

The Frequency and Versatility of Initial and Terminal Letters in English Words

ROBERT L. SOLSO

University of Idaho

CONNIE JUEL

University of Texas at Austin

AND

DAVID C. RUBIN

Duke University

A review of previous word and letter counts in addition to the applications of these counts were reported. A comprehensive count of initial and terminal letters and bigrams was compiled based on the Kučera and Francis (*Computational analysis of present-day American English*. Providence: Brown Univ. Press, 1967) corpus of English words. The count included frequency of occurrence and versatility, or number of different words in which letters or bigrams occurred. It was shown how such counts can be used to describe "Englishness" and make predictions as to the information load of letters in words and pseudo words.

Current research interest in the reading process and memory storage of words has necessitated an increased understanding of word orthography. This need has been facilitated by large capacity computers, which have replaced the tedious hand counts of English words and letter frequencies. In the present paper a review of current word and letter counts is presented, an analysis of contemporary and potential use of orthographic structure is made, and a comprehensive count of initial and terminal letters is presented.

The examination of orthographic structure in a language is usually preceded by an analysis of the occurrence of words. The first modern count of American words was begun in the 1920s by Thorndike and resulted in a massive word frequency count done by hand called the "Thorndike-

Lorge" count (see Thorndike & Lorge, 1944). The Thorndike-Lorge count served as the standard source of word frequencies for many years but is gradually being replaced by two more recent and more sophisticated samples of the frequency of American words.

A large portion of the work on American English orthography has been based on a word count gathered and analyzed by Kučera and Francis (1967). These researchers determined the frequency of words (called tokens) as well as the number of different words (called types) which occurred in their corpus selected from slightly over one million words of running text drawn from a diverse sample of American literature. The Kučera and Francis corpus is recorded on magnetic tape, thus allowing researchers to perform numerous complex operations based on a well-selected and large sample, including identification of specific letter characteristics of words. The corpus was drawn in 1961 from 15 different categories which included newspapers, religion, skills,

The authors wish to thank Karen Falke for writing the computer program for the present research. Address reprint requests to Connie Juel, Department of Curriculum and Instruction, University of Texas at Austin, Austin, Tex. 78712.

popular lore, scientific writings, fiction, love, humor, and so on.

A third contemporary count of words was done by Carroll, Davies, and Richman (1971) and called the American Heritage Intermediate Corpus. This computer based selection includes over 5 million words (tokens) drawn in 500 word samples from 1045 published materials and contains 86,741 different words (types). Of particular interest in this corpus are the categories for which frequency data are presented as they include, in addition to a count of the usage, a count by grade level and curriculum and library areas. These norms are particularly useful to workers interested in word frequencies by grade level but as of now have not been used to determine orthographic structure. A brief description of these three word frequency counts is shown in Table 1 along with counts of American English letter frequency.

There are many other word counts (e.g., Horn, 1926; Rinsland, 1945) and counts of spoken language (e.g., Wepman & Hass, 1969; Howes, 1966; and Beier, Starkweather & Miller, 1967) which may be of interest to some researchers.

Letter Counts

Historically, frequency counts of a language have been of greater interest to cryptoanalysts than psychologists, and several excellent reviews of those counts and their usage are available (see Kahn, 1967; Pratt, 1942; Smith, 1943). Although letter counts of many different languages have been known for centuries, modern counts of single letter frequency began to appear in the 1930s, and large scale counts in the 1960s (Baddeley, Conrad, & Thompson, 1960).

In some early psychological studies normative data on letter frequency were used as the basis for evaluating subjective judgments about the natural occurrence of events (Attneave, 1953) while in other studies the data were considered an integral part of a larger theory of frequency and

learning (Underwood & Schulz, 1960). A sample of some single letter counts is shown in Table 1.

One of the first comprehensive analyses of letter frequencies based on the Kučera and Francis corpus was done by Zettersten (1969). Solso and King (1976), who also used this corpus, reported in addition to single-letter frequencies (based on tokens), single-letter versatilities (based on types). From these data, it is possible to identify how frequently a specific letter appears and in how many different words a letter appears (versatility).

A count of letters in words that is position specific was also of interest to cryptographers (see Smith, 1943; Milliken, 1943). Recently, however, psychologists interested in the reading process or letter identification have found position-sensitive letter counts to be necessary components of an evaluation of orthography. Mayzner and Tresselt (1965) did the first large scale hand count of single- and multiple-letter positional frequency.

Solso and King (1976) reported positional frequencies and versatilities for four- and five-letter words based on a computer analysis of the Kučera and Francis corpus. This positional count was extended to six-, seven-, and eight-letter words by Solso (1979b). Recently, Solso and Juel (1980) reported positional bigram frequencies and versatilities, and Massaro, Taylor, Venezky, Jastrzembski, and Lucas (1980) reported positional bigram frequencies for words. Positional counts which have been generally used as a measure of orthography have provided researchers with a formidable arsenal with which to attack the problem of reading and lexical memory (cf., Juel & Solso, 1980, 1981; Mason, 1975; Massaro, Venezky, & Taylor, 1979; McClelland & Johnston, 1977). In the present paper, we extend these studies by reporting the frequencies and versatilities of initial and terminal letters of words. Our aim is to provide researchers in the field of verbal processing with a statistical basis for their

TABLE 1
A SELECTION OF MODERN COUNTS OF AMERICAN ENGLISH WORDS AND LETTERS

	Sample size	Sample characteristics
American word frequencies		
Thorndike and Lorge (1944)	About 20 million words	Magazines, books, juvenile books
Kučera and Francis (1967)	1,014,232 words	15 Different categories of reading material (e.g., newspapers, popular books, scientific writing, etc.)
Carroll, Davies, and Richman (1971)	5,088,721 words	School books; grades 3 through 9
Letter counts		
Praet (1942)	500 words	British English; single letter frequency bigrams from newspapers, letters, fiction, and so on.
Smith (1943)	Not reported	American English; single letters, bigrams, trigrams, doubles, initial letters, final letters, and so on.
Attneave (1953)	10,341 words	Half from San Antonio newspaper, half from popular magazines; single letter frequency
Underwood and Schulz (1960)	15,000 words	American novels, magazines, etc.; bigrams, trigrams
Fang (1966)	14,600 words	American newspapers; single letter frequency
Zettersten (1969)	About one million words	From Kučera and Francis; single letter frequency by specific sample of text, and consonant clusters
Solso and King (1976)	983,400 words	From Kučera and Francis; single letter frequency and versatility.
Solso, Barbuto, and Juel (1979)	983,400 words	From Kučera and Francis; bigram and trigram frequency and versatility

Positional frequency and versatility letter counts Smith (1943)	Not reported	Rank order of initial and terminal letters
Mayzner and Tresselt (1965)	20,000 words	From American novels, magazines, and so on single letter and bigram frequency by position for words of 3 to 7 letters in length
Mayzner, Tresselt, and Wolin (1965a, 1965b, 1965c)	20,000 words	Sample same as above; trigrams, tetragrams, and pentagrams frequency by position for words up to 7 letters in length
Solso and King (1976)	262,883 words	From Kučera and Francis; positional frequencies and versatilities for single letters in 4- and 5-letter words
Rubin (1978)	999,464 words	From Kučera and Francis; initial and terminal frequency and versatility of single letters and cluster types
Solso (1979)	213,902 words	From Kučera and Francis; positional frequency and versatility for single letters for 6-, 7-, and 8-letter words
Solso and Juel (1980)	897,435 words	From Kučera and Francis; bigram positional frequency and versatility of 2- through 9-letter words
Massaro, Taylor, Venezky, Jastrzembski, and Lucas, (1980)	612,193 words	From Kučera and Francis; single-letter and bigram positional frequencies for 3- to 7-letter words
Solso, Juel, and Rubin (1982)	983,400 words	From Kučera and Francis; frequency and versatility of initial and terminal letters and bigrams

interpretation of the effects of initial and terminal letters, as well as to provide values necessary to calculate several measures of orthography.

