A lot of them are used of course for purposes such as cryptography, the making and breaking of secrecy systems. In fact the cryptographer turns the Information Theory coin upside down, because he wants to obliterate the tell-tale statistical characteristics of text, so that the unauthorized eavesdropper hasn't got a handle with which to interpret the coded message. Shannon operated in the cryptology mode for quite a long time, and I always draw students' attention to the fact that though they always seem to be aware of one of his papers in the Bell Systems Technical Journal, they don't seem to be aware of his other paper about cryptology.

5.3 Use of Computers

Computers, which were first coming into prominence in the late 1940s, were ideal machines to calculate some of these statistical characteristics of language, and they have remained so ever since. There is still a counting-game going on and a very intricate analysis of many different languages — not only English. Other languages have been less thoroughly researched but there is still a vast quantity of data on file.

6 ORTHOGRAPHICAL WORDS

6.1 Shorthand

Anyone interested in spelling reform must inevitably take an interest in related fields, such as shorthand. Shorthand aims to record speech in writing with a high degree of accuracy as fast as it is uttered. A typist will reconstitute the shorthand notes or outlines later, normally within 24 hours, into running text. This brings me back to emphasize an obvious point I made earlier, which is the importance of context: unless the secretary reconstitutes the text fairly soon after it was taken down, it may no longer be possible to reconstitute it completely accurately, since the outlines merely serve to prompt the memory of what was said.

6.2 Word Divisions

The sample of shorthand and its transcription below shows words bound together with hyphens that orthographically should be separate.
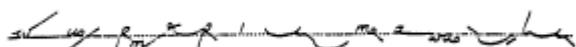

When using PitmanScript the writer must always try to-write neatly and-at-the same-time quickly, This may not seem very easy

That brings us to one of the major problems of any sort of linguistic analysis, including spelling reform, which is how to define the orthographic word. It is another area where the statistical behaviour of language is important. Consider the words *firewood, fire-engine, fire insurance*, which most people would probably spell respectively as one word, hyphenated and as two words. But there are very many expressions in English where there is no such consensus, and where different people will divide words in different ways. Then there are words like nevertheless, consisting visibly of elements that have the status of words in their own right. That is a problem both for spelling systems and for Information Theory. The next sample shows some shorthand outlines from the

field of politics and government which cover quite long chunks of material in which words are glued together.

Department of the Environment	Houses of Parliament
First Lord of the Treasury	in the House of Commons
Foreign and Commonwealth Office	in the House of Lords
freedom of the press	Labour Party
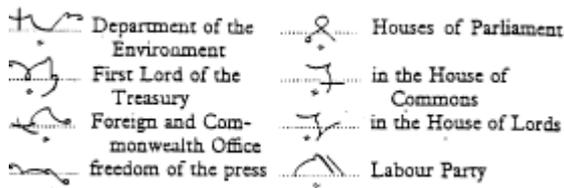
Words are of course normally separated from each other by a single space, but there is a kind of syntactic glue which crosses that space — between *fire* and *insurance* it is really quite strong, and has to be taken account of in Information Theory.

6.3 Syntactic Structures
This question is bound up with the nature of any given language and which of three methods it uses for structuring itself on the syntactic level. One method is inflection, tampering with word-endings usually, but sometimes with word-beginnings, and occasionally with word-middles. Another method is function words, like prepositions and so on in English. The third method is word order. All languages have to use a combination of these three methods to create syntactic meaning. Inflection of course doesn't figure very largely in English, which is particularly awkward in this respect because English words, so to speak, wear civilian clothes, rather than military uniforms with badges on them which say 'I am a noun', 'I am a verb', 'I am an adverb'. They are very gregarious and you don't know what they are until you have seen the company they keep. This has considerable implications for Information Theory, because what people often do when they are putting language into a restricted communication system as Shannon did is to use what we commonly refer to as telegraphic style, dropping words like articles, and carrying out other manipulations. Thus newspaper headlines are totally different from the ordinary grammar of English. The orthographic word problem is with us everywhere, and we cannot avoid it.

6.4 Non-Phonographic Shorthands
There are systems, some of them 2–300 years old, which try to contract running text to very short segments indeed that bring in a high level of conventionality. In Dutton Speedwords, for instance, the sentence *'there are hundreds of things to be done between now and the winter season'* appears as '*e cen d om e fad in nu & l pea peg'*, which has no   pronunciation equivalent indeed that is not its purpose; but it achieves a 62% reduction. That is however quite unlike the purpose of spelling reform, and indeed such systems do not seem to have been very successful over the last couple of centuries or so.

7 OMISSION

7.1 10% Letters Omitted
Let us now look at some texts illustrating the effect of varying percentages of omitted letters. The following text has a random 10% of its letters cut out:
D RING T E WAR THE APPLICATION OF RADA W RE PREDOMINANT Y OF MILTARY CHARACTER, UT SIN E HAT TIME ANY CIVIL A PLJCATIONS HAVE BE N FOUND. I WAS P OBABLY NAT RAL HAT THE USE F RADAR AS A AID 0 CIVIL A IATION AN MA INE AVIGATION S OULD FOLLOW Q ICKLY AF ER THE CE SATION OF HOS ILITIES AND HIS AS IND ED THE ASE. IT HAS FOUND A MA OR APPLICATION N THE F ELD OF IR TR FFIC CON ROL FOR INSTAN E.

It is possible to read this text with little difficulty, but it will be noticed that reading would be easier if the gaps within words had been closed. Cryptologists, after all, look especially for word endings

when deciphering text, but when composing coded text they try to baffle the reader by concealing the word boundaries and bunching their messages up into groups of five characters. If you know where the word beginnings are, the character combination TH is so common in English that it often gives the game away.

7.2 25% Letters Omitted
Now for a text in which a random 25% of the letters have been cut:

SINE HE GE 0 FIF E N POET Y AS BE N M R LING PA SION AN IONALL E EVE IN EN IONALL NDERTA EN AN ASK OR O MED A Y R LA IO SHIPHAT AS SEEM INCON IS ENT IT POET C RINC PLE , W ICH HA S METI ES ON M TH RE TATIO OF AN E CENT IC. A HE AG F SIXTY- I E I A TILL MUSED A T E PA ADO OF OETRY' OB TIN TE ONTI UA CE N TH PR S NT HAS OF CIVIL AT ON.

The full text reads:
Since the age of fifteen poetry has been my ruling passion and I have never intentionally undertaken any task or formed any relationship that has been inconsistent with poetic principles, which has sometimes won me the reputation of an eccentric. At the age of sixty-five I am still amused at the paradox of poetry's obstinate continuance into the present phase of civilization.

It is characteristic of the reading-process that on now looking back at the reduced text, patterns become clear that were not so before. One letter in four was missing, but with some persistence it was possible to read the text. If two letters out of five go, then the task becomes as arduous as a crossword puzzle — it might take an hour to solve. In fact crosswords are interesting from the point of view of Information Theory because they contain a lot of contextuality, and they wouldn't be possible if language didn't have this inbuilt redundancy.

7.3 Systematic Omission
A question that now arises is whether it makes any difference if letters are systematically rather than randomly omitted. We find for instance that if the word beginnings get obliterated, reading becomes very much harder, but it makes much less difference if the letters omitted are not pronounced anyway, or if only vowel-letters are omitted. Consider how little difficulty the large number of omissions in the following sentences make to our comprehension.

1 Tk the bk to the grl in the clsrm.

2 I shl go hm at fv.

3 Pt the bx on the tp shlf in the rm.

4 I wnt into the rm to se the grl.

5 I tk the bs to the vlg to mt my fthr.

But there is a paradox here regarding spoken language, because just as you can delete vowels (but not consonants) from a written text without rendering it incomprehensible, so you can remove consonants (but not vowels) from recorded spoken language, and it doesn't have too serious an effect. Technically the reason is that vowels have three phases, an onset, a hold and an offset phase, and there are enough acoustic clues in the onsets and offsets to give an almost cast-iron guarantee as to what the adjacent consonants would have been.

7.4 Hebrew
This brings us to Hebrew. Written Hebrew has some similarity to that form of English used in the five sentences above. The consonants remain, indeed a consonantal skeleton has to remain, but they are insufficient to reconstitute the pronunciation. To do that, certain other symbols have to be added. A speaker of Hebrew would be able to read off the sentence

<div dir="rtl">

הזיקה
שבין השלטון המרכזי
לשלטונות המקומיים
יש בה ניגוד נוסף.
</div>

(from right to left)

which transliterated gives (reading from left to right):

hzyqh   <u>sh</u>byn h<u>sh</u>lTvn hmrkzy   l<u>sh</u>lTvnvt hmqvmyym   y<u>sh</u> bh nygvd nvsp

while phonemically it produces (reading from left to right with literal translation beneath each line):

/hazika
the-relationship

ʃebeyn	haʃilton	hamerkazi
which-between	*the-authority*	*the-central*
laʃiltonot	hamekomiin	
to-the-authorities	*the-local*	

yeʃ	ba	nigud	nosaf/
there-is.	*in-it*	*contrast*	*additiona.l*

This sentence, *'the relationship between the central and local authorities constitutes another contradiction'*, is actually a quotation from Mao Tse-tung. The operation of reading in Hebrew is rather different from that in English. An English text leads the reader to make a phonetic projection in the mind which yields a semantic interpretation; but in Hebrew this is not possible. The text provides only the consonantal skeleton, and the reader has to supply the vowels; but there are many ambiguities, and reading straight off is not an easy exercise. A semantic interpretation first has to be struggled for, which then contextualizes to give a unique phonetic representation. There is an orthographical reform lobby for modern Hebrew in Israel, which would like to make things easier for the reader by inserting characters which will give a more secure prompt as to which word is really intended. It should be remembered though that while the above sentence is the normal way of writing Hebrew, sacred texts and sometimes poetry have special status and can be written in what is called vocalized form, which includes extra dots above and below the letters, called the Nikud system. There is a Hebrew word which means a teacher, and is pronounced /moreh/; in Hebrew orthography סורה but the written form can also mean a female teacher, and then it is pronounced /mora/. There is no difference in the spelling. The syntax or context may give a clue, but if you wanted to write down *'I like my teacher'* in Hebrew, a reader would not know how to pronounce it without knowing whether a male or female teacher was meant. One way round this is to add extra information as to the sex at the end of the sentence. But the Hebrew spelling system is an excellent example of how redundancy can be dispensed with, even though it antedates Shannon's formulation of redundancy by a couple of thousand years.

7.5 Arabic & Abbreviations

Arabic, being also a semitic language, is rather similar. If you ask an Arab to read this word (right to left) تعلم he will be unable to. It could be /ta'lamu/, meaning *you know*, or /ta'allama/ (*he studied*), or /ta'allam/ (study!, imperative), or /ta'allum[un]/ (*study, instruction*), or /tu'allimu/ (*you are teaching*), or /tu'limu/ (*you are announcing, informing*). It is only when you have the context that you can resolve the ambiguity of these abbreviated forms. There is nothing unusual about abbreviations, of course; we use them all the time. Why bother saying *Union of Soviet Socialist Republics* when you could say just *USSR*. In small ads, in an attempt to reduce the cost of the ad, VGC can stand for 'very good condition'. But all abbreviations are conventional, and to understand them you have to know what the conventions are.

8 ORDERS OF APPROXIMATION

8.1 Zero Order Approximation
Another statistical aspect of information theory is illustrated by the following sequence of letters:

ZEWRTZYNDADXESYJRQY-WGECIJJ-CBVKRBOPQZB-YMBUABWLBTQCNIKFMP-MVUUGBSAHXLHISIE -M...

This is a random extraction of letters from the English alphabet with absolutely no structure. Each letter had an equal chance of being extracted, so it represents what in Information Theory is called zero order approximation to the structure of English words.

