A REVISED SPELLING RULE

A. ROSS ECKLER
Morristown, New Jersey

In the February 1974 issue of Word Ways, Ramona J. Quincunx examined in some detail the ability of the venerable rule "Place I before E except after C, or when sounded like A as in neighbor or weigh" to predict correctly the spelling of a myriad of IE- or EI-diphthong words in English. She concluded that a new jingle, probably longer than the original one, was needed to more accurately describe the vagaries of English orthography. It is the purpose of this article to present such a jingle, and to compare its performance to the original one in a more-or-less objective manner.

In evaluating the performance of a rule, not all words should be given equal weight. Very few people have the need to spell teiglach or ophicleide, whereas everybody needs to know how to spell veil or foreign. In short, it seems plausible to weight words according to their frequency of use in English text. Fortunately, a large body of suitable data is at hand: Henry Kucera and W. Nelson Francis's Computational Analysis of Present-Day American English (Brown University Press, 1967), a sample of one million words taken from a varied collection of writings by American authors that were first published in the United States in 1961, listed with their observed frequencies of occurrence. From this list, I extracted over 800 words: all words containing EI- or IE-diphthongs which appeared two or more times in the sample of a million (extraction of the single-occurrence words would have nearly doubled the search-time, but would have increased the total word-usage a negligible two per cent). Although the statistics on the usage of individual words are rather unreliable for the low-occurrence words (one that was sampled twice in a million might, in reality, be represented as much as 5 or 10 times per million words of text), the collective statistics on larger groups are more stable.

How good is the old rule? Ms. Quincunx claims that the first half of the rule was meant to apply only to diphthongs with a long E sound, and the last half of the rule to diphthongs with a long A sound. If this is true, it presents the rule in its most favorable light, for otherwise the word their, weighted with 2670 occurrences, would be an exception to the rule (most dictionaries distinguish between the sound of A in mate and the sound of A in care). This strict interpretation of the rule includes only about 65 per cent of EI- and IE-diphthongs by usage; the other 35 per cent, in addition to their, have sounds such as short I (foreign), long I (Eisenhower), long U (review) and so on. Any rule worth its salt ought to strive for more complete coverage.
Within this limited class, the rule does moderately well:

I before E: 6219 occurrences in favor, 575 against
Except after C: 454 occurrences in favor, 258 against
Or when sounded like A: 664 occurrences in favor, 2 against

The one word violating the last part of the rule was grenier, a French term for an attic used as a studio. Collecting these statistics, we find that the rule is satisfied in 7337 out of 8172 cases, or approximately 90 per cent of the time.

To replace the old rule, I propose the following jingle consisting of five rhymed couplets:

EI at start, IE at end; inside a word, IE defend.
Except to this rule abound: with C in front, EI is found
In ceiling and receive (and kin); for their and heir, put EI in;
And, if one places GANTZ behind, EI's the pair to keep in mind
(But waive this rule and use IE for patient, friend, deficiency.)

To the uninitiated, this jingle may seem somewhat opaque. However, it was designed not as a first course in spelling, but as a memory-aid or mnemonic for the person who has previously been instructed in the vagaries of IE and EI spelling.

Let us examine the rule closely. "EI at start, IE at end" codifies the fact that most words start with EI rather than IE, and most words end with IE rather than EI. If the diphthong is inside a word, the chances favor it being IE, so this is presented as the policy to follow if none of the exceptions discussed in the last four lines apply (more precisely, these lines describe three types of exceptions, and give in parenthesis an exception to the final exception):

1) As Ms. Quincunx noted, there are only five common English words (and their derivative forms) which contain the trigram CEl: ceiling, receive, perceive, deceive, conceive. However, a host of words contain the trigram CIE: fancied, policies, etc. The last half of the second line and the first half of the third succinctly remind us of these facts.

2) Their is a major exception to the rule. It proved impossible to handle this case by suggesting that an H in front converted IE to EI, for then common words such as shield, chief and thief were made exceptions; similarly, it was impossible to suggest that an R following the diphthong converted IE to EI.

3) If the diphthong is followed by any of the letters G, A, N, T or Z, change IE to EI (as in eight, pharmacopeia, veined, either, seize). However, certain classes of words do not obey the GANTZ exception: words having the tetragram IENT (patient, ancient, sentient, resilient, convenient, omniscient), the tetragram IEND (friend, fiendish), or the tetragram IENC (efficiency, deficiency, sufficiency, proficiency). The words given in the fifth line were picked as examples to alert the reader to these three classes.
It is worth noting that this rule does not depend in any way upon
the vagaries of pronunciation for its application; all parts of the rule
depend only upon the position of the diphthong within the word and the
letters immediately preceding or following it.

How well does the new rule perform? Not only does it include the
35 per cent of EI- and IE-diphthongs ignored by the old rule, but it
works on more than 98 per cent of all word-occurrences: out of 14278
occurrences, only 257 violated the rule. For the record, I list all the
Kucera and Francis exceptions below:

18 Andrei
13 frieze
11 leisure
10 Lucien, Steichen, weird
9 chien
8 veil
7 Heiser
6 Buchheister, Heidenstam, Reich, siege, surveillance, Weider
5 Blenheim, Dietrich, Hotel, leisurely, monel
4 Neiman-Marcus, sheik
3 apartheid, Durkheim, Fleischmanns, friezes, Heidenstam's,
   hygiene, ierulli, matel, Neil, O'Neil, Reid, unveiled,
   veils, Viet
2 Andrei's, Antietam, besieged, besiegers, conseil, feint,
   Fleischman, Heidegger's, Heilman, lien, Neisse, Nietzsche,
   Reichenberg, Riegger, seidel, Sergei, tien, Wangenheim, Wei,
   weil, Zeiss, Ziegfeld

The number at the beginning of each row indicates the number of times
the word occurred in the million-word corpus. A substantial fraction
of the exceptions are proper names, which are often derived from for-
egian sources unpredictable in their orthography. The most well-known
exceptions on this list are probably frieze, leisure, weird, veil, siege,
hygiene and sheik; other reasonably well-known exceptions not appear-
ing two or more times in Kucera and Francis are geisha, heifer, heist,
poltergeist, clothier, mien and klieg light.