The original counts of initial and terminal letters were done by cryptanalysts (see Gaines, 1956; O'Haver, 1933; and Yardley, 1931) as a methodological tool in decoding secret information. Several modern attempts have been made to count initial and terminal letters (e.g., Zettersten, 1969; Rubin, 1978). The Zettersten frequency count restricted its clusters to those consisting totally of consonants. Rubin's count provides values for both frequency and versatility but provides values for only those clusters with frequencies greater than 30. The present tabulation of initial and terminal letters and bigrams based on all existing words in the Kučera and Francis corpus will provide workers with a comprehensive index of this feature of English orthography which is also compatible with existing norms (e.g., Solso, Barbuto, & Juel, 1979; Solso & King, 1976).

Initial and terminal letters are of special interest to workers interested in reading and memory as they begin or end words, phonemes, syllables, morphemes, and usually a spelling pattern (Fries, 1963; Gibson, Pick, Osser, & Hammond, 1962). Initial letters also provide a salient cue in retrieval of information from memory as attested to by crossword puzzle workers, by psychological research (Brown & McNeill, 1966; Rubin, 1975; Solso & Biersdorff, 1975), and a powerful mnemonic (Morris & Cook, 1978).

The popular use of an initial letter as a mnemonic technique attests to the efficacy of initial letters of words as a retrieval cue. Some of these systems are widely known. The cranial nerves are frequently remembered by the rhyme "On old Olympus' towering top; A Finn and German vault and hop" in which the first O is a cue for the olfactory nerve, the second O a cue for the optic nerve, and so on (see Solso, 1979a, and Young & Gibson, 1974, for other

examples). However, the reason initial letters serve as an efficacious retrieval cue is not well understood, but several hypotheses have been suggested. Some have suggested that the primacy effect may be operating in initial letters. Others hypothesize that initial letter of efficacy may be due to the fact that initial letters are not susceptible to as much visual masking as central letters. The hypothesis we find most intriguing (especially as related to the present norms) is the notion that knowledge of an initial letter allows the reader to predict with increased certainty the next letter(s). Yet, until the present norms were available, it was extremely difficult to accurately measure the information conveyed by a specific initial letter.

The current counts of initial and terminal letters may also be of special interest to researchers who wish to study the word recognition processes of beginning readers. There is abundant evidence that kindergarten and first-grade children rely on individual letters as their primary cue in recognizing words (Marchbanks & Levin, 1965; Samuels & Jeffrey, 1966; Fisher & Price, 1970; Timko, 1972; Williams, Blumberg, & Williams, 1970; Rayner & Hagelberg, 1975; Rayner, 1976). Initial letters, in particular, appear to be the most salient cue these young readers utilize. At some point, however, children must abandon their reliance on first letter cues (i.e., "c" for "cat" fails the child once "cap" is introduced). There is evidence that older children increasingly focus on higher-order units, such as letter clusters, which are formed by orthographic structures and sequential letter redundancies (Lott & Smith, 1970; Mason, 1975; Massaro et al., 1979; Massaro & Taylor, 1979; Juel & Solso, 1980, 1981). In some of these studies, bigram frequency and versatility counts have been employed. With the present norms researchers may find it easier to untangle the beginning and more advanced stages in word recognition processing by devising stimuli words that represent orthogonally various levels of *both*

initial/terminal letter counts and bigram frequencies/versatilities.

METHOD AND RESULTS

All frequency and versatility counts are based on the Kučera and Francis (1967) frequency count of about one million words in the English language. All hyphenated words, words containing apostrophes, and digits were excluded from the analysis. From the original corpus of 50,406 distinct graphic elements (types), 40,181 different words with frequency of 983,400 occurrences (tokens) were used. Single words were identified, and the frequencies and versatilities of the first and last letter of each word were determined.

Initial and terminal frequencies and versatilities for single letters are contained in Table 2. Since some single letters also appear in the corpus as types, their influence may distort the distribution of initial and terminal values. Therefore, an additional count is provided in which single-letter types are removed.

A second count was undertaken in which initial and terminal bigram frequencies and versatilities are reported. These counts and the percentage of occurrence are shown in Table 3. The percentage of occurrence is based on total frequency of occurrence of the initial or terminal letters. Thus, the percentage occurrence of the initial bigram CR is 4.68%. That percentage was found by finding the ratio between the frequency of all CR words (2204) and the frequency of all C words (47,141) longer than single-letter types. The same procedure was applied to the percentages of versatility counts. The data for any specific terminal bigram may be found by reversing the notations. Thus the occurrence of words ending in ED may be found by finding DE in Table 3 ($F = 34.4\%$, $V = 30.0\%$). This procedure was done so the total occurrence of any specific letter (in the example, *E*) would be grouped together. To find the percentage of occurrence for any series of letters a base of $F = 953,443$ and $V = 40,156$ could be used.

DISCUSSION

One advantage of the present count is that it can be used in conjunction with previously published counts (Solso et al., 1979; Solso & King, 1976) to calculate orders of approximation to English (Miller & Selfridge, 1950; Rubin, 1980), or Englishness (Travers & Olivier, 1978) for any word. For this purpose, the initial and terminal letters listed in Table 2 can be viewed as representing bigrams followed by or following a blank. Similarly, the initial and terminal bigrams listed in Table 3 can be viewed as trigrams beginning or ending in blank.

By using the enclosed norms, it is possible to quantify the amount of information contained in initial letters. It is presumed that the word is encoded in a number of attributes which, in addition to semantic and acoustic attributes, include orthographic attributes. These orthographic attributes contain knowledge of letter sequencing.

We would like to introduce a measure of approximation to English words in which the component of positional orthography is maintained. Consider the word *chair*. In zero-order approximation to English, all 27 characters, the 26 letters plus *blank*, are equally likely, and the word *chair* would have a probability of being created equal to drawing a blank, *c*, *h*, *a*, *i*, *r*, and *blank*, or $(1/27)^7$. For first-order approximation, the $1/27$ values would be replaced by the actual probabilities of the individual characters calculated from Solso and King (1976) (e.g., $c = .031 \times h = .055 \times a = .076 \times i = .071 \times n = .062$ or about .00000056).

For second-order approximation, the probabilities of drawing the following items would be multiplied: *blank*, *c* given *blank*, *h* given *c*, *a* given *h*, *i* given *a*, *r* given *i*, and *blank* given *r*. For third-order approximation, the probabilities of drawing the following items would be multiplied: *blank*, *c* given *blank*, *h* given *blank*, *c*, *a* given *ch*, *i* given *ha*, *r* given *ai*, and *blank* given *ir*. The values of second- and third-order approxi-

TABLE 2
INITIAL AND TERMINAL FREQUENCIES AND VERSATILITIES FOR SINGLE LETTERS

	ALL WORDS					WORDS WITH SINGLE LETTER REMOVED			
	INITIAL		TERMINAL			INITIAL		TERMINAL	
	F	V	F	V		F	V	F	V
A	114948	2556	28427	956	A	91711	2555	5190	955
B	45646	2447	1010	102	B	45529	2446	893	101
C	47271	3850	5030	601	C	47141	3849	4900	600
D	28952	2379	103926	4780	D	28862	2378	103836	4779
E	24192	1673	201016	4939	E	24091	1672	200915	4938
F	40418	1695	42069	152	F	40348	1694	41999	151
G	16805	1350	30851	3112	G	16755	1349	30801	3111
H	53557	1512	25959	719	H	53483	1511	25885	718
I	67085	1660	6048	347	I	61892	1659	875	346
J	5267	414	135	6	J	5143	413	11	5
K	4854	468	9052	521	K	4834	467	9032	520
L	23135	1335	31201	1550	L	23080	1334	31146	1549
M	38843	2161	16506	696	M	38759	2160	16422	695
N	21283	816	87253	3265	N	21242	815	87212	3264
O	71304	935	42308	537	O	71249	934	42253	536
P	38608	3074	6014	308	P	38525	3073	5931	307
Q	1915	163	13	6	Q	1913	162	11	5
R	25835	2274	58716	2111	R	25769	2273	58650	2110
S	68101	4505	122784	8867	S	67966	4504	122649	8866
T	159006	1991	91760	2433	T	158974	1990	91728	2432
U	11405	953	3756	100	U	11314	952	3665	99
V	6426	649	272	50	V	6395	648	241	49
W	60251	1084	9125	180	W	60167	1083	9041	179
X	17	9	1253	110	X	16	8	1252	109
Y	8085	154	58634	3642	Y	8074	153	58623	3647
Z	211	74	282	85	Z	211	74	282	85
TOTAL	983400	40181	983400	40181		953443	40156	953443	40156

mations can be calculated from Solso et al (1979) and Tables 2 and 3. In order to convert the frequency values given in these tables and papers to probabilities, one needs to divide by the number of occurrences. The total number of characters, bigrams, and trigrams in the combined sample of letters plus blanks is 5.6×10^6 , 5.6×10^6 , and 4.6×10^6 , respectively. The frequency of a *blank* is 983,400. In order to convert the product of the probabilities to an information measure, the logarithm to the base 2 is taken. Travers and Olivier (1978) provide more details on the calculation of a similar measure. Rubin (1980) provides some applications of these measures.