8.2 First Order Approximation
If on the other hand we choose letters in the frequency with which they would normally occur in text, the result is different. For every 1,000 letters used in English text, about 150 Es occur, but only about 3 or 4 Qs and Zs. If you take these statistics into account, you get a first order approximation, as here:

AI-NGAE-ITF-NNR-ASAEV-OIE-BAINTHA- HYR-OO-POER-SETRYGAIETRWCO-EHDUARU-EU-C-F-T-NSREM-OIY-EESE -F-O-SRIS-R-UNNASHOR….

This is still not very like English.

8.3 Second Order Approximation
However if you then take the frequency with which any letter occurs after any given preceding letter, we get a second order approximation, as here:

URTESHETHING-AD-E-AT-FOULE-ITHALIORT-W-ACT-C-STE-MINTSAN-OLINS-TWID-OULY-TE-T-HIGHE-CO-YS-TH-HR-UPAVIDE-PAD-CTAVED....

Just occasionally quite plausible strings of English-sounding syllables appear here: *THING, ACT* as a word, *TH* together, *LY* at the end of a word, *GH* as in high, *ED* as the end of a sort of verb, and so on. This arises from the frequency with which particular pairs of letters occur. By taking letters in pairs, this order of approximation gives you a set of pairs comprising 26 x 26.

8.4 Third Order Approximations
The step after that gives a set of 26 x 26 x 26, from A to Z in triads, which looks much more plausible as English.

IANKS-CAN-OU-ANG-RLER-THATTED-OF-TO-S-H-OR-OF-TO-HAVEMEM-A-I-MAND-AND-PUT-WHISS-ITABLY-THERVEREER-FIGHTS-TAKILLIS-TA….

It doesn't mean anything, but if that had been a garbled telegram, one would have had no doubt that it was English that had been thus garbled. Although there is no intelligence in it, it does show the great stability of the statistics of language. If you then apply third-order statistics to French, the result indeed looks Re French:

JOU-MOUETAS-DE-MONNERNAISSAINS-BEME-US-VREH-PRE-TU-DE-TOUCHEUR-DIMMERE-LIES-MAR-FLAME-RE-Á-VER-IL-DOUVENTS-SOC....

German looks as follows:
BET-F-REINER-SOMMEIT-SINACH-OSST-BETT-O-ER-ZUM-WIE-BEST-ALLIENDER-TAUSICHELLE- LAUFURCHT-ER-BLEINDESE1T-UEBER-KONN….

Spanish:
RAMA-DE-LLA-EL-GUIA-IMO-SUS-CONDIAS- S-U-E,-UNCONDADADO-DEA-MARE-TO-BUERBALI-A-NUE-Y-NERARSIN-DE-SE-SUS-SUPAROCEDA....

Latin:
ET-LIGER-CUM-SITECI-LIBEMUS-ACELEREN-TE-VICAESCERUM-PE-NON-SUM-MINUS-UTERNE-UT-IN-ARION-POPOMIN-SE-INQUENEQUE-IRA....

Russian:
лоя-б-и-ечлечним-маш-фунсткий-иг-
даль-стой-предатягнащёки-этузко-
быложут-часы-орглиц-пъю-вленает…

8.5 Fifth Order Approximation
The final example is based on the frequency of words rather than of letters, and is derived rather like a guessing game. If you start with the word *the*, you then look for the first word that follows *the* in whatever long text you choose as your source; if the word you find after *the* happens to be *head*, you then look for whatever word follows the next occurrence of *head*, which happens to be *and*, and so on. Again, it doesn't mean anything, but it is indicative of a sort of structure, indeed that is the only point to this fifth order approximation to English syntax.

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED....

9 LETTER REDUNDANCY

9.1 Left to Right Branching in Spelling
Statisticians and computer scientists know that creating say an eighth order approximation to the behaviour of letters of the alphabet is a massive computing job, so guessing games are often used instead: subjects are presented with a certain letter and asked which letter they think should follow. If we start with the letters CONF, for instance, and you have to bet which letter would occur next, you might say E if you were thinking of *conference* or U if you were thinking of *confusion*. But let us enter the letter I (CONFI), which might suggest *confidence* or *confiscate*. This choice between several alternatives can be seen as a branching effect as we proceed from left to right through the word. If we add D (CONFID), then an A or an E could follow. E could be the end of the word, in which case a space would follow. Alternatively, one could add N, then T followed by either space or TAL. That too could be the end of the word, or you could add IT+Y to produce CONFIDENTIALITY, or IT+IES to produce CONFIDENTIALITIES.

9.2 No Choice = Redundant Information
This method can be used in the form of a very fine algorithm by the unintelligent computer to split orthographic words up automatically into their constituent morphemes, according to the statistics that govern each threshold, and it shows Information Theory can be applied not just to letters, but to fragments and segments of words. It is all very well to say T occurs so many times in the middle of words, but from the orthographical point of view we want to know exactly where, in what environments, since this is relevant to pronunciation and the tension that exists between speech

and script. Shannon would conclude from this that redundancy arises whenever there is no choice as to the next letter in a word: where it is obligatory, it can be left out.

9.3 Cutting Spelling

With these points in mind, let us consider three versions of a text. The first is in standard English but capitalized and using the dot as the separator between words:

TASS.A'I'I'ACKED.INDEPENDENT.PEACE.ACTIVISTS.
IN.THE.SOVIET.UNION.YESTERDAY.AS.
ANTISOVIET.RENEGADES.CRIMINALS.AND.
IMPOSTERS.IT.SAID.THE.GROUP.FOR.THE.
ESTABLISHMENT.OF.TRUST.BETWEEN.THE.USSR.
AND.THE.USA.WAS.MANIPULATED.BY.
WESTERN. INTELLIGENCE.SERVICES.

The same text in Cut Spelling is obviously an improvement in Information Theory terms:

TASS.ATAKD.INDEPENDNT.PECE.ACTIVISTS
.IN.TH.SOVIET.UNION.YESTRDAY.AS.
ANTISOVIET.RENEGADES.CRIMINLS.AND.
IMPOSTRS.IT.SAID.TH.GRUP.FOR.TH.
ESTABLISHMNT.OF.TRUST.BETWEEN.TH.USSR.
AND.TH.USA.WAS.MANIPULATED.BY.
WESTRN.INTELIGENCE.SERVICES.

Then the same text with vowels removed as in Hebrew, but indicated by a mark (this certainly gives pause for thought as to how meaning is preserved in such a script):

T_SS._TT_CK_D._ND_P_ND_NT.P_C_._CT_V_STS.
_N.TH_.S_V_T._N_N.Y_ST_RD_Y._S.
_NT_S_V_T.R_N_G_D_S.CR_M_N_LS._ND.
_MP_ST_RS._T.S.__D.TH_.GR_P._F_R.TH_.
_ST_BL_SHM_NT._F.TR_ST._B_TW_N.TH_._SSR.
_ND.TH _._S_._W_S.M_N_P_L_T_D.BY.W_ST_RN.
_NT_LL_G_NC_._S_RV_C_S.

10 CALCULATING REDUNDANCY

10.1 Binary Analysis

Shannon basically approaches all Information Theory questions by requiring a yes-or-no answer. This binary approach, which modern school children are familiar with through their new maths, is the calculus that underlies his computation of redundancy figures. The basic principle he uses is that if something is highly probable, then its informativity is low, and if it is improbable, then its informativity is high. It took him some time to find the right sort of mathematical treatment for this, but he did it in the end, as I think is acknowledged by all workers in the field.

10.2 Single Letter Redundancy

The table below gives the letters of the alphabet in frequency order as they occur in an extended version of the first text in 9.3 above, with doT representing space, as the most common, and some calculations carried out with binary logarithms on these probabilities, summed up in the right hand column to yield a value which is explained underneath. Shannon says that ideally all symbols used should be equiprobable, in which case, any deviation from equiprobability is some sort of

redundancy. On that basis we can show that the source text contains almost 14% redundancy. Because some letters are used more than others,   they don't carry an equal burden.

No.	Let		Prob	-LB(Prob)	Term	Sum
1	:	.	- 0.16022	2.64190	0.42328	0.42328
2	:	E	- 0.11881	3.07333	0.36513	0.78841
3	:	T	- 0.08215	3.60568	0.29619	1.08460
4	:	A	- 0.06517	3.93958	0.25675	1.34135
5	:	S	- 0.06517	3.93958	0.25675	1.59810
6	:	N	- 0.06449	3.95469	0.25505	1.85316
7	:	I	- 0.06042	4.04881	0.24463	2.09779
8	:	R	- 0.05703	4.13222	0.23565	2.33344
9	:	O	- 0.05227	4.25776	0.22257	2.55601
10	:	D	- 0.03870	4.69165	0.18155	2.73756
11	:	H	- 0.03055	5.03269	0.15375	2.89131
12	:	C	- 0.02512	5.31509	0.13351	3.02482
13	:	U	- 0.02308	5.43708	0.12550	3.15032
14	:	M	- 0.02105	5.57035	0.11723	3.26755
15	:	L	- 0.02105	5.57035	0.11723	3.38478
16	:	P	- 0.02037	5.61765	0.11441	3.49919
17	:	V	- 0.01697	5.88069	0.09981	3.59900
18	:	Y	- 0.01494	6.06511	0.09059	3.68958
19	:	G	- 0.01290	6.27662	0.08096	3.77054
20	:	B	- 0.01222	6.35462	0.07765	3.77054
21	:	W	- 0.01222	6.35462	0.07765	3.84820
22	:	F	- 0.01154	6.43708	0.07429	4.00014
23	:	K	- 0.00611	7.35462	0.04494	4.04508
24	:	Z	- 0.00272	8.52454	0.02315	4.06823
25	:	J	- 0.00204	8.93958	0.01821	4.08643
26	:	Q	- 0.00136	9.52454	0.01293	4.09937
27	:	X	- 0.00136	9.52454	0.01293	4.11230

Maximum possible entropy for a set of 27 symbols: 4.75489 bits
Actual entropy for this source: 4.11230 bits. Relative entropy for this text-source: 86.48570%
Redundancy for this text-source: 13.51431%

10.3 Redundancy of Letter-Pairs
If you then apply the same technique to letter-pairs, redundancy goes up to nearly 24%, which is because the letter-pairs are not all equally exploited. Many are very familiar, such as TH, AN, IN. The 55 most common in the full version of the text with their frequencies, were:

1	E.	41	20	NT	17	38	.W	11	**11**	AN	24	30	P	12	48	O.	8
2	.A	38	21	ST	17	39	OV	11	**12**	ND	23	31	DE	12	49	OF	8
3	.T	35	22	Y.	16	40	AS	10	**13**	ER	22	32	AC	12	50	NO	8
4	S.	33	23	VI	16	41	IS	10	**14**	N.	20	33	OR	12	51	EA	8
5	TH	30	24	M	15	42	RS	9	**15**	S	20	34	ON	11	52	F.	8
6	D.	30	25	.O	14	43	.B	9	**16**	RE	19	35	AR	11	53	ME	8
7	T.	29	26	.I	14	44	.H	9	**17**	ES	18	36	SE	11	54	OU	8
8	HE	27	27	ED	14	45	CE	9	**18**	TE	17	37	ET	11	55	LI	8
9	IN	26	28	TI	14	46	TA	8	**19**	R	17					
10	EN	24	29	IE	13	47	WE	8	**11**	AN	24	30	P	12	48	O.	8

With letters in threes, or triads, redundancy exceeds 40% for this text. Many letter-combinations, like THQ, of course never occur, and that represents a sort of systemic constraint.

10.4 Bits of Information
Taking that approach, you can say that the standard orthography transmitted a certain number of 'bits' of information, 'bit' being a technical term in Information Theory and computer science, a stump word from 'binary digit'. Now a text in Cut Spelling contains the same linguistic information, because the message is the same, but it saves a number of bits of information in its transmission, and that number can be used as an index of improved efficiency. Comparative figures for the three versions of the full text are as follows:

Text l (t.o. with dots for spaces): 1473 symbols, information transmitted, 380.26 bits.
Text 2 (Cut Spelling) 1352 symbols, information transmitted, 345.59 bits.
Text 3 (Vowels deleted): 1002 symbols, information transmitted, 249.22 bits.
Ratios: Text 2 : Text 1 = 0.91; Text 3 : Text 2 = 0.72; Text 3 : Text 1 = 0.66.