From the positional frequency norms collected by Solso and his colleagues, the occurrence of letters by position may be determined. In the case of *c-h-a-i-r*, we can calculate the probability of *c* in the first position of five-letter words by dividing the frequency of *c* in the first position of five-letter words (6733, see Solso & King,

1976) by the total occurrence of letters in the first position of five-letter words (107,302). That value is .063. Using the same procedure for *h* in the second position, and *a*, *i*, *r*, in the third, fourth, and fifth positions, respectively, we find $18,960/107,302$, $12,769/107,302$, $7200/107,302$, and $12,029/107,302$, which when multiplied times each other is equal to $.063 \times .177 \times .119 \times .067 \times .112 = .000099$. A similar procedure could be used to calculate positional versatilities. An even more sophisticated measure of approximations to English orthography could be calculated by the same procedure using frequency or versatility norms for positional bigrams (see Solso & Juel, 1980).

In order to illustrate the use of the norms in memory research, we will use an example. Suppose you are trying to recall from memory the word *prank*. The probability a random selection of a letter will be the same as the initial letter of any given word, such as *prank*, would be $1/26$, or about a 3.85%

“chance” of identifying the letter *p*. The amount of “information” conveyed by the correct letter, by this means, is the reciprocal of these estimates, *viz.*, 25/26 letters or 96.15% of all possible initial letters are excluded. An initial letter serves as a powerful retrieval cue because so many letter alternatives are excluded.

The above analysis lacks the precision of estimating letter information available in the present analysis. Given the same problem (retrieve *prank*), it is possible to estimate the frequency of any specific letter based on total letter frequency as well as an estimate of the versatility of any letter (see Solso & King, 1976). These norms indicate that the letter *p* occurs about 2.03% of usage (frequency) and in about 2.97% of words (versatility). Although the frequency measure may be a useful index of the number of times a certain letter or letter combination is experienced (which may to some degree affect letter availability), versatility may prove to be more closely related to word retrieval as it is from a corpus of words in memory that specific words are retrieved.

An even more accurate measure of the expected frequency and versatility of any letter can be derived from an analysis of letter occurrence in which positional considerations are maintained. The positional frequency and versatility of each letter can be determined in four- and five-letter words (Solso & King, 1976), and six-, seven-, and eight-letter words (Solso, 1979b). In the case of *prank*, for example, we find that the frequency of the letter *p* in the first position of five-letter words is 4687 out of a total occurrence of initial letters in five-letter words of 107,302. Using the same norms, we can determine the versatility of *p*. That letter was the initial letter in 242 five-letter words out of a total of 4066 five-letter words. By calculating the ratios (4687/107,302 and 242/4066), we find that the frequency of *p* is about 4.37% and its versatility about 5.95%. Similar operations can be performed for any letter in any position. From these calculations it is possible to determine the information load of each let-

ter with regard to its positional frequency and versatility.

Another useful index may be to determine the expected frequency of second or later letters in any given letter sequence. One estimate would be based on an assumption of uniform distribution of letters: that estimate would be 1/26 or 3.85% of each letter. The fallacy of that estimate is obvious as letters are not uniformly distributed in English words. Another technique is similar to the above-mentioned one in which the occurrence of the second letter is divided by the total occurrence. In the word *prank*, the occurrence of *r* in the same position is 8173 of 107,302 occurrences and 368 of 4066 different words of 7.62% (frequency) and 9.05% (versatility).

An even more sophisticated measure of positional frequencies and versatilities is based on the probability of the occurrence of a letter given one or more letters in a word. In the example of *prank*, we can estimate the probability of the *r* given *p* from the norms presented in this paper by the following means. First find the total frequency and versatility of the initial letter *p* (Table 2). These values are 38.525 and 3073. Then find frequency and versatility of *pr* (Table 3). These values are 11,849 and 809. The estimated probability of *r* given the initial letter *p* among *all* English words is equal to the ratio between the total of *pr* and the totals for *p* or 11,849/38,525 and 809/3073, which yields about 31 and 26% for frequency and versatility. These values are for *all* words, not only five-letter words, and this fact should be borne in mind when using the ratios. However, another estimate of the initial bigram *pr* in five-letter words may be obtained from norms collected by Solso and Juel (1979). The procedure for finding this estimate is to (1) sum all initial bigrams starting with *p* for five-letter words; (2) divide the occurrence of *pr* by the sum found in the first operation. In the case of *pr*, the sum of the frequency and versatility of all five-letter words with an initial letter of *p* is 4687 and 242. The frequency and versatility of *pr* is 631 and 26. Performing the above operation, we find a

TABLE 3
INITIAL AND TERMINAL BIGRAM FREQUENCIES AND VERSATILITIES

	INITIAL		TERMINAL	
	F	V	F	V
AA	9(0.01)	3(0.12)	0(0.00)	0(0.00)
AB	3198(3.49)	151(5.91)	58(0.06)	14(0.55)
AC	3688(4.02)	224(8.77)	69(0.08)	27(1.06)
AD	2528(2.76)	229(8.96)	7463(8.20)	95(3.74)
AE	113(0.12)	26(1.02)	61(0.07)	22(0.87)
AF	1794(1.96)	76(2.97)	38(0.04)	10(0.39)
AG	2452(2.67)	80(3.13)	156(0.17)	24(0.95)
AH	140(0.15)	9(0.35)	149(0.16)	45(1.77)
AI	739(0.81)	55(2.15)	26(0.03)	18(0.71)
AJ	2(0.00)	1(0.04)	2(0.00)	2(0.08)
AK	12(0.01)	3(0.12)	344(0.38)	41(1.62)
AL	7709(8.41)	274(10.72)	13954(15.33)	810(31.93)
AM	2207(2.41)	124(4.85)	1550(1.70)	105(4.14)
AN	37525(40.92)	313(12.25)	13810(15.17)	596(23.49)
AO	4(0.00)	2(0.08)	35(0.04)	10(0.39)
AP	2346(2.56)	176(6.89)	260(0.29)	32(1.26)
AQ	22(0.02)	7(0.27)	3(0.00)	1(0.04)
AR	7915(8.63)	264(10.33)	4643(5.10)	200(7.88)
AS	9667(10.54)	200(7.83)	20762(22.81)	156(6.15)
AT	7051(7.69)	117(4.58)	20009(21.98)	98(3.86)
AU	1035(1.13)	118(4.62)	115(0.13)	26(1.02)
AV	669(0.73)	49(1.92)	20(0.02)	6(0.24)
AW	799(0.87)	32(1.25)	853(0.94)	34(1.34)
AX	73(0.08)	12(0.47)	289(0.32)	17(0.67)
AY	7(0.01)	6(0.23)	6367(6.99)	147(5.79)
AZ	7(0.01)	4(0.16)	2(0.00)	1(0.04)
BA	4481(4.84)	466(19.05)	77(0.60)	17(9.29)
BB	0(0.00)	0(0.00)	31(0.24)	10(5.46)
BC	1(0.00)	1(0.04)	1(0.01)	1(0.55)
BE	18118(39.79)	459(18.77)	6683(52.00)	38(20.77)
BH	1(0.00)	1(0.04)	0(0.00)	0(0.00)
BI	1614(3.54)	165(6.75)	27(0.21)	7(3.83)
BL	1552(3.41)	236(9.65)	0(0.00)	0(0.00)
BN	0(0.00)	0(0.00)	1(0.01)	1(0.55)
BO	4166(9.15)	379(15.49)	17(0.13)	10(5.46)
BR	3309(7.27)	389(15.90)	1(0.01)	1(0.55)
BS	0(0.00)	0(0.00)	280(2.18)	56(30.60)
BT	0(0.00)	0(0.00)	127(0.99)	2(1.09)
BU	6927(15.21)	329(13.45)	5(0.04)	1(0.55)
BY	5360(11.77)	21(0.86)	5601(43.58)	39(21.31)
CA	9060(19.22)	674(17.51)	308(1.07)	42(3.28)
CC	2(0.00)	1(0.03)	2(0.01)	1(0.08)
CE	1989(4.22)	132(3.43)	10858(37.89)	451(35.26)
CF	8(0.02)	1(0.03)	8(0.03)	1(0.08)
CH	5748(12.19)	480(12.47)	9192(32.07)	193(15.09)
CI	1302(2.76)	95(2.47)	9(0.03)	7(0.55)
CK	0(0.00)	0(0.00)	3269(11.41)	224(17.51)
CL	3235(6.86)	287(7.46)	0(0.00)	0(0.00)
CM	22(0.05)	2(0.05)	21(0.07)	1(0.08)
CO	21963(46.59)	1653(42.95)	258(0.89)	44(3.44)
CP	2(0.00)	1(0.03)	0(0.00)	0(0.00)
CR	2204(4.68)	324(8.42)	0(0.00)	0(0.00)
CS	0(0.00)	0(0.00)	669(2.33)	100(7.82)
CT	1(0.00)	1(0.03)	3063(10.69)	105(8.21)
CU	1469(3.12)	166(4.31)	2(0.01)	2(0.16)
CY	126(0.27)	27(0.70)	999(3.49)	106(8.29)
CZ	10(0.02)	5(0.13)	2(0.01)	2(0.16)
DA	3047(10.56)	185(7.78)	262(2.06)	50(6.10)
DD	0(0.00)	0(0.00)	141(1.11)	6(0.73)
DE	9136(31.65)	814(34.23)	4362(34.36)	246(30.00)
DG	0(0.00)	0(0.00)	3(0.02)	1(0.12)
DH	1(0.00)	1(0.04)	0(0.00)	0(0.00)
DI	7324(25.38)	751(31.58)	66(0.52)	16(1.95)
DJ	7(0.02)	3(0.13)	0(0.00)	0(0.00)
DL	0(0.00)	0(0.00)	3(0.02)	1(0.12)
DN	1(0.00)	1(0.04)	5(0.04)	3(0.37)
DO	5395(18.69)	251(10.56)	1492(11.75)	50(6.10)
DR	2386(8.27)	205(8.62)	193(1.52)	2(0.24)
DS	0(0.00)	0(0.00)	4106(32.34)	335(40.85)
DT	0(0.00)	0(0.00)	27(0.21)	14(1.71)
DU	1370(4.75)	122(5.13)	77(0.61)	2(0.24)
DV	1(0.00)	1(0.04)	0(0.00)	0(0.00)
DW	65(0.23)	17(0.71)	1(0.01)	1(0.12)
DY	129(0.45)	27(1.14)	1958(15.42)	93(11.34)

TABLE 3—Continued

	F	INITIAL	V	F	TERMINAL	V		
EA	2551(10.59)	67(4.01)	726(0.56)	32(0.35)
EB	13(0.05)	8(0.48)	33(0.03)	7(0.08)
EC	517(2.15)	40(2.39)	83(0.06)	14(0.15)
ED	842(3.50)	55(3.29)	40585(31.29)	3872(41.76)
EE	6(0.02)	3(0.18)	2517(1.94)	106(1.14)
EF	946(3.93)	31(1.85)	421(0.32)	17(0.18)
EG	99(0.41)	19(1.14)	112(0.09)	12(0.13)
EH	9(0.04)	2(0.12)	18(0.01)	9(0.10)
EI	567(2.35)	20(1.20)	63(0.05)	16(0.17)
EJ	8(0.03)	4(0.24)	5(0.00)	1(0.01)
EK	31(0.13)	5(0.30)	488(0.38)	27(0.29)
EL	1510(6.27)	161(9.63)	1854(1.43)	189(2.04)
EM	1184(4.91)	138(8.25)	3029(2.34)	36(0.39)
EN	3958(16.43)	320(19.14)	14625(11.28)	427(4.61)
EO	1(0.00)	1(0.06)	33(0.03)	10(0.11)
EP	145(0.60)	46(2.75)	735(0.57)	27(0.29)
EQ	555(2.30)	40(2.39)	3(0.00)	2(0.02)
ER	302(1.25)	59(3.53)	28492(21.97)	1490(16.07)
ES	1012(4.26)	80(4.78)	22293(17.19)	2435(26.26)
ET	238(0.99)	29(1.73)	4712(3.63)	226(2.44)
EU	247(1.03)	22(1.32)	15(0.01)	8(0.09)
EV	3446(14.30)	90(5.38)	116(0.09)	15(0.16)
EW	2(0.01)	2(0.12)	3319(2.56)	43(0.46)
EX	5327(22.11)	408(24.40)	378(0.29)	29(0.31)
EY	573(2.38)	21(1.26)	4987(3.85)	198(2.14)
EZ	2(0.01)	1(0.06)	47(0.04)	23(0.25)
FA	5057(12.53)	289(17.06)	9(0.26)	3(1.40)
FB	0(0.00)	0(0.00)	1(0.03)	1(0.47)
FC	3452(8.56)	186(10.98)	1175(34.40)	30(14.02)
FD	2(0.00)	2(0.12)	1029(30.12)	61(28.50)
FE	0(0.00)	0(0.00)	1(0.03)	1(0.47)
FF	6281(15.57)	252(14.88)	1(0.03)	1(0.47)
FG	2(0.00)	1(0.06)	0(0.00)	0(0.00)
FH	1538(3.81)	260(15.35)	0(0.00)	0(0.00)
FI	14961(37.08)	318(18.77)	0(0.00)	0(0.00)
FJ	7248(17.96)	261(15.41)	0(0.00)	0(0.00)
FK	0(0.00)	0(0.00)	93(2.72)	29(13.55)
FL	14(0.03)	1(0.06)	895(26.20)	39(18.22)
FM	1791(4.44)	122(7.20)	1(0.03)	1(0.47)
FN	2(0.00)	2(0.12)	211(6.18)	48(22.43)
FO	1723(10.31)	255(18.90)	75(0.65)	21(2.92)
FP	0(0.00)	0(0.00)	1(0.01)	1(0.14)
FQ	2765(16.50)	156(11.56)	4502(39.27)	249(34.58)
FR	0(0.00)	0(0.00)	41(0.36)	5(0.69)
FS	58(0.35)	19(1.41)	2922(25.49)	46(6.39)
FT	1667(9.95)	94(6.97)	10(0.09)	8(1.11)
FU	717(4.28)	138(10.23)	0(0.00)	0(0.00)
FV	12(0.07)	1(0.07)	25(0.22)	3(0.42)
FW	13(0.08)	9(0.67)	515(4.49)	14(1.94)
FX	4275(25.51)	147(10.90)	1105(9.64)	33(4.58)
FY	4433(26.46)	374(27.72)	1(0.01)	1(0.14)
FZ	0(0.00)	0(0.00)	1827(15.94)	267(37.08)
GA	1008(6.02)	135(10.01)	0(0.00)	0(0.00)
GB	1(0.01)	1(0.07)	0(0.00)	0(0.00)
GC	78(0.47)	20(1.48)	441(3.85)	72(10.00)
GD	16013(29.94)	436(28.86)	48(0.05)	20(5.28)
GE	0(0.00)	0(0.00)	1(0.00)	1(0.26)
GF	17521(32.76)	333(22.04)	82527(90.61)	51(13.46)
GG	0(0.00)	0(0.00)	1(0.00)	1(0.26)
GH	0(0.00)	0(0.00)	3(0.00)	2(0.53)
GI	12288(22.98)	152(10.06)	44(0.05)	17(4.49)
GJ	0(0.00)	0(0.00)	1(0.00)	1(0.26)
GK	0(0.00)	0(0.00)	28(0.03)	5(1.32)
GL	2(0.00)	2(0.13)	58(0.06)	8(2.11)
GM	0(0.00)	0(0.00)	389(0.43)	16(4.22)
GN	5820(10.88)	352(23.30)	2288(2.51)	12(3.17)
GO	7(0.01)	2(0.13)	18(0.02)	7(1.85)
GP	0(0.00)	0(0.00)	418(0.46)	62(16.36)
GQ	0(0.00)	0(0.00)	4359(4.79)	95(25.07)
GR	1558(2.91)	152(10.06)	28(0.03)	5(1.32)
GS	4(0.01)	1(0.07)	0(0.00)	0(0.00)
GT	270(0.56)	81(5.36)	872(0.96)	76(20.05)
H	0(0.00)	0(0.00)	977(1.17)	175(8.16)
HA	7(0.01)	4(0.24)	11(0.01)	6(0.28)