But there are some problems there that I can't give a satisfactory explanation of, even as someone who has thought about these things for a long long time. All the pieces of the jigsaw don't easily fall into place.

11 ENCODING INFORMATION

11.1 Huffman Coding
One ingenious method of handling Information Theory concepts came from Huffman, who said quite correctly that one has to accept as a fact of life that the symbols are used with different frequencies, as Samuel Morse knew when he developed the morse code. Morse went into a printer's workshop and saw the compositor's trays full of slugs: he noticed the container for Es was massive, but only a few Zs were needed. On this basis he decided that the code elements for the common letters needed to be short, while those for the less frequent letters could be longer. Below is a Huffman coding for the English alphabet in descending order of letter-frequency (read down the columns, left to right); the less frequent the letter, the longer its coding:

.	111	D	11001	G	001001
E	011	H	01011	B	000111
T	1101	C	0010	W	001000
A	1010	U	00010	F	000110
S	1011	M	00001	K	0101010
N	1001	L	00000	Z	01010110
I	1000	P	110001	J	010101110
R	0100	V	110000	Q	0101011111
O	0011	Y	010101	Z	0101011110

11.2 Equal Length Coding
However this system gives rise to a serious problem for computer scientists, because they don't like codes of varying length. An experiment to reverse this was carried out by Mike Lynch of Sheffield University, with whom I worked for 10 years. He devised 256 codes, each with 8 digits, to represent all the letters of the alphabet, numbers from 0 to 9, and punctuation marks, as well as a large number of common letter-pairs, triads, foursomes and even a few combinations of short words. Here is a small selection of the codes:

*	00000001	6	00010001
?	00010110	A	00010111
Z	00110000	'S	00110010

AN	00111100	IS	01100100
NG	01111000	Q	10000111
CON	10110110	ING	10111110
TH	11011001	ATIO	11100001
FROM	11110100	WILL	11111000
AND	11111010	TO THE	11111111

To convert our text into this system, we enclose in brackets whatever character or string of characters can be encoded by a single code-group, as shown here for the opening of the text:

( )(TA)(SS)(AT)(TA)(CK)(ED)(IN)(DE)(PE)(ND)(EN)(T)
(PE)(AC)(E)(AC)(TI)(VI)(ST) (S)(IN THE)
(SO)(VI)(E)(T)(UN)(IO)(N)(Y)(ES)TER)(DA)(Y)

The brackets split up the text so that it can be stored in the computer as a continuous sequence of ones and zeros:

000000001001100111010101001111110011001010
001111011011101100010010011001000010001101
10010100001001100010000100001101100100111100
0110110100111001010011010010110000110111001
10001010000101100011010010111010100

This kind of coding enables text then only to take up half the space inside computers that would be required if they were put in letter by letter. It is a very impressive piece of applied Information Theory. However it is now unnecessary, because computer memory, once so expensive, is now very cheap.

11.3 Braille

The braille alphabet, which is a series of dots or non-dots on paper, is likewise a kind of binary code. It has some close analogies to Information Theory, even though it pre-dates it by a century; for instance it comes in both a contracted and a non-contracted form. To save the blind reader the effort of scanning too many cells with the finger tips, contracted braille uses a number of conventional dot configurations, which have to be learnt. The text is thereby enormously reduced, and the reading process speeded up. It might be of interest to the Simplified Spelling Society to examine what sort of stump spellings and conventions the RNIB and its equivalent in other countries use in their systems.

Here is a selection of the braille dot configurations:
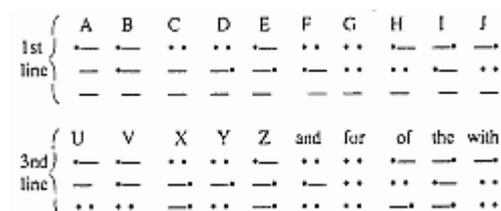
12 COMPUTERS AND SPELLING
12.1 Trends towards Text-Processing

Already today computers are processing massive amounts of text — in fact that is the real growth area with computers. At the moment the ratio of numerical to textual material processed is 7:3, but by 1990 that ratio is expected to be reversed, in addition to the absolute growth that will occur

anyway. A lot of time and money is being invested in intricate software for spelling correction, and one has to say that it is a shame such a thing is felt to be necessary.

12.2 Algorithms and Dictionary Look-up

Computers like a diet of algorithms, which one could compare with the recipe for baking a cake: if you take a certain sequence of steps, correctly and in the right order, the desired result will emerge. Computers are highly adept at carrying out such sequences of steps, at high speed and repetitively. But with our present spelling, algorithms have to be supplemented with checklists, a dictionary look-up procedure to deal with all the exceptions. An example from outside spelling: in order to deal with a word like went in text you don't use an algorithm to relate it to go, you use a dictionary look-up procedure. But looking up individual words is a very untidy business as far as computing is concerned. However, one shouldn't worry too much about the effect of spelling irregularities on computing, because what the computer can't compute, it can look up, and vice versa.

13 SOME QUOTABLE STATISTICS
13.1 Savings by Cutting Redundant Letters

Let me finish with some quotable statistics. Before the Russian Revolution, the Russian orthographical system included a lot of redundant letters, the most redundant being called the hard sign. One feature of the Russian language is that consonants can be palatalized or non-palatalized, and the distinction used to be marked by either a soft sign or a hard sign. The hard sign was used in particular phonetic contexts at the end of words wherever there wasn't a soft sign, and it was thus clearly redundant. Lenin himself largely got rid of it in the orthographical reform that took place shortly after the October Revolution, and as a result Anna Karenina became 35 pages shorter. The hard sign still exists but is restricted to some very unusual situations, where it acts as a separator to prevent a consonant being contaminated by a following vowel, where that is required.

13.2 Simplifying Doubled Consonants

Another example, also from Russian, is the word for communist, which is of course very common. In Russian it is spelt with MM although only a single M is pronounced; but in other slavonic languages, such as Polish (komunistyczny), the word is written with only one M. If one of the Ms were dropped in Russian, 2.35 tonnes of printing ink would be saved every year in the USSR.

14 RELEVANCE OF INFORMATION THEORY TO SPELLING REFORM

Although the finer technical details of Information Theory are no doubt not central to spelling reform as such, nevertheless as a whole it is not entirely on the periphery, and those who are concerned with designing improved orthographies should perhaps have a general awareness of its implications. In Information Theory the concept of redundancy is chiefly applied to the achievement of maximum efficiency for machines; but in terms of the psychology of reading, it is clearly of the greatest importance to consider the question of efficiency of text-processing by the human brain, which, though it operates in a different way from machines and has different needs, nevertheless should ideally also be enabled to perform the functions of literacy as quickly, as easily and as accurately as possible.

FURTHER READING
J. Campbell, *Grammatical Man*, Penguin, 1984
C.Cherry, *On Human Communication*, MIT, 1968
J. R. Pierce, *Symbols, Signals and Noise*, Harper, 1961
J. Singh, *Great Ideas in Information Theory, Language and Cybernetics*, Dover, 1966
Braille Primer, RNIB, 1966.

4. Morphological Considerations in the Creation of Rational Orthographies
Edgar A. Gregersen

[Professor Gregersen teaches and researches in the Department of Anthropology at Queens College and Graduate Center in the City University of New York. He has been writing on aspects of orthography since 1977. This article is an edited version of the address he gave to the Society's **Fourth International Conference in July 1985**].

"English orthography, despite its often cited inconsistencies, comes remarkably close to being an optimal orthographic   system for English." (*The Sound Pattern of English*, 1968, p.49)

This assertion by Noam Chomsky (perhaps the most influential linguist of the 20th century) and Morris Halle (another linguistic giant) is probably the most serious intellectual blow the spelling reform movement has yet suffered.

It doesn't matter that many linguists, especially in Europe, are no longer particularly entranced by Chomskyan theory. The reality is that Chomsky's prestige in intellectual circles in the United States alone suffices to give his views tremendous clout. This is especially true because all students on the college or university level are exposed to his ideas if they take courses in linguistics. For example, in a widely used college textbook for elementary linguistics courses, An Introduction to Language by Victoria Fromkin and Robert Rodman (1983, 3rd ed.), the question of spelling reform for English is considered — and largely rejected, using Chomskyan arguments (see pp.156–159). Similarly, in an essay on English orthography in *The English Heritage Dictionary of the English Language* (much beloved by American English teachers), Wayne O'Neill makes the same argument.

Spelling reformers must take up the challenge of Chomsky's critique and see how they can use it to best advantage.

First however, let us consider the relevant points in Chomsky's position. His major point is that variations in the related forms of a word (or, more technically, the allomorphs of a morpheme) should not be indicated if predictable. For example, stress is very often predictable in English words and when it is, it should not be shown. Compare *phOtograph-photOgrapher-photogrAphic* or *hIstory-histOrical-historIcity* with a totally regular pattern of stress on the third syllable from the end. However, the morpheme {histori} consequently has three different allomorphs: /hIstori, histOri, historI/ (capital letters indicate stress; other variations are not shown).

Similarly with vowel reduction. The vowel /o/ in the allomorphs of {histori} often is reduced to shwa when not stressed, thus /hIstəri/ or /hIlstri/, and /histərI/ as in *historicity*.

The single form from which all the variations can be derived is of course hypothetical. But Chomsky sets it up as an underlying representation and he thinks that this is what should be the basis for conventional orthographies.

He notes also that in different dialects the underlying representations will be the same even tho the actual forms used (the surface representations) may be quite different.

The implications for spelling reform in English are the following.

(1) Related words that can easily be shown as related should not be broken up. The most egregious example of breaking up a morpheme unwarrantedly is the 1971 proposal adopted by the Simplified Spelling Society (SSS) that *your* and *yours* be written *yoor* and *yoors* but *you* should be written *U*. The older 1948 forms in classic *Nue Speling* were superior (tho not wonderful): *U, uer, uerz*. I myself prefer something on the order of *yu, yuur, yuurz*. The form *yu* is in fact currently sponsored by the Simplified Spelling Society of Canada and was suggested by the now defunct Simplified Spelling Board (an American group).

(2) Inflections should as much as possible be written without variation. A notable example is the regular English plural, which in traditional orthography is written as <-s>. In *Nue Speling* (of whatever vintage) it is broken up into <-s> and <-z>, as in *kats, dogz*. The recent proposal (1985) put forth by the SSS's working party that the plural always be written as <-z> is a big improvement and should be adopted immediately. (Interestingly enough, in the book by Fromkin and Rodman mentioned earlier, the authors assert:

"It is doubtful that anyone would suggest that the plural morpheme should be spelled *S* in cats and *Z* in dogs. The sound of the morpheme is determined by rules..." (P. 158).

The authors have clearly not done their homework since virtually all spelling reformers have hitherto broken up the plural morpheme. As a matter of fact, I know of no exceptions apart from the recent SSS proposal.)

(3) Vowels in unstressed syllables must reflect the stressed vowels in related words, e.g. *symbol-symbolic, substance-substantial*. This is precisely what is done (for the most part) in *Nue Speling* (*simbol-simbolik, susbtans-substanshal*) and should be retained, perhaps even being made more consistent (cf. the classic 1948 *Nue Speling* forms *leprus* but *leprosy*; also *kalus, jenerus* but *kalosity, jenerosity*.

(4) Regular vowel alternations in stressed syllables reflecting the Great Vowel Shift of Early Modem English have been called by Chomsky and Halle 'without doubt the pivotal process of Modern English phonology' (*The Sound Pattern of English* p.187). These suggestions are exemplified in the following often quoted examples (the phonemic values are given in the traditional Daniel Jones transcription for RP):

/ei/	- /ae/	sane	- sanity
/i: /	- /e/	serene	- serenity
/ai/	- /i/	divine	- divinity
/ou/	- / /	verbose	- verbosity
Compare	also:		
/(j)u:/	- /Λ/	presume	- presumption

Some proposals for the reform of English spelling would obscure these alternations by using the 'international' values of vowels, e.g. *sein (*for *sane), devain (*for *devine), verbous (*for *verbose).* It is really these proposals that Chomsky's main displeasure is directed against.