TABLE 3—Continued

	INITIAL		TERMINAL	
	F	V	F	V
IC	70(0.11)	10(0.60)	4548(5.46)	536(24.99)
ID	784(1.27)	54(3.25)	4247(5.10)	104(4.85)
IE	3(0.00)	1(0.06)	837(1.00)	153(7.13)
IF	2200(3.55)	2(0.12)	2229(2.67)	6(0.28)
IG	104(0.17)	15(0.90)	504(0.60)	23(1.07)
IH	1(0.00)	1(0.06)	1(0.00)	1(0.05)
II	1(0.00)	1(0.06)	30(0.04)	8(0.37)
IJ	1(0.00)	1(0.06)	4(0.00)	2(0.09)
IK	5(0.01)	2(0.12)	30(0.04)	16(0.75)
IL	359(0.58)	43(2.59)	1627(1.95)	97(4.52)
IM	2305(3.72)	243(14.65)	3033(3.64)	43(2.00)
IN	33686(54.43)	1097(66.12)	25770(30.92)	423(19.72)
IO	87(0.14)	22(1.33)	330(0.40)	57(2.66)
IP	1(0.00)	1(0.06)	1001(1.20)	89(4.15)
IR	334(0.54)	81(4.88)	3537(4.24)	42(1.96)
IS	10745(17.36)	50(3.01)	23636(28.36)	213(9.93)
IT	11162(18.03)	24(1.45)	10653(12.78)	122(5.69)
IU	1(0.00)	1(0.06)	0(0.00)	0(0.00)
IV	32(0.05)	4(0.24)	2(0.00)	2(0.09)
IX	0(0.00)	0(0.00)	314(0.38)	17(0.79)
IZ	4(0.01)	2(0.12)	20(0.02)	10(0.47)
JA	842(16.37)	101(24.46)	7(6.03)	1(7.14)
JE	583(11.34)	82(19.85)	4(3.45)	2(14.29)
JJ	83(1.61)	21(5.08)	12(10.34)	5(35.71)
JO	1524(29.63)	101(24.46)	15(12.93)	3(21.43)
JR	74(1.44)	1(0.24)	74(63.79)	1(7.14)
JU	2037(39.61)	107(25.91)	4(3.45)	2(14.29)
KA	267(5.52)	75(16.06)	91(1.58)	32(6.84)
KC	2(0.04)	1(0.21)	2(0.03)	1(0.21)
KE	1073(22.20)	91(19.49)	3802(66.19)	138(29.49)
KH	81(1.68)	10(2.14)	6(0.10)	2(0.43)
KI	1189(24.60)	115(24.63)	57(0.99)	21(4.49)
KL	21(0.43)	15(3.21)	0(0.00)	0(0.00)
KM	1(0.02)	1(0.21)	1(0.02)	1(0.21)
KN	1961(40.57)	58(12.42)	0(0.00)	0(0.00)
KO	140(2.90)	55(11.78)	6(0.10)	6(1.28)
KR	66(1.37)	29(6.21)	0(0.00)	0(0.00)
KS	0(0.00)	0(0.00)	1499(26.10)	191(40.81)
KT	0(0.00)	0(0.00)	1(0.02)	1(0.21)
KU	15(0.31)	9(1.93)	8(0.14)	4(0.85)
KW	6(0.12)	4(0.86)	0(0.00)	0(0.00)
KY	12(0.25)	4(0.86)	271(4.72)	71(15.17)
LA	4961(21.49)	320(23.99)	359(0.66)	78(2.31)
LB	23(0.10)	2(0.15)	26(0.05)	4(0.12)
LD	0(0.00)	0(0.00)	8915(16.32)	117(3.47)
LE	5060(21.92)	270(20.24)	10963(20.06)	886(26.24)
LF	0(0.00)	0(0.00)	1682(3.08)	23(0.68)
LH	0(0.00)	0(0.00)	1(0.00)	1(0.03)
LI	7109(30.80)	301(22.56)	90(0.16)	49(1.45)
LK	0(0.00)	0(0.00)	400(0.73)	17(0.50)
LL	11(0.05)	2(0.15)	10926(20.00)	225(6.66)
LM	0(0.00)	0(0.00)	199(0.36)	15(0.44)
LN	0(0.00)	0(0.00)	49(0.09)	3(0.09)
LO	5189(22.48)	280(20.99)	158(0.29)	42(1.24)
LP	1(0.00)	1(0.07)	327(0.60)	7(0.21)
LS	0(0.00)	0(0.00)	2939(5.38)	351(10.40)
LT	13(0.06)	2(0.15)	1237(2.26)	48(1.42)
LU	601(2.60)	131(9.82)	19(0.03)	7(0.21)
LY	112(0.49)	25(1.87)	16348(29.92)	1502(44.49)
LZ	0(0.00)	0(0.00)	1(0.00)	1(0.03)
MA	12575(32.44)	714(33.06)	369(2.40)	71(13.00)
MB	0(0.00)	0(0.00)	118(0.77)	17(3.11)
MC	3(0.01)	3(0.14)	1(0.01)	1(0.18)
MD	5(0.01)	1(0.05)	7(0.05)	3(0.55)
ME	6781(17.50)	375(17.36)	9280(60.28)	144(26.37)
MF	1(0.00)	1(0.05)	0(0.00)	0(0.00)
MG	17(0.04)	1(0.05)	17(0.11)	1(0.18)
MI	4692(12.11)	391(18.10)	66(0.43)	15(2.75)
ML	11(0.03)	1(0.05)	11(0.07)	1(0.18)
MM	20(0.05)	4(0.19)	60(0.39)	11(2.01)
MN	0(0.00)	0(0.00)	162(1.05)	8(1.47)
MO	8449(21.80)	425(19.68)	33(0.21)	19(3.48)
MP	1(0.00)	1(0.05)	238(1.55)	37(6.78)
MR	1373(3.54)	2(0.09)	839(5.45)	1(0.18)