However, the traditional *Nue Speling* conventions in fact preserve the relationships in the most practical way possible (other than by introducing some diacritic), viz, by indicating the so-called 'long' value of vowels by vowel plus <-e>: *saen, divien, verboes.* The actual design of *New Speling* conventions thus accommodates the regular vowel alternations quite adequately and is a decided virtue.

The recent proposal by the SSS working party to substitute <y> for <ie> would break up the regular alternations. This proposed use of Y should be abandoned.

No orthography can be totally morphemic if only because scholars disagree on what that would entail. Even tho traditional English orthography often does ignore differences between allomorphs, it is not entirely consistent; cf. *five* but *fifty* and *fifth* (one could write *fivty and fivth* but no one to my knowledge has suggested doing so). It is therefore clear that if major alternations such as the Vowel Shift are dealt with, other details can be safely ignored. *Nue Speling* does this admirably.

(5) Stress is for the most part predictable, according to Chomsky and Halle. But the rule they offer is so complicated that it has little relevance for a practical orthography. The 1968 version of their rule is known also to be inadequate. However, we can abstract from it a general rule that will cover most cases and then mark the exceptions, say, with a <'> over the stressed vowel.

Altho most spelling reformers in English have ignored the question of stress — largely, I suppose, because it would introduce a diacritic — nevertheless, I believe irregular stress should be shown. Stress is probably the single most unstable element in English pronunciation and traditional stress patterns have been lost in the last few generations or are changing significantly at present simply because of a failure of the standard orthography to indicate it (cf. the pronunciations *of éxquísite, hóspítable, lábóratory,* and at least in the United States *háráss, áfflúent,* etc., where the diacritics indicate alternate stresses). In Spanish or Greek, where stress is consistently marked, no comparable variation *occurs.*

The rule *I* suggest for regular stress is an abbreviated version of the Chomky-Halle rule:

Stress occurs as far from the end of a word as possible, up to the third syllable (the ante-penult), unless one of the syllables contains a 'long' vowel; then, the first long vowel from the end is stressed.

Hence, words like *history-historical-historicity, general, optimal, element, spelling, alive, linguist, obtain* would receive no stress mark because their stress is regular. But *phonétics, segméntal, succéss, éxercise* would be marked. Monosyllabic words would normally not be marked (except perhaps to make differences between homonyms, e.g. w*ún 'one' vs. wun 'won').*

(6) Altho it would be desirable for an orthography to be devised so that a speaker of the language would find a general one-to-one correlation between his own pronunciation and the spelling, for various reasons this is seldom completely practical. However, a speaker should be able to deduce an appropriate pronunciation for his own dialect from the spelling. In short a spelling may show more differences than are phonemic for any one dialect. It should never show fewer.

The recent proposals by the SSS to reduce phonemic contrasts that are almost universally observed (e.g. between *cam* and *calm, fool* and *full)* or very often kept *(whine-wine)* should be abandoned. Furthermore, the universally observed distinction between the initial consonants in *thy-thigh* must be shown as in classic *Nue Speling*, with <dh> vs. <th>.

In point of fact some other distinctions should be introduced into *Nue Speling* which have not been observed before.

One of these is the distinction generally maintained thruout Scotland in the traditional form of Standard English spoken there between the vowels of *fir-fern-fur.* Altho some Scots apparently have adopted the RP merger of all these vowels into one, most speakers seem to have preserved

the distinction. Furthermore, if these distinctions are not recognized, Scots nationalists might well reject *Nue Speling* altogether as a form of English imperialism. So for both linguistic and nationalistic reasons, these distinctions must be recognized. By the way, they are observed in Sir James Pitman's i.t.a. and seem to cause little trouble for students who do not make them**.** The OED recognizes two distinctions, lumping <ir> and <er> together. In ***[Braeking dhe spel: an apeel to komon sens](#)*** (an SSS publication from 1942), a two-way distinction is at least tolerated (see p.39).

A much more complicated issue involves the vowel of the final syllable of *history, every,* etc. Spelling reformers and orthoëpists have almost always regarded this vowel as short and as phonemically the same vowel as in pit. This was true even in the United States where dictionary makers always indicated the pronunciation of such Ys as a short I. That is, until 1961 *Webster's Third New International Dictionary* was published, which listed as the most common pronunciation one equivalent to the <ee> in *see* (in Jonesian transcription /i:/, in American tradition /iy/ or IPA /ij/).

Since that time American dictionaries recognize the <ee> pronunciation as the preferred one.

The linguistic reality is in fact reflected in *Webster's Third,* and it is shocking that it took so long for dictionaries to come to terms with it. It is also ironic that when the British Simplified Spelling Society and the American Simplified Spelling Association decided in 1956 to compromise and reach unanimity in their decisions, the Simplified Spelling Association insisted on showing the <-y> ending and derived forms always with a short <-i>.

This dialect difference between RP and perhaps most dialects of English outside of North America, and dialects of the General American type poses more of a difficulty than is usually understood. In General American the words *candid* and *candied* are different. (A pronunciation of <-y> and related forms with /i/ also occurs in America but tends to be very regional and has little prestige; it is used by some Rock singers in imitation of Black English dialects.) Following the general rule implicitly suggested earlier that when dialects differ, the maximally differentiated dialect should serve as the basis for the written norm, we might assume that the American pronunciations should be taken as the norm: we would thus get *kandid 'candid', kándeed* 'candied', and *kandee*, 'candy'.

But the short /i/ speakers make another difference: whereas the <ee> speakers *never* have a contrast between unstressed <ee> and <i> finally, the short /i/ speakers have. In RP, for example, where unstressed <ee> is unusual and /i/ tends to be ousting it (e.g. words from Italian with a written plural in **<-i>** are now pronounced either with <ee> or /i/, with /i/ forms perhaps now more common: *libretti, banditti),* nevertheless an <ee>-/i/ distinction is maintained in final unstressed syllables. A pronunciation with <ee> is the norm in words derived from Latin written with <-ae>: *alumnae, amoebae, antennae, formulae, larvae, nebulae.* The singular *stele* ends with /i/, the plural *stelae* with /i:/. Many words recently borrowed from French and some other foreign languages also have final <-ee>: *causerie, chassis, debris, gaucherie, jalousie, lingerie, précis, spahi* (these words are sometimes pronounced quite differently in the United States: *debris* is usually stressed on the last syllable, and *lingerie* usually — to the consternation of purists — rimes with *day).* Note that whereas in American English *mammy* and *mammae* 'mammary glands' are pronounced identically (unless a restored Latin pronunciation is attempted so that the <-ae> is pronounced like the <ai> *in aisle),* in the rest of the English speaking world they are generally differentiated as *mami vs mámee.* The usual American pronunciation of the song made famous by Al Jolson containing the words 'my little mammy' would perhaps be misconstrued by an Englishman as at once a lament by a preoperative transsexual and a plea for more silicone.

In this instance an out-and-out compromise must be made. The problem is not final <-y> since even if written with <-i> speakers of General American could simply interpret it as a variant of <-

ee>. The problem occurs when the sound is not final, usually before an inflectional ending, where <i> would not do.

One solution has been suggested by Martin Joos: write what in General American is <ee> as <i> in unstressed syllables, and /i/ as <e>. This means *kanded* for 'candid' and *kandid* for 'candied'. The major drawback is that the ending <-ing> would have to be written as <-*eng>: singeng, etc.*

A better solution has been proposed by Edward Rondthaler who recommends that words with General American unstressed <ee> be written with Y both finally and elsewhere: using other *Nue Speling* conventions we would get k*andy* 'candy'*, kandyd* 'candied', *kandyz* 'candies'. This solution in fact simply extends present written usage: altho normally Y is changed to I before inflectional endings, it is not before the possessive <'s> nor with proper names, e.g., *Mary's, the two Germanys.* (Note also that the plural of fly 'carriage' can be *flys*; the spelling *flies* refers to insects.) All we have to do is drop the change-the-Y-to-I rule. The major drawback with this is that it gives two values to Y, one of which is not in accordance with international conventions.

Let me here suggest a third solution, which may be the most far out but which has several advantages from a linguistic point of view.

The main problem with writing <-i>, as I've said before, is before an inflectional ending. If we could overcome that problem we would have the best solution. In the first place, we would want to have the vowel count as short (no matter how it's actually pronounced) so that the stress rule given earlier would not treat *history, etc.,* as irregular (writing <i> would indicate that stress is indeed regular). Secondly, <i> would be in keeping with international conventions, whereas <y> would not, and in particular spellings like <-ny> (as in *money, honey, pony)* would suggest a pronunciation as in *canyon* or the Ñ in *mañana.*

We could accomplish our goal if we had a device that could indicate that the inflectional endings somehow didn't count as a full-fledged part of the word and were quasi-autonomous. In part the apostrophe used in possessives does this quite well: *lady's,* etc., which could quite adequately be written as laedi'z. Unfortunately we cannot use the apostrophe to indicate the plural as well (altho Sinclair Eustace was surely on an interesting track when he suggested that it be used in a whole range of phenomena he called junction). If we had yet another 'boundary marker' we would do quite nicely. The 1956 version of *Nue Speling* used a raised dot or full stop (period) to indicate the separateness of vowel sequences (for the earlier hyphen): *medi.eeval, poe.em, kwie.et. I* suggest that this be used with the plural and other inflectional endings. Thus, the plural of lady would be *laedi.z. [*In *Webster's Dictionary,* compiled by John Gage, 1981 (Baltimore, MD: Ottenbeimer), the final <-y> is always indicated in the pronunciation respelling by the compromise /i./.]

If we generalize this not only to plurals but to regular past inflections in verbs, we get an unexpected and highly desirable reward. One of the distinctive characteristics of English spoken in Scotland is that traditionally long vowels   (in the technical phonetic sense) are shortened when a consonant follows. This means that the vowel in *week* is as short as the one in wick and the two have to be represented phonemically by separate letters, and not as being in the relationship found in other dialects, e.g. RP /i:/-/i/ in Daniel Jones's conventions.

However, before the past inflection <-d> the traditionally long vowels are long. Thus,

/rod/	'road' but	/ro:d/	'rowed'
/brud/	'brood'	/bru:d/	'brewed'
/rud/	'rude, rood'	/ru:d/	'rued'
/grid/	'greed'	/ə'gri:d/	'agreed'

Similarly, the vowel in *side* is /ə+i/ but that in *sigh is* /ai/, and the past tense form keeps the /ai/: /səid/ 'side' /said/ 'sighed'.

In short in the standard English of Scotland (as described by William Grant in *The Pronunciation of English in Scotland,* 1970, College Park, Maryland: McGrath), vowels sometimes differ before inflectional endings and create contrasts not otherwise found. If we extended the use of the stop <.> to the past tense marker, we could deal with these contrasts perfectly.

In at least some dialects of American and Australian English, vowels are also sometimes lengthened before the past. So *side* and *signed* contrast but in a way different from the Scots forms; the realizations are /said/ and /sa:id/, respectively. In other dialects, there is a short and long form for /oi/, as in /ə'void/ 'avoid', /to:id/ 'toyed'. A true minimal pair in my own speech is /toid/ (pseudo-Brooklynese pronunciation of 'third', which is more accurately /təid/) and /to:id/ 'toyed'.