TABLE 3—Continued

	F	INITIAL	V	F	TERMINAL	V		
MS	1(0.00)	1(0.05)	2378(15.45)	166(30.40)
MT	6(0.02)	2(0.09)	9(0.06)	5(0.92)
MU	3155(8.14)	195(9.03)	1(0.01)	1(0.18)
MY	1669(4.31)	43(1.99)	1807(11.74)	45(8.24)
NA	2300(10.83)	154(18.90)	506(0.49)	114(1.99)
NC	0(0.00)	0(0.00)	35(0.03)	5(0.09)
ND	1(0.00)	1(0.12)	36314(35.51)	241(4.21)
NE	5943(27.98)	219(26.87)	7916(7.74)	417(7.28)
NF	0(0.00)	0(0.00)	1(0.00)	1(0.02)
NG	2(0.01)	2(0.25)	29538(28.89)	2961(51.69)
NH	0(0.00)	0(0.00)	3(0.00)	3(0.05)
NI	986(4.64)	98(12.02)	95(0.09)	52(0.91)
NJ	1(0.00)	1(0.12)	0(0.00)	0(0.00)
NK	2(0.01)	1(0.12)	1081(1.06)	56(0.98)
NN	0(0.00)	0(0.00)	131(0.13)	35(0.61)
NC	10944(51.52)	258(31.66)	2323(2.27)	42(0.73)
NQ	0(0.00)	0(0.00)	1(0.00)	1(0.02)
NS	0(0.00)	0(0.00)	7843(7.67)	968(16.90)
NT	0(0.00)	0(0.00)	13001(12.71)	724(12.64)
NU	1050(4.94)	74(9.08)	14(0.01)	5(0.09)
NW	2(0.01)	1(0.12)	2(0.00)	1(0.02)
NX	0(0.00)	0(0.00)	16(0.02)	4(0.07)
NY	11(0.05)	6(0.74)	3433(3.36)	94(1.64)
NZ	0(0.00)	0(0.00)	6(0.01)	4(0.07)
OA	50(0.07)	15(1.61)	25(0.02)	8(0.30)
OB	1174(1.65)	96(10.28)	327(0.30)	17(0.64)
OC	647(0.91)	53(5.67)	42(0.04)	7(0.27)
OD	112(0.16)	14(1.50)	2519(2.30)	90(3.41)
OE	37(0.05)	3(0.32)	143(0.13)	26(0.99)
OF	38376(53.86)	50(5.35)	36561(33.45)	18(0.68)
OG	7(0.01)	4(0.43)	141(0.13)	19(0.72)
OH	158(0.22)	3(0.32)	124(0.11)	5(0.19)
OI	137(0.19)	12(1.28)	5(0.00)	5(0.19)
OK	42(0.06)	7(0.75)	1241(1.14)	47(1.78)
OL	845(1.19)	37(3.96)	1244(1.14)	57(2.16)
OM	56(0.08)	17(1.82)	5669(5.19)	81(3.07)
ON	12567(17.64)	33(3.53)	26329(24.09)	1532(58.05)
OO	6(0.01)	4(0.43)	864(0.79)	18(0.68)
OP	1864(2.62)	72(7.71)	774(0.71)	55(2.08)
OR	5915(8.30)	154(16.49)	17766(16.25)	280(10.61)
OS	66(0.09)	29(3.10)	348(0.32)	101(3.83)
OT	2139(3.00)	14(1.50)	6251(5.72)	86(3.26)
OU	4237(5.95)	120(12.85)	3322(3.04)	9(0.34)
OV	1681(2.36)	151(16.17)	94(0.09)	23(0.87)
OW	982(1.38)	19(2.03)	4862(4.45)	97(3.68)
OX	118(0.17)	15(1.61)	159(0.15)	17(0.64)
OY	20(0.03)	5(0.54)	499(0.46)	37(1.40)
OZ	13(0.02)	7(0.75)	6(0.01)	4(0.15)
PA	6633(17.22)	586(19.07)	86(2.10)	10(2.70)
PE	5053(13.12)	428(13.93)	1049(25.61)	61(16.44)
PF	10(0.03)	5(0.16)	8(0.20)	3(0.81)
PH	1143(2.97)	143(4.65)	205(5.00)	25(6.74)
PI	1685(4.37)	221(7.19)	51(1.25)	6(1.62)
PL	3543(9.20)	209(6.80)	2(0.05)	1(0.27)
PM	0(0.00)	0(0.00)	5(0.12)	1(0.27)
PN	3(0.01)	1(0.03)	0(0.00)	0(0.00)
PO	5802(15.06)	403(13.11)	21(0.51)	9(2.43)
PP	9(0.02)	1(0.03)	16(0.39)	6(1.62)
PR	11849(30.76)	809(26.33)	0(0.00)	0(0.00)
PS	173(0.45)	37(1.20)	1440(35.16)	164(44.20)
PT	23(0.06)	5(0.16)	940(22.95)	48(12.94)
PU	2560(6.65)	207(6.74)	0(0.00)	0(0.00)
PV	1(0.00)	1(0.03)	0(0.00)	0(0.00)
PY	38(0.10)	17(0.55)	273(6.67)	37(9.97)
QN	0(0.00)	0(0.00)	3(60.00)	1(50.00)
QU	1913(100.00)	162(100.00)	0(0.00)	0(0.00)
QW	0(0.00)	0(0.00)	2(40.00)	1(50.00)
RA	3277(12.72)	319(14.03)	469(0.92)	90(3.33)
RB	0(0.00)	0(0.00)	84(0.16)	11(0.41)
RC	0(0.00)	0(0.00)	45(0.09)	3(0.11)
RD	3(0.01)	1(0.04)	3346(6.56)	212(7.85)
RE	15851(61.51)	1301(57.24)	21453(42.09)	378(14.01)
RF	1(0.00)	1(0.04)	18(0.04)	8(0.30)
RG	0(0.00)	0(0.00)	143(0.28)	44(1.63)

TABLE 3—Continued

	INITIAL		TERMINAL	
	F	V	F	V
RH	200(0.78)	27(1.19)	2(0.00)	1(0.04)
RI	2154(8.36)	176(7.74)	95(0.19)	32(1.19)
RK	0(0.00)	0(0.00)	1757(3.45)	63(2.33)
RL	0(0.00)	0(0.00)	306(0.60)	13(0.48)
RM	0(0.00)	0(0.00)	1220(2.39)	41(1.52)
RN	0(0.00)	0(0.00)	1830(3.59)	86(3.19)
RO	2875(11.16)	288(12.67)	349(0.68)	43(1.59)
RP	5(0.02)	1(0.04)	100(0.20)	8(0.30)
RR	0(0.00)	0(0.00)	37(0.07)	10(0.37)
RS	0(0.00)	0(0.00)	9064(17.78)	1048(38.83)
RT	2(0.01)	1(0.04)	3249(6.37)	143(5.30)
RU	1367(5.30)	148(6.51)	21(0.04)	5(0.19)
RV	0(0.00)	0(0.00)	9(0.02)	3(0.11)
RX	0(0.00)	0(0.00)	8(0.02)	1(0.04)
RY	34(0.13)	10(0.44)	7361(14.44)	454(16.82)
RZ	0(0.00)	0(0.00)	3(0.01)	2(0.07)
SA	6399(9.41)	368(8.17)	92(0.27)	33(1.68)
SC	2593(3.82)	318(7.06)	6(0.02)	1(0.05)
SE	9163(13.48)	445(9.88)	10294(29.73)	331(16.84)
SF	2(0.00)	2(0.04)	0(0.00)	0(0.00)
SH	8193(12.05)	419(9.30)	1630(4.71)	192(9.77)
SI	4698(6.91)	281(6.24)	16(0.05)	15(0.76)
SK	567(0.83)	95(2.11)	390(1.13)	19(0.97)
SL	1234(1.82)	165(3.66)	1(0.00)	1(0.05)
SM	1199(1.76)	75(1.67)	731(2.11)	191(9.72)
SN	361(0.53)	90(2.00)	3(0.01)	2(0.10)
SO	8831(12.99)	330(7.33)	3124(9.02)	25(1.27)
SP	3634(5.35)	356(7.90)	41(0.12)	9(0.46)
SQ	312(0.46)	53(1.18)	4(0.01)	1(0.05)
SA	4(0.01)	1(0.02)	4(0.01)	1(0.05)
SS	1(0.00)	1(0.02)	5399(15.59)	570(29.01)
ST	11908(17.52)	751(16.67)	12481(36.04)	514(26.16)
SU	6965(10.25)	519(11.52)	12(0.03)	7(0.36)
SV	3(0.00)	3(0.07)	0(0.00)	0(0.00)
SW	813(1.20)	130(2.89)	1(0.00)	1(0.05)
SY	1070(1.57)	99(2.20)	399(1.15)	51(2.60)
SZ	16(0.02)	3(0.07)	1(0.00)	1(0.05)
TA	3252(2.05)	257(12.91)	506(0.78)	77(2.98)
TC	3(0.00)	2(0.10)	58(0.09)	1(0.04)
TD	0(0.00)	0(0.00)	5(0.01)	1(0.04)
TE	3799(2.39)	301(15.13)	7027(10.82)	597(23.14)
TH	109651(68.97)	351(17.64)	11601(17.87)	177(6.86)
TI	2841(1.79)	164(8.24)	66(0.10)	33(1.28)
TJ	1(0.00)	1(0.05)	0(0.00)	0(0.00)
TL	0(0.00)	0(0.00)	1(0.00)	1(0.04)
TO	31054(19.53)	244(12.26)	28280(43.55)	74(2.87)
TR	4936(3.10)	466(23.42)	0(0.00)	0(0.00)
TS	35(0.02)	9(0.45)	10739(16.54)	976(37.83)
TT	0(0.00)	0(0.00)	196(0.30)	72(2.79)
TU	1144(0.72)	117(5.88)	10(0.02)	7(0.27)
TV	1(0.00)	1(0.05)	0(0.00)	0(0.00)
TW	1784(1.12)	41(2.06)	0(0.00)	0(0.00)
TY	473(0.30)	36(1.81)	6377(9.82)	532(20.62)
TZ	0(0.00)	0(0.00)	70(0.11)	32(1.24)
UA	0(0.00)	0(0.00)	14(0.06)	9(0.74)
UB	2(0.02)	1(0.11)	205(0.83)	15(1.23)
UC	0(0.00)	0(0.00)	7(0.03)	1(0.08)
UD	7(0.06)	2(0.21)	204(0.82)	26(2.13)
UE	0(0.00)	0(0.00)	1780(7.17)	104(8.53)
UF	0(0.00)	0(0.00)	3(0.01)	2(0.16)
UG	31(0.27)	4(0.42)	145(0.58)	20(1.64)
UH	11(0.10)	3(0.32)	26(0.10)	15(1.23)
UI	0(0.00)	0(0.00)	7(0.03)	6(0.49)
UK	4(0.04)	2(0.21)	11(0.04)	5(0.41)
UL	148(1.31)	19(2.00)	1105(4.45)	123(10.09)
UM	23(0.20)	7(0.74)	812(3.27)	154(12.63)
UN	5457(48.23)	775(81.41)	626(2.52)	34(2.79)
UD	0(0.00)	0(0.00)	48(0.19)	4(0.33)
UP	2688(23.76)	59(6.20)	2431(9.79)	35(2.97)
UR	225(1.99)	31(3.26)	3030(12.20)	70(5.74)