Similar phenomena have not been mentioned for RP altho Daniel Jones does talk about certain other morpheme boundary lengthenings. For example, he notes that the /ai/ in *bi-plane* is longer than in *pipeline.* Furthermore, he indicates that the word *highly* has a long /ai/ when used to mean 'in a high manner' (i.e. 'she praised him highly'), but a short one in the meaning of 'very' or 'very much '('he was highly praised'), and also in the name Haile (as in Haile Selassie). (In the fourteenth edition of Jones's *English Pronouncing Dictionary* (1977) revised by the late A. C. Gimson, indications of such length are dropped altogether because they are 'generally evident from the orthography and from the meaningful segmentation (morpheme boundaries) of the word' (p.xiv) — an unfortunate decision based on a circular argument.)

Let us assume that such distinctions do not occur in RP and a great many other dialects, before inflectional endings. Nevertheless, if we choose to mark such boundaries, this would be no great burden for speakers of any dialect since part of the linguistic competence of all English speakers is to recognize such boundaries for grammatical reasons if not because of pronunciation.

Furthermore, to introduce such a distinction would add considerable clarity to the reformed system since forms traditionally written separately would be distinguished: *daze-days, seize-seas/sees, size-sighs, nose-knows, use-ewes ; wade-weighed, tide-tied, road-rowed,* etc. (in modified *Nue Speling*: *daez-dae.zl seez-see.z, siez-sie.z, noez-noe.z, uez-ue.z, waed-wae.d, tied-tie.d, roed-roe.d,* respectively).

As a matter of fact this convention of <-.s> and <-.d> suggests even further that a single regular ending could be adopted for all inflections involved. We find something comparable in traditional orthography where the possessive is normally written <'s> even tho there are 3 variants (all of them shown, unfortunately, in *Nue Speling*: *kat's, dogz, hors'ez).* The traditional way is better and modified *Nue Speling* forms should be *kat'z, dog'z, hors'z.* But with <.> we can also drop the <-e> in plurals, third-person present tense forms, and regular past inflections. Thus:

plural:	*kat.z*	*dog.z*	*hors.z*
present:	*hop.z*	*rob.z*	*wish.z*	*hunt.z*
past:	*hop.d*	*rob.d*	*wish.d*	*hunt.d*

All in all, classic *Nue Speling* (the 1948 variety) tends to meet the general criteria Chomsky has proposed for a desirable orthography for English. With a few modifications, *Nue Speling* can counter most of the remaining objections Chomskyites have raised against spelling reform.

Altho English alone has been considered here, it should be clear that similar considerations could be used in the creation of rational orthographies for any and all languages in the world.

[Madhukar Gogate: see **Journals**, **Newsletters**]

5. A World Script is Necessary
Madhukar N. Gogate

[Madhukar Gogate is founder and Director of Roman Lipi Parishad, the movement for a common romanizcd script for Indian languages, and he submitted this paper to the 1985 Simplified Spelling Society Conference. He would welcome correspondence from readers.]

English dictionary is like a holy book. It gives thousands of commandments such as FRIEND/ SELL/ CHEMICAL/ THROUGH in place of rational spellings FREND/ SEL/ KEMIKAL THRU. The question is whether the spellings can be reformed and how.

To answer above question, it will be useful to know impressions of non-English persons. I am an engineer, and not a linguist, so my views may differ from what the experts feel. Having clarified this, I go ahead. My mother tongue is Marathi, spoken in Maharashtra state in Western India. The language uses devnagri script. When we learn a language, we learn its script, spellings, grammar, vocabulary, idioms and so on. I find English grammar relatively simpler. For instance adjective WHITE is applicable to horse, horses, cow, cows, paper, papers. In Marathi, the adjective changes according to gender, number and also the case. Numbering system in English is far simpler. After first twenty numbers, further numbers run in cycles, such as twenty one, twenty two, twenty three... thirty one, thirty two, thirty three... But in Marathi, all hundred numbers have non-repeating names. Thus 52 is read in Marathi with no bearing how numbers 5, 50 and 2 are read. The disadvantage of irrational spellings in English is thus compensated elsewhere.

Standard Oxford and Webster dictionaries are used in later life, but in schools we use local dictionaries, with English words, Devnagri pronunciations, and Marathi meanings. Since Devnagri is reasonably phonetic, with clear sound-symbol relations, our task becomes easier in the pronunciation part. As regards mugging and reciting spellings, that ordeal has to be gone through.

A script acts as a beautiful filter or armour. Take words LAMP (single L), ALLA (double L, both pronounced), CHURCHILL (double L, but one L is redundant pronunciationwise). In phonetic Devnagri, the extra L in the last word is oniitted. This shows that one way of solving the spelling problem is to use Devnagri script for English! But that would be unpractical.

Incidentally, Devnagri is not totally phonetic and often its phonetic quality is not properly used. Assume A = A in Alone and AA = A in Army. In Devnagri, word SARKAAR (=government) is actually written equivalent to SARAKAARA. A male name SHAAM is written equivalent to SHYAAM, with a silent Y consonant added. In my name too, due to Devnagri influence, a silent A has crept in, and my surname rhymes with COLGATE, WATERGATE, when I go abroad. I do not respell it to GOGTE, because it becomes a problem of certificates, passport, phone book, books written by me and so on. In no Indian language is India called India, we call it Bharat. Electron is both a particle and a wave. So I am called both GOOTE and GOGATE. Fine. I do not mind this duality and distortion. Such distortions occur also for Indian names with retroflex (N'), as English has dental (N). Even within Devnagri, Hindi and Marathi pronunciations differ for one symbol as GYA and DNYA.

Some of the symbols in Devnagri script are queer. For instance, MURTI is written in Devnagri in sequence MUITR. This can be rationalized by changing some stroke lines but the reform is not

accepted due to public inertia. We find that it is very difficult to change even a single dot in our scripts.

The foregoings should console the spelling reformers that other languages and scripts too have some absurdities. Spelling reforms, if they are to be made, need not be therefore 100% phonetic and 100% rational. A script, or a spelling system, is not a hifi taperecorder. People change fashions. People migrate from villages to cities. But people are reluctant to change spellings or script. I think the reason is simple. Whether to wear moustache or not, is a man's personal choice. But a change from FRIEND to FREND, howsoever rational, involves whole society, its books, its dictionaries, its roadsigns.

And yet, I am working for using the Roman script for Indian languages. Last year, along with some like-minded persons, I started an institution called Roman Lipi Parishad to popularize the Roman scripts (lipi). In India we have 15 languages written in 12 different scripts. Multiplicity of scripts results in barriers among people speaking different languages. A common script is necessary for things like phone books and train reservation charts. Otherwise there will be chaos, and we have been using the Roman script (English) for the purpose. In India we use English for higher business and higher education. Industries, the main consumers of typewriters, teleprinters, word processors, computers need machines only for English. Our scripts, though phonetically good, are not linear like Roman script. Their complex symbols can be accommodated in big printing presses, but not in above compact office equipment. India's economy is not such as to have luxury of machines for every script. On these considerations, we feel that a common Roman script will be useful for India.

Roman Lipi Parishad has held several meetings to consider sound-symbol relations, and within few months they would be finalised. Booklets etc. in different languages, with Roman script. Meanwhile we would like to know whether English spelling reformers have developed a master plan for sound-symbol relations.

I shall indicate the trend of our thinking, regarding the sound-symbol relations. These symbols are for Indian consumption, English spelling reformers may like to reserve A for vowel in APPLE and digraph AU for vowel in AUTHOR. But no Indian language requires vowel of APPLE, except for words of English origin. We have many words wherein vowel A is followed by vowel U, and vowel of AUTHOR may be shown by us as AW. We may use A as in ALONE, AA as in Army, AE as in Apple. We may take D for TH in They, and D' for D in Dog. Thus words INJECTION, DOCTOR, COLOUR (COLOUR) may be written in our language as INJEKSHAN, D'AWKT'AR, KALAR. In day to day writing, we may omit apostrophe marks. We are inclined to use a colon sign (:) after personal names, and highly technical words like POTASSIUM. It is not necessary to respell them. The colon sign would indicate that the preceding word is spelled in scheme not accepted by us.

When dealing with English words, for internal usages in our languages, we may either reform or deform spellings. Every language, including English, deforms words absorbed from other languages. Moreover, our languages are not international. So whatever formula we adopt will not upset the world.

But with English spelling reforms, there would be objections both at national and international levels. Suppose the reformers agree to respell FRIEND, SELL as FREND, SEL. What happens next? No student or typist will implement these changes unless authorities in different places accept them. Newspapers and book publishers would not accept them, as they are interested in selling their publications nationally and internationally. Non-English countries who teach English as a second language, may not approve the reforms. They too have made investment on current spellings and basically they are interested in their own languages, and not English. Without reference dictionaries, it may be difficult to use the reformed spellings for any serious work. They can be used for say some poems and nothing further. I am interested to know how the spelling reformers would handle popularization. In the case of Indian languages, romanization is likely to

get boost from the machine infrastructure. Economic and integration compulsions would drive us towards the Roman script. Romanization does not clash with existing script and a parallel language can be built. What are the compulsions behind spelling reforms movement? Are they life-or-death issues? My personal feeling is that English spelling reforms are possible if an international, logical script is developed usable by all languages. With this script, even names should be respelled and local feelings respected, naming JAPAN as NIPPON etc. The world will take interest in this project and an international pressure would be built up. Publishers and governments in English speaking countries would be electrified into action. The script would be logical like the metric system.

This new script may consist of lower case letters, without any capitals, with suffix symbols like apostrophe, colon and with slant stroke in place of full stop, to distinguish sentences. This script is available on existing machines.

EPILOGUE
Possibly after 50 years, 50% of mankind may know English. An attractive thought for world brotherhood. But to help future generations, today nobody will accept chaos. If spelling reforms were that essential, American government would have sunk millions of dollars on the project. Why are they indifferent? Probably because their economy is highly developed and interwoven. Everyday, millions of business letters and technical drawings are issued. Everyday, millions of newspapers and books are printed. The superpower cannot change spellings, which are bolts and nuts of its huge industrial apparatus.

A world script too may not be realized, due to political reasons. But an interesting possibility is that the spelling reformers may get thousands of sympathizers in India, when India accepts the Roman script.

Today, there are no Bengali-Marathi, Hindi-Tamil, Urdu-Kannada dictionaries etc. It is hard to locate complex symbols in one's own script, let alone other scripts. People will soon see the advantage of a linear Roman script in achieving mutual understanding and in economy of machines — the same machine for any language — and then there will be a big surge towards romanization.

Since India makes active use of English, Indians will start asking why English spellings are not rationalized. Today, about 20 million Indians know English. 20 millions, over one third population of Britain.

It must be understood that India will take trouble to romanize to uplift India and not to uplift English. Indians have great respect for English, but their hearts are with their mother tongues. Indian language literature is of fine quality. It awakens us to our social problems. We are more comfortable with our languages than with English. We have no patriotic songs or prayers in English.

Naturally, Indian sound-symbol relations will be such as to suit their languages, prepared by their near-phonetic scripts. They are likely to reshape English words as follows (if absorbed within their languages): America = A/ME/RI/KAA. Bible = BAA/Y/BAL. Unity = YU/NI/T'I. Freedom = FREE/D'A/M. Air = E'/A/R. Rod = RAW/D' = RAWD'. Road = RO'/D' = RO'D'. Pump = PA/MP. Some people may like to reserve A' for AA, since A is presently used for both A and AA. Perhaps, to accommodate English, Indians may take U' = (U in UP), I' = (I in BIBLE), reserving U for (U in PUT) and I for (I in IT). With a single apostrophe sign, no other sound can be considered. I wonder how far spelling reformers and romanizers may cooperate.

Whether Indian sympathizers will help English spelling reform — only history can tell.

6. When Noah Missed the Boat
Harvie Barnard

[Harvie Barnard is associated with Edward Rondthaler's *Typographic Council for Spelling Reform,* and submitted this paper to the Society's **1985 Conference**.]

It has been a well kept secret for the past 200 years that Noah Webster, our highly respected orthographer, once published a very significant study of the English language, his *Dissertations,* in which he recommended a complete overhaul of English spelling.

In addition to discussing the grammar and syntactical difficulties of English, Webster's *Dissertations* included an Appendix primarily related to the peculiarities of the spelling handed down to us by Samuel Johnson in his *Dictionary of the English Language,* publisht in 1755.

In his *Dissertations* Noah states: 'I will here subjoin what Doctor Franklin (Benjamin) has done and written to effect a reform in our mode of spelling.' This surprising statement supports Webster's previous confession that 'I once believed that a reformation of our orthography would be unnecessary and impractical. This opinion was hasty, being the result of a slight examination of the subject. I now believe with Dr. Franklin that such a reformation is practicable and highly necessary'. Such a statement from the great-grandfather of the millions of Webster's dictionaries publisht in the past 200 years will undoubtedly generate disbelief for many academics, printers, teachers and others committed to the inviolable sanctity of 'Webster's Dictionaries', many, if not most of which were edited and publisht without permission or knowledge of Noah Webster, his heirs or assigns. They were essentially copies, plagiarisms, taken much as Noah himself 'borrowed' Samuel Johnson's spellings, spellings based upon the pronunciations of those Anglo-Saxons who spoke, according to Johnson, in a 'wild barbarous jargon', a veritable potpourri of all the languages brought to England over a period of at least two thousand years.

Yet in spite of the chaotic confusion of many tongues which provided fodder for Johnson to organize, weed out and standardize, he certainly accomplisht what needed doing, and Webster recognized that he had an excellent base with which to work.

Still there were many imperfections, and in this view it appears that Webster had the support of several distinguisht writers of his time. With some pride he states, 'In the singularity of spelling certain words, I am authorized by Sidney, Clarendon, Middleton, Blackstone, Ash and other eminent writers, whose authority being supported by good principles and conveniences deemed superior to that of Johnson, whose pedantry has corrupted the purity of our language, and whose principles would in time destroy all agreement between the spelling and pronunciation of words.' Time and events have proved that Webster's dire prophecy was fulfilled, since the best intentions of both Franklin and Noah were never realized.

Somehow the orthography of Samuel Johnson prevailed, perhaps because his *Dictionary of the English Language* was recognized as the most thoroly researched publication of its kind at that time. Also his definitions and classical illustrations of the use of words and their preferred meanings were outstandingly superior. Yet the Johnson dictionary exposed two serious weaknesses pointed out by Webster and other writers who discerned the lack of dependable relationships between Johnsonian spellings and their sounds when spoken. The *Dictionary of the English Language* ignored pronunciation, for at that time there were no diacritical markings in use to aid the uncertain seeker for clues to proper enunciation.

As Webster succinctly stated, '...the same letters often represented different sounds, and the same sounds were expressed by different letters', a situation which has been carried over to the present as a stumbling block to both readers and pupils who are required to accept non-fonetic spelling as an insult to their sense of logical reasoning and rationality.

Webster's *Dissertations* were publisht in 1789, and this scholarly volume contained the sum of his views on the English language, especially in the Appendix in which he proposes a reform of English spelling. But thirty-nine years elapsed between the printing of the *Dissertations* and the publications of his famed *American Dictionary of the English Language.*

During these years Webster's mentor and admired friend Ben Franklin passed on, and Noah, 51 years younger than Franklin, became involved in writing and publishing to earn a living. A change of heart may have developed over this time, and perhaps his involvment with the famed Blue Backed Speller, publisht in 1783, and widely used thruout the American colonies and the new United States, may have resulted in a modification of his original plans for spelling reform. Perhaps it was the outstanding success of the Blue Backed Speller, which became the universally used primer in American schools and homes for at least 100 years, convinced Noah that regardless of its peculiarities English spelling could be mastered if the pupil was determined and the hickory stick was applied frequently and with sufficient force.

In his *Dissertations* Webster had said earlier that 'America is in a situation most favorable for great reformations, and the present time is, in a singular degree auspicious. The minds of men in this country have been awakened...'But the financial success of the Blue Backed Speller, which actually ran into many millions of copies — even tho it sold for only 14 cents per copy — could have been a deciding factor. Noah wasn't about to argue with success. The old spellings had been disseminated far and wide, had become part of the establishment.

Yet before achieving financial independence, Webster had proclaimed with fervor and enthusiasm: 'Now is the time and this is the country in which we may expect success in attempting changes favorable to language, science and government. Delay in the plan here proposed may be fatal; under a tranquil general government, the minds of men may again sink into indolence; a national acquiescence in error will follow, and posterity be doomed to struggle with difficulties, which time and accident will perpetually multiply'.

It is a great misfortune for English speaking peoples that altho Webster was 'right on', and his predictions for the dire consequences of delay have come to pass, we have continued to use Websterian spellings much as they were 'borrowed' directly from Samuel Johnson's original orthography. The result has been that during the past 200 years posterity has struggled and suffered with the queerly inconsistent spellings which have confused and frustrated many generations of English speaking children. Their children, grandchildren, and all those obliged to conform to the spellings declared sacrosanct by our dictators of diction and orthography, are still victims of a peculiarly insidious form of psychological child abuse which has resulted in behavior problems, antagonistic attitudes, academic failure, and a too high percentage of functional illiterates after years of forced 'feeding' in our public schools.

Yet Noah Webster, in his early exuberance of intelligent idealism, had preached a course of creative linguistic development when he proclaimed: 'Let us then seize the present moment and establish a national language as well as a national government. Let us remember that there is a certain respect due the opinions of other nations. In short, let it be impressed upon the minds of every American, that to neglect the means of commanding respect abroad is treason against the character and dignity of a brave and independent people.'

Webster concluded his *Dissertations* by quoting a letter from Benjamin Franklin, who writes:
    'in short whatever the difficulties and inconveniences now are, they will be more easily
    surmounted now than hereafter; and some time or other it must be done, or our writing will
    become the same as the Chinese as to difficulty of learning and using it. And it would already
    have been such, if we had continued the Saxon spelling and writing used by our forefathers,
    I am, my dear friend, Yours affectionately, *B.FRANKLIN*
    London, Craven Street, Sept 28, 1768.

Need more be said? In publishing his widely copied *American Dictionary of the English Language,*
using substantially the same spellings of Samuel Johnson's 1755 dictionary (there were a few
changes made, such as *honor for honour and center for centre),* our American spellers became
locked into the orthography of a by-gone period of English history. In abandoning his original
determinations and common-sense opinions of the weirdly inconsistent spellings espoused by the
'pedantry' of Samuel Johnson, the conclusion must be that 'Noah Missed the Boat' — the passage
from an archaic outmoded English to a truly American spelling.

That was in 1828. Times have changed many things. The language has greatly expanded and
modified both as to usage and pronunciation. Thousands of new words have been coined and
expressions been developed. In spite of the Johnsonian influence there has been a trend toward
fonic spelling, spellings which more nearly represent the spoken language. As Noah said in 1789
— perhaps he is more correct now than then — 'Now is the time and this is the country'.

And again quoting B. Franklin, '...some time or other it must be done', and 'whatever the difficulties
and inconveniences now are, they will be more easily surmounted now than hereafter.'

The enthusiasm of the younger Webster and the wisdom of the older Franklin have earned respect
and validity. Our problem now is the implementation of a rational reform of our American-English
spelling. After over 50 years of orthographic research, a logical program of fonetically acceptable
spelling has been developed, computerized, and is ready for use: AMERICAN, an *alternative*
spelling of the English language which is neither inflexible nor lacking fonetic integrity.

Change and improvements, like fashions, are inevitable. Over the centuries, modes of speech
have undergone change and will undoubtedly continue to undergo further modifications. Updating
the spelling and diction of our language may very properly become the responsibility and
prerogative of a national commission of linguistic experts. The fact that the use of AMERICAN, the
rational alternative, permits a degree of choice, yet within well defined fonic rules, provides
practicality with flexibility heretofore unavailable in any proposed program of modernized
orthography.

Because AMERICAN provides tremendous advantages in simplifying the teaching of reading,
writing, as well as spelling, the efficiency of public education may be substantially enhanced.
Therefore AMERICAN merits serious investigation and due consideration. Its potential in
revolutionizing and implementing communication on a world-wide basis also merits thoughtful
recognition.

Thus by providing a basically utilitarian means for world-wide cooperation thru understanding and
mutual comprehension, it becomes reasonable to accept AMERICAN, the alternative English
spelling, as an opportunity for friendly commercial coexistence, which is coming as close to 'peace'
as may be possible in this age of intense industrial competition.

(It is hoped the next issue of the Newsletter *will present an overview of AMERICAN alternative
spelling* — Editor.)

7. Spelling Reform 1 — And Nothing Else!
Garry F. Jimmieson

[Garry Jimmieson is Secretary of the Australian Spelling Action Society, which advocates as a first stage spelling reform (SRI) that the phoneme /e/ should always be spelt E. This paper was submitted to the Simplified Spelling Society's Conference accompanied by an audio tape.]

Of late SRI has increasingly been criticized from several quarters as:

a   not being an effective spelling reform on its own, due to it not being too conspicuous in day to day written articles

b   not being effective enough in altering those difficult words encountered early on by children

c   being a method of spelling reform which has tried but has achieved very little.

This paper addresses these very issues in a systematic and documented way with the aim of presenting the concept *of SRI — and nothing else as a viable proposition to be adopted by all.*

A. Spelling Reform 1 is an effective stage reform by itself

1.   SR1 is a spelling reform which gains great strength in the fact that it can be taught in schools due chiefly to it being a reform based on a *phonetic principle;* that is, SR1 selects certain letter patterns in t.o. which contain irrelevant letters.

2.   Studies by Peter Heggart, Lincoln Institute (Carlton, Victoria, Australia, 1978) resulted in these findings:

'Presentation of Data

This paper presents a list of what might be considered the 500 most commonly occurring content words in the English language — the 200 most commonly occurring nouns, the 200 most commonly occurring verbs and the 100 most common adjectives. The words were selected from amongst the most widely-used word lists available, word lists which included frequency counts, basal reader core vocabularies, spelling lists and spoken language corpora. The lists were considered to be representative of English-speaking populations and were chosen for their popularity, utility and representativeness, including both dated and recently compiled lists and those which have their origins in the language samples of British, Canadian, American and Australian populations. In all, twelve lists were consulted and from these a compilation was made of the 500 most commonly occurring content words in the English language.'

When applying the principles of SR1 to this list it is found that:

a. of the 200 most common nouns in the English language, 2% of these would be transformed

b. of the 200 most common verbs in the English language, 1% of these would be transferred

c. of the 100 most common adjectives in our language, 5% of these would be transformed.

3. Studies by Ves Thomas, University of Calgary (Canada 1977) resulted in these findings:

Comparative results as % of total word count

No.of Difrnt. Words	New S. Wales (Gr.2-6)	Alberta (Gr. 1-6)	USA (Gr.1-8)
50	44%	48%	40%
100	54%	58%	60%
500	75%	78%	82%
1000	82%	83%	89%
2000	90%	89%	95%

In general, the results shown continue to support the previous findings that a small core of words dominates children's writing. The 100 most frequently used words (and their repetitions) continue to account for over one half of all the writing which children produced whether this was in Australia, Canada or the United States of America. However, it must not be assumed that the same 100 words form the core for each study. Actually there is only about an 80% overlap between and among each of the three lists of 100 words. It is also worth noting that the first 2000 words in order of frequency of usage continue to account for approximately 90% of all the writing done by children in grades 1–6. It is this core of 2000 words which has for many years provided the basis for the core curriculum in spelling.

The major purpose of this study was to identify those 2000 words which form the core of the writing needs of primary school students in New South Wales. Although such a core of words was identified, the limited size of the total writing sample tends to reduce the validity and the reliability of those words which appear at the lower end of the frequency scale. With this concern in mind and using Rinsland's criterion for excluding words which have a frequency of less than three, a composite list of 1500 words has been prepared for publication. Furthermore, this list has been derived from an analysis of all writing samples without reference to grade. The words are presented in descending order of frequency.'