TABLE 3—Continued

	INITIAL		TERMINAL	
	F	V	F	V
US	2514(22.22)	24(2.52)	4158(16.74)	490(40.20)
UT	203(1.79)	24(2.52)	10002(40.28)	78(6.40)
UU	0(0.00)	0(0.00)	1(0.00)	1(0.08)
UX	1(0.01)	1(0.11)	86(0.35)	22(1.80)
UY	0(0.00)	0(0.00)	122(0.49)	3(0.25)
UZ	0(0.00)	0(0.00)	9(0.04)	2(0.16)
VA	1619(25.32)	142(21.91)	96(0.91)	34(7.89)
VD	0(0.00)	0(0.00)	4(0.04)	1(0.23)
VE	1676(26.21)	174(26.85)	10175(96.80)	364(84.45)
VI	2019(31.57)	215(33.18)	2(0.02)	2(0.46)
VL	1(0.02)	1(0.15)	0(0.00)	0(0.00)
VO	1009(15.78)	100(15.43)	19(0.18)	9(2.09)
VP	2(0.03)	1(0.15)	2(0.02)	1(0.23)
VR	3(0.05)	3(0.46)	0(0.00)	0(0.00)
VS	16(0.25)	1(0.15)	19(0.18)	3(0.70)
VT	0(0.00)	0(0.00)	1(0.01)	1(0.23)
VU	47(0.73)	10(1.54)	1(0.01)	1(0.23)
VY	3(0.05)	1(0.15)	192(1.83)	15(3.48)
WA	15360(25.53)	278(25.67)	19(0.23)	8(3.96)
WD	0(0.00)	0(0.00)	64(0.79)	3(1.49)
WE	9348(15.54)	169(15.60)	2694(33.05)	15(7.43)
WH	14225(23.64)	151(13.94)	2(0.02)	1(0.50)
WI	13037(21.67)	234(21.61)	1(0.01)	1(0.50)
WK	0(0.00)	0(0.00)	15(0.18)	2(0.99)
WL	0(0.00)	0(0.00)	61(0.75)	15(7.43)
WN	0(0.00)	0(0.00)	2919(35.81)	69(34.16)
WO	7041(11.70)	166(15.33)	1412(17.32)	1(0.50)
WR	1128(1.87)	68(6.28)	2(0.02)	1(0.50)
WS	0(0.00)	0(0.00)	941(11.54)	75(37.13)
WT	1(0.00)	1(0.09)	11(0.13)	4(1.98)
WU	4(0.01)	4(0.37)	1(0.01)	1(0.50)
WW	0(0.00)	0(0.00)	1(0.01)	1(0.50)
WY	23(0.04)	12(1.11)	9(0.11)	5(2.48)
XA	1(6.25)	1(12.50)	1(0.18)	1(5.00)
XE	4(25.00)	3(37.50)	9(1.66)	3(15.00)
XI	0(0.00)	0(0.00)	16(2.95)	1(5.00)
XT	0(0.00)	0(0.00)	492(90.77)	4(20.00)
XX	1(6.25)	1(12.50)	2(0.37)	2(10.00)
XY	10(62.50)	3(37.50)	22(4.06)	9(45.00)
YA	270(3.34)	44(28.76)	21(0.98)	12(6.28)
YC	0(0.00)	0(0.00)	1(0.05)	1(6.52)
YD	1(0.01)	1(0.65)	19(0.89)	5(2.62)
YE	2445(30.28)	40(26.14)	158(7.38)	17(8.90)
YF	2(0.02)	1(0.65)	0(0.00)	0(0.00)
YI	77(0.95)	8(5.23)	2(0.09)	2(1.05)
YK	0(0.00)	0(0.00)	5(0.23)	2(1.05)
YL	0(0.00)	0(0.00)	23(1.07)	10(5.24)
YM	0(0.00)	0(0.00)	9(0.42)	4(2.09)
YN	0(0.00)	0(0.00)	45(2.10)	14(7.33)
YO	5242(64.92)	43(28.10)	25(1.17)	6(3.14)
YP	0(0.00)	0(0.00)	6(0.28)	1(0.52)
YR	4(0.05)	1(0.65)	13(0.61)	3(1.57)
YS	1(0.01)	1(0.65)	1798(84.02)	110(57.59)
YT	1(0.01)	1(0.65)	11(0.51)	2(1.05)
YU	28(0.35)	11(7.19)	3(0.14)	1(0.52)
YV	2(0.02)	1(0.65)	0(0.00)	0(0.00)
YY	1(0.01)	1(0.65)	1(0.05)	1(0.52)
ZA	8(3.79)	8(10.81)	47(5.08)	17(10.83)
ZE	97(45.97)	22(29.73)	646(69.84)	109(69.43)
ZH	3(1.42)	3(4.05)	0(0.00)	0(0.00)
ZI	43(20.38)	15(20.27)	18(1.95)	3(1.91)
ZL	1(0.47)	1(1.35)	0(0.00)	0(0.00)
ZQ	47(22.27)	18(24.32)	20(2.16)	9(5.73)
ZT	0(0.00)	0(0.00)	1(0.11)	1(0.64)
ZU	9(4.27)	5(6.76)	5(0.54)	3(1.91)
ZH	3(1.42)	2(2.70)	0(0.00)	0(0.00)
ZY	0(0.00)	0(0.00)	73(7.89)	12(7.64)
ZZ	0(0.00)	0(0.00)	115(12.43)	3(1.91)

probability of occurrence based on frequency to be 13.46% and a versatility 10.74%.

A final way to estimate the information load of a letter(s) is to view it as reducing the number of possible letters which could follow it. Each letter in the alphabet may be an initial letter of a longer lexical unit or word. Disregarding (for the purpose of illustration) the uneven distribution of initial letters, the likelihood of any specific letter starting a word is 1 in 26. In our example of *prank*, the amount of information conveyed by *p* is 1 in 26 by this index. Given *p* as an initial letter, the occurrence of a specific second letter is restricted in our sample of orthographic types and tokens, as *p* combines with only 16 letters (see Table 3). If we extend this logic to larger lexical units, we generally find an ever diminishing probability of letter combinations and an enhanced predictable value for subsequent letters. This rule is rarely violated. An exception might be in the case of the letter *q* which is almost always followed by *u*. For example, we find (Solso & Juel, 1979) that only six *pr* trigrams appear in the English language regardless of positional constraints (*pra*, *pre*, *pro*, *pru*, and *pry*).