When applying the principles of SRI to the published list it is found that:
a.   Of those words that account for 54% of all words written, *1%* of these would be changed
b.   Of the remaining 1400 words which comprise another 36%, *1.7%* of these would be changed
c.   *Overall 1.6% of words in this list would be altered by implementing SRI.*

4   As Peter Heggart claimed in 1978:
'Evidence has shown that the most commonly implemented word list used by teachers in Australian schools is the Dolch Basic Sight Vocabulary (Dolch 1941); American teachers prefer the Heritage List (Carroll et al, 1971); while children in the United Kingdom are usually exposed to the frequencies of the Ladybird Key Words (Murray, 1963).'

However when the Dolch Basic Sight List is viewed, we transgress from not only examining a child's core list but also we view an adult's core vocabulary list. For as Dolch explains as an introduction to this list:

'These words are so selected as to include those words which make up 95% or more of all the words written by the average person.'

When applying these principles of SR1 to this list, it is found that *1.7%* of these words would be affected.

5. Obviously, each person's written vocabulary list is a highly individual one; however when one considers the written articles by the printed media, consideration of the core list concepts' must be made. As I have shown, the occurrence of SRI words can be seen to vary somewhat between 1%–5%, depending on the information-source. In his book *Spelling Reform A New Approach,* Harry states that SR1 would affect '...only one word in 165' (p.128). He also believes that'... if a spelling reform does not get established in newspapers, it will never get established' (p.19). The importance of this last assertion cannot be overemphasized, when one considers that '90% of the Australian population read little outside newspapers and popular magazines' (Dr R McAvaney, Language Centre, University of Sydney, N.S.W., Australia, 1977). When Harry's word count for SR1 is considered in this context, as the Printed Media use such high frequency SR1 words *as sed, many, helth, frend,* etc., I would imagine that Harry's word count to be highly underestimated.

6. *Thus I assert that* by adopting SR1alone, the number of words affected world wide is enough so that:

a.   it is small enough to satisfy the principle of 'least disturbance'

b.   it is large enough to change enough of the 'high frequency words'

c.   it is a 'teachable reform' within the education system of eny country.

B. Spelling Reform 1 does affect those difficult words encountered early on in school by children.

1. In Queensland the Department of Education has divided the Spelling Program into 3 sections. When SR1 is applied to these separate lists, it is found that:

a.   of the 100 words in the Basic Core Spelling List, 1% of words would be affected

b.   of the 2000 words in the Body of Content List (words to be taught to children in Years 2–6), 1.85% of words would be affected.

c.   of the 100 words in the Words of Special Difficulty (Demons) List, 10% would be affected. Of these words, 5 are encountered in Year 2, 2 are encountered in Year 3, 2 are encountered in Year 4, 1 is encountered in Year 6.

WORDS OF SPECIAL DIFFICULTY (DEMONS)

ache	here	could	read	too
again	country	hoarse	*ready*	trouble
always	dear	hour	*said*	truly
among	doctor	*instead*	*says*	Tuesday
answer	does	just	seems	two
any	done	know	separate	used
been	don't	knew	shoes	very
beginning	early	laid	since	wear
believe	easy	loose	some	Wednesday
blue	enough	lose	straight	week
break	every	making	sugar	where
built	February	*many*	sure	whether
business	forty	*meant*	tear	which
busy	*friend*	minute	their	whole
buy	grammar	much	there	women
can't	*guess*	none	they	won't
choose	half	often	though	would
colour	having	once	through	write
coming	hear	piece	tired	writing
cough	heard	raise	tonight	wrote

2. 1 suspect that if other Word Frequency Lists were examined so as to determine the difficult words, SR1 words would feature as prominently as they have in this Queensland list. Hence I advocate that research in this area could only benefit the cause of spelling reform world wide.

C. Spelling Reform 1 is not ded: it is very much alive and kicking.

1. Harry Lindgren's Spelling Reform 1 has achieved:

a.   as Valerie Yule has acknowledged, the concept of SRI broke out of the 'thought mould' about spelling reform in that a principle rather than a list of sweeping changes might be the best way to start

b.   might I be so bold to refer to SR1as the forerunner of the 'stage' or 'fase' concepts?

c.   as Chris Jolly has commented, SR1 is probably the most accepted Spelling Reform today

d.   in its 'hedy days' around 1975, over 700 people Australia-wide were members of the Spelling Action Society — making it the largest spelling reform movement of its kind

e.   in 1975 the Australian Teachers' Federation adopted SR1 as a policy. As this body is changing to the Australian Teachers' Union, this commitment has yet to be reaffirmed,

2. There are moves afoot to decentralize the workings of the S.A.S, and thus giving it the potential to achieve its aim sooner.

8. 'Spelling Reform Now'
H. W. Herbert

[H.W.Herbert, Brisbane, writes for the Queensland *Sunday Mail* and is convenor of the Advocates of Simplified Spelling Australia. He submitted this paper to the Society's **1985 Conference**.]

'Spelling Reform NOW!' What does that intriguing word NOW mean? Does it mean 1985? Does it mean that this is the year of opportunity for spelling reform, that after 80 years in the wilderness the promised land is within sight?

Many of us think it is. I will now outline the 'toe in the door' strategy that we believe can succeed.

Spelling reformers are devoted intellectuals who often forget how conservative the general populace is. Hence spelling reformers have chronically been too ambitious, proposing changes the public will not accept, such as Bernard Shaw's 42-letter alphabet.

'Toe in the door' strategy means proposing changes that the public will accept right away, while foreshadowing further changes. After intense discussions, spelling reform groups in Canada, the United States, Australia, and perhaps Britain are agreed on a list of 20 <ough> and <augh> words for immediate reform, as listed:

	Present Spelling	*New Spelling*		Present Spelling	*New Spelling*
1	enough	*enuf*	11	ought	*aut*
2	rough	*ruf*	12	bought	*baut*
3	tough	*tuf*	13	brought	*braut*
4	though	*tho*	14	nought	*naut*
5	dough	*do*	15	sought	*saut*
6	through	*thru*	16	thought	*thaut*
7	thorough	*thuro*	17	caught	*caut*
8	plough	plow	18	taught	*taut*
9	drought	*drowt*	19	draught	*draft*
10	cough	*cof*	20	laugh	*laf*

It so happens that the above list of 20 coincides exactly with the question I asked my newspaper readers at New Year. In place of my regular column on economics, I wrote on the advantages of spelling reform, and requested readers to write in, saying whether they agreed or disagreed that the <ough> and <augh> words should be made phonetic. Over 100 people wrote.

As far as I know this is the first time a specific spelling reform has been tried out on the general public, so we should pay attention to the result. A surprisingly high 85% were in favour of making <ough> and <augh> phonetic.

Of the 15% who wanted no change from present spelling several said the answer was to discipline the children more firmly — make them learn the traditional spelling. This is of course just what spelling reformers want to avoid — children (and teachers) wasting time and effort learning unphonetic spelling. Others objected to the strangeness of the phonetic spellings. True, new forms take some getting used to.

Traditional spelling is less founded in scholarship than its adherents imagine. The early type-setters were not scholars, they were semi-literates who spelt any way they fancied, including using <ough> for several sounds. So when Dr. Johnson froze our spelling '<ough> had 7 different pronunciations and still has.

Traditionalists do not object to having to use a glossary to read Chaucer; they do not think any less of Milton's splendid epitaph on Shakespeare because Milton spells in pre-Johnson form dramaticke and conceaving. Sympathize with the traditionalists and assure them they will soon get used to the new spelling.

Spelling reform *NOW* and now is 1985. One small toe in the door, one giant step for spelling reform.

In announcing the first step reform of 20 <ough> and <,lugll> we must also foreshadow changes to be included in Step 2. Here we must exercise restraint, selecting unphonetics that will have the greatest appeal to the public.

My selection of foreshadowed changes would be three groups:– ridiculous unphonetic words like *one, once, tongue, yacht*; SR1 (*hed* for *head*) because it is a simple rule; and silent letters as in *would, dumb*. Foreshadowed changes do not have to be given in full detail with Step 1.

Let us now look at the three solid propositions on which spelling reform rests. Everyone here is familiar with the first — that spelling reform would save an immense amount of time and effort that children and teachers now put into learning unphonetic spelling, time that could be spent learning useful arts and crafts.

The second proposition on which reform rests is the need to make it easier for people with non-English mother tongues to learn to read and write English. Many of these people now speak English surprisingly well, especially in Northern Europe. English has also become the business language of Asia. Learning to read and write it, however, is quite a different and difficult matter.

These people in Europe and Asia can hardly credit that such a simple, expressive language, so uncomplicated in its structure — no genders, no suffixes — should be spelt in such a crazy ridiculous way.

It is nearly as bad as the frustrated German who complained 'You inglisch haf this name C-H-O-L-M-O-N-D-E-L-Y and you call it 'Marchbanks'. Progress will be slow, but continuous, as we progressively phoneticize English spelling and thus help foreigners write it.

The third proposition is the urgent need to help the under-achievers. There is a widening gap between the under-achievers and the average and above-average school-children. The underachievers suffer much more unemployment; only a few small countries have managed to

correct this. In the United States unemployment among blacks and Latinos is shockingly high and persistent.

Many of these ethnics, and other under-achievers in Canada, Britain and Australia are rebels against normal social values, frustrated rebels, who demonstrate it occasionally in orgies of destruction. In Australia it takes the form of vandalizing and befouling classrooms and equipment and setting fire to wooden schools.

I cannot of course prove it but I suggest that there is a link between the under-achievers being hounded through unphonetic stupid spelling at an early stage, and the destructive behaviour they develop — they hate schools.

Years of low unemployment (until 1974) hid the problem of illiteracy, and the under-achievers were slotted into process work and other jobs requiring low literacy.

No longer. Economic development is widening the gap between the average and above-average people, who get jobs, and the under-achievers, who don't.

There is no easy answer to this problem. We are going to have to provide jobs for the under-achievers, simple jobs as reafforestation, manufacturing goods at present imported in large quantities, fetching and carrying for service industries like tourism. The greater the degree of literacy the under-achievers can acquire, through phonetic spelling, the easier they will find it to read labels and instructions, and write messages. A slow process, but certainly worth doing.

How can we ensure that Spelling Reform *NOW* has a big impact on the public in 1985? It would be unprecedented and newsworthy for groups in 4 English-speaking countries to agree on a program of reform. This would also be newsworthy — a Boston Teaparty in reverse, with the erstwhile colonies trying to persuade the mother country to have greater representation!

How do we follow up the first impact of the <ough> and <augh> reform? I like the idea of putting the list of 20 into verse (with as few other words as possible). Such lines as 'Enough's enuf, ruf tuf tho do is', 'cof caut braut to naut'. Is there a word-weaver among us? Can we find one? Do we ask word-master Alan J Lerner, who owes Bernard Shaw a debt for My Fair Lady?

Then we could put the verse to a memorable tune — like Blowing in the Wind. It would help children and foreigners to keep the list of 20 well in mind.

Spelling reform NOW, and NOW is 1985. Seize the opportunity.

I must finish with a warning. Miss this opportunity and editors will write off spelling reform as a good idea that failed.

We won't get another chance. Our efforts will contract into interminable debate on spelling niceties, in which the public isn't the slightest bit interested.

9. Dropping Useless E
Gilbert Rae

[Gilbert Rae, a member of the Society for several years, has previously published articles in the Newsletter. His interest in spelling reform dates back. to a chance meeting with Mont Follick, the M.P. and enthusiastic simplified speller. He was taught phonetics from his first day in school and later learned several languages.]

In simplifying English spelling, the advantage of dropping useless E seems strikingly clear as there are so many. However, the uses of E must be examined first. Letter E has two uses. It is a vowel with the usual short and long forms. It also functions as a retroactive diacritic letter affecting the short vowel that precedes it, converting that vowel to a long one, and affecting the consonants C and G, converting them from hard to soft sounds S and J. As in all English pronunciation there are exceptions to the rule, e.g. *gem* is correctly pronounced 'jem', but G in *get* remains hard as in *got.*

Letter I is the only other vowel behaving to some extent like E. At present it has short and long forms, and is frequently pronounced in the French way in words that were originally French (Fr. <i> = Eng. <ee>). The letter I also acts as a retroactive diacritic, mainly used with the vowel A as in *gain,* and occasionally with U as in *fruit,* and it softens letters C and G. There are many exceptions to these rules which with other irregularities probably make English pronunciation appear to foreigners as complete chaos.

In the cases of soft C and G, E and I must be placed on the right of the consonant, e.g. *certain, civil, general, engine.* There are exceptions of course, such as *give* and *ogive,* pronounced /giv/ and /odjaiv/.

In general, short vowels are basic, and can be made long by placing E or I immediately to the right of them. It is also allowed to have a consonant between the converted vowel and the letter diacritic. When E diacritic is the last letter of a word, it is usually silent, e.g. *made, here, ripe, hope, jute.* To have more than one consonant between the vowel and the diacritic letter destroys the power of the diacritic letter. From the foregoing it follows that in any word where the letter E is silent, and where it does not act as a retrospective diacritic either, then it is useless, and can be dropped.

Some words can confuse as regards spelling because the vowels are not strictly long or short, or are exceptions to the rules. These are some samples: *more awe are were course come done oven minute private.*

More The dictionary pronunciation is /mōr/, so the E is a retroactive diacritic, and not to be dropped.

Awe Because the pronunciation is accepted as AW which is the sound of a modified short /o/, produced in the throat, it can be accepted as practical spelling. Many Londoners pronounce *awe* as O-AH which is not far out, but amazingly they pronounce the word *or* in the same way! The letter E in *awe* is not just useless, it is completely misleading. It does not act as a diacritic, nor is it sounded. It is useless and should be thrown out.

Are, Were Dictionary pronunciations are /ar, wer/. Neither final E is sounded, and they do not act as diacritics. Drop the useless Es in both.

Course Dictionary pronunciation is /kors/, but its Latin origin suggests that once it must have been pronounced as in France. The two consonants in front of the E render it useless and it can be dropped.

Come, done, oven Pronunciations ar /kʌm, dʌn, ʌvn/. As we already have the short U giving us the explosive sound which is also found sometimes with the letter O, e.g. *son, sun,* it would seem an advantage to return to the genuine short O sound where letter O appears. There are enough changes to be made in English spelling to keep spelling reformers busy for a long time to come, and consideration of changing O to U can be postponed. Meanwhile it can carry on is now as an alternative pronunciation. Letter E in such words as *come, done, oven* does not act as a normal retroactive diacritic, and it is not sounded in *come and done.* In *oven,* E ought to be sounded, and therefore should be retained.

Two more words to examine for useless E:
Minute Dictionary pronunciation is /minit/ for all meanings of the word except where it signifies 'very small', when it is sounded as /mainju:t/ and E acts as a diacritic and cannot be dropped. Respell *minute* (60 seconds) as *minit.*

Private Dictionary pronunciation is /praivit/ or /praiveit/. English is pronounced carelessly, unless clear pronunciation is required by the audience. If spelling is to be changed to conform to careless speech with short I or its equivalent schwa dominating the language, so be it, but we shall still have to know how to pronounce the stressed vowels, both primary and secondary. Meanwhile we still have to drop useless Es. With *private,* the E is acting as a diacritic and therefore cannot be dropped.

All silent Es however can be dropped if we use our other system of long and short vowel diacritics, instead of the letters E and I. This E & I system is very ingenious, but it is not so simple or efficient as the diacritic marks of more recent years. In this article both systems are used. The Roman system is used every time we write in English, and the true diacritic marks are used where pronunciation has to be indicated. For the most part of course only the long sign in required. (See *private* and other examples above...)

10. Phonics and Spelling
David Stark

[David Stark aims to combine skills and experiences as an Adult Literacy Tutor and as a designer, to help formulate a practical spelling reform. This article is the second in a series, of which the first *Alphabetic Consistency in Reading,* appeared in the **Summer 1985 *Newsletter*,** Item 13. He has also contributed to most other recent issues of the *Newsletter.*]

In my last article I argued that it was understandable that the educational establishment could ignore the benefits of a more alphabetically consistent orthography for reading, since the reading process is basically morphographic. That is, the workings of our brain require 'that whole words and groups of words be processed at a time, these forming potentially meaningful units which can be combined to reveal the meaning of text. Also, where an unfamiliar word does require to be broken down into smaller units in order to identify it by its pronunciation, it is more efficient to use syllables rather than individual phonemes.

Educationalists can argue that it is easier to make up for any deficiency in English orthography by extra teaching, which is their speciality, than to contemplate tampering with the established orthography, thus causing incalculable confusion for existing literates and children caught in the transition period.

Many educationalists employ similar arguments for the writing process. Judging reform to be a nonstarter, even if it is considered, they point to the fact that the children who are best at spelling have developed the ability to visualize spellings before writing them, and they advocate the teaching of skills which encourage this. They play down the phonic approach which they, quite correctly in traditional orthography, regard as unreliable and limited.

One must share a certain amount of sympathy with such an approach for a number of reasons. Firstly, as with reading, it is more efficient to process whole words or at least syllables at a time. When one is writing, one must hold in one's short term memory the individual words in sentences, which form part of the overall message, long enough to write them down. If one has visually memorized the spellings of individual common words, and all the common syllabic elements for uncommon words, one will be able to write a sentence before its part in the overall meaning is forgotten. One might argue that plodding through individual phoneme/grapheme relationships would place a greater burden on the short-term memory, with a tendency to allow only less complicated or shorter messages to be written at a time.

Secondly, the only reason one ever writes anything is for it to be read, and if reading is more efficient when it contains morphographic clues, it follows that these have to be built into text when it is written. The only way to remember homophones is to relate meaning directly to the visual transcription of words, and if inflections are morphographic, the writer will have to ensure that his spellings are meaningful, rather than following an imagined pronunciation which might be unreliable or perceived differently by the reader.

Thirdly, spelling requires much more precision than normal pronunciation can give. Actual pronunciation is too variable and imprecise to act as a wholly reliable standard for an orthography. (A fuller discussion of this follows later.) In any language this will render the oral word memory somewhat deficient placing some reliance on our ability to remember what words look like.

One uses one's knowledge of the appearance of words and possible letter strings or groupings to check if spellings look correct once they are written. However, when I was being taught spelling, I was not allowed to write down a word during a dictation exercise to see if it looked correct, and then score it out and try again if it did not. I had to get it right first time or it was considered a mistake. This encouraged me to develop my recall of the visual appearance of words, or as it was said, 'be able to see the word on the inside of my eyelids'.

I learned many new words by their visual appearance rather than their pronunciation, and in the literary tradition of language learning, I was taught French in a similar way, with less importance being given to pronunciation. In the past, before widespread travel between countries became easy, it was the norm to learn to read the literature of a foreign language without much concern for the living pronunciation. Perhaps this was just as well for French and English where the relationship between the written word and pronunciation is unreliable.

When I decided to learn conversational Italian for travel abroad, I found I was unable to learn it 'the natural way, as children learn it', by absorbing the pronunciation of words. The literary approach was so much fixed in me that I had to see a new word written down first in order to memorize it. The pronunciation was derived from the appearance of the word or learned after it.

There may be an advantage in reforming English orthography to aid learning of English as a foreign language by being taught pronunciation first and then relating, this to the written word. However, it would most certainly be highly beneficial for native speakers of English to relate their wide knowledge of the oral language more directly to the   written word. Educationalists who advocate that traditional orthography can be easily and effectively learned using primarily visual methods are ignoring the fact that, in spite of its irregularities, English orthography is overwhelmingly alphabetic. Children work this out without being taught it, even if they cannot learn by themselves how to use all the alphabetic clues in the best way.

In practice we use a mixture of phonic and non-phonic cues when we spell English, the non-phonic ones being used in words or bits of words which we have learned are not phonetically regular. For example, one might use phonics to remember the Bs and T in the words *bough and bought,* but the <ough> is remembered as a letter string, independently of pronunciation.

Spelling reformers point to the fact that there are at least 2 million adults in Britain who can be described as illiterate, and the vast majority of die rest have problems at some time with spelling. They argue that since all the efforts to teach our way out of the situation, including a 10 year programme of Adult Literacy Training, have been completely ineffective in reducing the overall problem, the only solution is to reform the orthography so that phonic cues can be increased and non-phonic ones decreased. The premise is that children remember pronunciations of words more easily than their visual appearance.

If we, as spelling reformers, can lead sceptics through the minefield of argument to reach this conclusion, we will be rewarded with entry to another. Spelling reformers have spent a great deal of time in the past inventing sets of graphemes for revised orthographies when the greater problem involves the definition of a set of phonemes and a pronunciation standard to link to the lexicon.

Up until about the 1950's, reformers on this side of the Atlantic could argue that British Received Pronunciation (RP) should be used for this purpose. It had been widely studied and defined for some time, and had become established as the verbal currency of the BBC. As such it reached the homes of most Britons, and through the British Empire and in connection with the Second World War it was heard in many countries throughout the world.

However the days are long gone since the BBC hired only RP speakers as newsreaders and presenters. A large number of accents are heard over die media as the central sources of broadcasting become more democratized, local broadcasting develops, and the influx of American and other non-British accents continues. There is no standard accent which prevails in the world and there is no trend towards one. To try and impose an accent like RP as the standard for a revised orthography is totally impractical. Apart from anything else, one would have to force media presenters, teachers and all others who could be used as a point of reference to take elocution lessons so that their accents became standardized and frozen.

In addition to this, even if such a standard accent could be established, it would not be an exact model for an orthography. No writing system is a phonological record of speech, nor could it be. The variability of human speech, displaying emotions stress and intonation, does not produce a machinelike consistency. When we hear human speech, we are so used to this variability that we only use the sounds heard as a rough guide to the words spoken. We rely as much on our hypothesis of whether the possible words made sense in context in order to identify them. The individual jigsaw piece only makes sense when we know die meaning of a large section of the puzzle.

This is why voice recognition computers are so ineffective at present. To program powerful 5th generation computers to deal with processing human speech, linguists and computer scientists will have to define what is meant by 'meaning', and teach the computer how to recognize it. In the statement 'the man played (a round/around) with his secretary', a computer would need to appreciate context to determine whether to write 'a round' or 'around', or deduce whether the subject referred to is a keen golfer or a dirty old man.

We would have the same sort of problem transcribing a foreign language which we did not understand. Even if the person we were listening to spoke the standard dialect and the orthography were as alphabetically consistent as it could be, our transcription would be hopelessly wrong. We would not even be able to separate individual words effectively.

If one studies the definition of the pronunciation of words using the International Phonetic Alphabet as displayed in a large dictionary, one would gain the impression that phonetics is a precise science. However, the precision of a scientific device like a voice spectrograph in studying a particular accent is useless when confronted with the variability of a human voice. Phonetic researchers do not even trust tape recordings of speech, preferring to study subjects at first hand, so that they can observe the various parts of the mouth and vocal tract as words and phonemes are pronounced. Even a trained linguist will be afraid that his own preconception of how he thinks a word should sound will affect the way he perceives the word being spoken.

The conclusions to be drawn from all this are: that it is impractical to choose and fix a standard living pronunciation; that normal speech in such an accent would not be consistent enough to translate directly into written text; and that people do not possess an ability to define individual phonemes accurately, having found the development of comprehension skills as rewarding a way of identifying words.

To find a practical way forward for a revised orthography, we must explore the nature of the pronunciation of words which can be used in a Phonetic approach to spelling, and how this relates to die various living English accents. However that must be left for another article.

References
Jorm, A F, *The Psychology of Reading and Spelling Disabilities,* Routledge and Kegan Paul, 1983.
Peters, Margaret, *Spelling: Caught or Taught?* Routledge and Kegan Paul, 1985.