Further refinements of letter predictability can be made from the normative data presented in this paper in conjunction with other compatible data. Some of these refinements would include the probability of a designated letter occurring in a specific location. The probability of occurrence could be based on frequency of letters or the versatility of letters by position.

REFERENCES

- ATTNEAVE, F. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 1953, **46** (2), 81–86.
- BADDELEY, A. D., CONRAD, R., & THOMSON, W. E. Letter structure of the English language. *Nature (London)*, 1960, **186**, 414–416.
- BEIER, E. G., STARKWEATHER, J. A., & MILLER, D. E. Analysis of word frequencies in spoken language of children. *Language and Speech*, 1967, **10**, 217–227.
- BROWN, R., & MCNEILL, D. The "tip-of-the-tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*. 1966, **5**, 325–337.
- CARROLL, J. B., DAVIES, P., & RICHMAN, B. *Word frequency book*. New York: American Heritage, 1971.
- FANG, I. It isn't ETAOIN SHRDLU; it's ETAONI RSHDLC. *Journalism Quarterly*, 1966, **43**, 761–762.
- FISHER, V. L., & PRICE, J. H. Cues to word similarity used by children and adults: Supplementary report. *Perceptual and Motor Skills*, 1970, **31**, 849–850.
- FRIES, C. C. *Linguistics and reading*. New York: Holt, Rinehart & Winston, 1963.
- GAINES, H. F. *Cryptanalysis*. New York: Dover, 1956.
- GIBSON, E. J., PICK, A. D., OSSER, H., & HAMMOND, M. The role of grapheme–phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, **75**, 554–570.
- HORN, E. *A basic writing vocabulary: 10,000 words most commonly used in writing*. Iowa City: State Univ. of Iowa, 1926.
- HOWES, D. A word count of spoken English. *Journal of Verbal Learning and Verbal Behavior*. 1966, **5**, 572–604.
- JUEL, C., & SOLSO, R. L. Effects of orthographic and phonic structures on word identification. In M. L. Kamil & A. J. Moe (Eds.), *Perspectives in reading research and instruction: 29th yearbook of the National Reading Conference*. Washington, D.C.: The National Reading Conference, 1980. Pp. 154–161.
- JUEL, C., & SOLSO, R. L. The role of orthographic redundancy, versatility and spelling-sound correspondences in word identification. In M. L. Kamil (Ed.), *Directions in Reading: Research and Instruction, 30th yearbook of the National Reading Conference*. Washington, D.C.: The National Reading Conference, 1981, pp. 74–82.
- KAHN, D. *The codebreakers: The story of secret writing*. New York: Macmillan, 1967.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis of present-day American English*. Providence: Brown Univ. Press, 1967.
- LOTT, D., & SMITH, F. Knowledge of intraword redundancy by beginning readers. *Psychonomic Science*, 1970, **19**, 343–344.
- MARCHBANKS, G., & LEVIN, H. Cues by which children recognize words. *Journal of Educational Psychology*, 1965, **56**, 57–61.
- MASON, M. Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency. *Journal of Experimental Psychology: General*, 1975, **1**, 146–166.
- MASSARO, D. W., & TAYLOR, G. A. *Reading ability and the utilization of orthographic structure in reading*. (Tech. Rep. 515.) Madison, Wisconsin Research and Development Center for Individualized Schooling, July 1979.

- MASSARO, D. W., TAYLOR, G. A., VENEZKY, R. L., JASTRZEMBSKI, J. E., & LUCAS, P. A. Letter and word perception. Amsterdam: North-Holland, 1980.
- MASSARO, D. W., VENEZKY, R. L., & TAYLOR, G. A. Orthographic regularity, positional frequency, and visual processing of letter strings. *Journal of Experimental Psychology: General*, 1979, **108**, 107-124.
- MAYZNER, M. S., & TRESSELT, M. E. Table of single-letter and digram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1** (Whole No. 2).
- MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of pentagram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1** (Whole No. 5). (a)
- MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of tetragram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1** (Whole No. 4). (b)
- MAYZNER, M. S., TRESSELT, M. E., & WOLIN, B. R. Tables of trigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1** (Whole No. 3). (c)
- MCCLELLAND, J. L., & JOHNSTON, J. C. The role of familiar units in the perception of words and non-words. *Perception and Psychophysics* 1977, **22**, 249-261.
- MILLER, G. A., & SELFRIDGE, J. A. Verbal context and the recall of meaningful material. *American Journal of Psychology*, 1950, **63**, 176-185.
- MILLIKEN, D. D. *Elementary cryptography and cryptanalysis*. New York: New York Univ. Press, 1943. 3rd ed.
- MORRIS, P. E., & COOK, N. When do first letter mnemonics aid recall? *British Journal of Educational Psychology*, 1978, **48**, 22-28.
- O'HAVER, M. E., 1933. Cited in H. F. Gaines, *Cryptanalysis*. New York: Dover, 1956.
- PRATT, F. *Secret and urgent, the story of codes and ciphers*. Garden City, N.Y.: Blue Ribbon Books, 1942.
- RAYNER, K. Developmental changes in word recognition strategies. *Journal of Educational Psychology*, 1976, **68**, 323-329.
- RAYNER, K., & HAGELBERG, E. M. Word recognition cues for beginning and skilled readers. *Journal of Experimental Child Psychology*, 1975, **20**, 444-455.
- RINSLAND, H. D. *A basic vocabulary of elementary school children*. New York: Macmillan, 1945.
- RUBIN, D. C. Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 392-397.
- RUBIN, D. C. Word-initial and word-final ngram frequencies. *Journal of Reading Behavior*, 1978, **10**, 171-183.
- RUBIN, D. C. 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior*, 1980, **19**, 736-755.
- SAMUELS, S. J., & JEFFREY, W. E. Discriminability of words and letter cues in learning to read. *Journal of Educational Psychology*, 1966, **57** (6), 337-340.
- SMITH, L. D. *Cryptography*. New York: Norton, 1943.
- SOLSO, R. L. *Cognitive psychology*. New York: Harcourt, Brace, Jovanovich, 1979. (a)
- SOLSO, R. L. Positional frequency and versatility of letters for six-, seven-, and eight-letter English words. *Behavior Research Methods & Instrumentation*, 1979, **11**, 355-358. (b)
- SOLSO, R. L., BARBUTO, P. F., & JUEL, C. L. Bigram and trigram frequencies and versatilities in the English language. *Behavior Research Methods & Instrumentation*, 1979, **11**, 475-481.
- SOLSO, R. L., & BIERSDORFF, K. K. Recall under conditions of cumulative cues. *Journal of General Psychology*, 1975, **93**, 233-246.
- SOLSO, R. L., & JUEL, C. L. Positional frequency and versatility of bigrams for two- through nine-letter English words. *Behavior Research Methods & Instrumentation*, 1980, **12** (3), 297-343.
- SOLSO, R. L., & KING, J. F. Frequency and versatility of letters in the English language. *Behavior Research Methods & Instrumentation*, 1976, **8**, 283-286.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Teacher's College Press, Columbia University, 1944.
- TIMKO, H. G. Letter position in trigram discrimination by beginning readers. *Perceptual and Motor Skills*, 1972, **35**, 153-154.
- TRAVERS, J. R., & OLIVIER, D. C. Pronounceability and statistical "Englishness" as determinants of letter identification. *American Journal of Psychology*, 1978, **91**, 523-538.
- UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. New York: Lippincott, 1960.
- WEPMAN, J. M., & HASS, W. *A spoken word count*. Chicago: Language Research Association, 1969.
- WILLIAMS, J. P., BLUMBERG, E. L., & WILLIAMS, D. V. Cues used in visual word recognition. *Journal of Educational Psychology*, 1970, **61**, 310-315.
- YARDLEY, H. O., 1931. Cited in H. F. Gaines, *Cryptanalysis*. New York: Dover, 1956.
- YOUNG, M. N., & GIBSON, W. B. *How to develop an exceptional memory*. North Hollywood, Calif.: Wilshire, 1974.
- ZETTERSTEN, A. *A statistical study of the graphic system of present-day American English*. Lund, Sweden: Studentlitteratur, 1969.

(Received August 4, 1981)