

ADVANCES
IN
PSYCHOLOGY

4

Letter and Word Perception

D.W. Massaro, G.A. Taylor,
R.L. Venezky,
J.E. Jastrzembski, and P.A. Lucas

NORTH-HOLLAND

LETTER AND WORD PERCEPTION

ADVANCES
IN
PSYCHOLOGY

4

Editors

G. E. STELMACH

P. A. VROON



NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

LETTER AND WORD PERCEPTION

*Orthographic Structure and
Visual Processing in Reading*

Dominic W. MASSARO

Glen A. TAYLOR

Richard L. VENEZKY

James E. JASTRZEMBSKI

and

Peter A. LUCAS

*Wisconsin Research and Development Center
for Individualized Schooling
University of Wisconsin*



1980

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM · NEW YORK · OXFORD

Copyright © 1980 – The Board of Regents of the University of Wisconsin System

Copyright is claimed until 1990. Thereafter all portions of this work covered by this copyright will be in the public domain. This work was developed under a grant from the National Institute of Education, Department of Health, Education, and Welfare. However, the content does not necessarily reflect the position or policy of that Agency, and no official endorsement of these materials should be inferred.

ISBN: 0 444 85493 2

Publishers:

NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM•NEW YORK•OXFORD

Sole distributors for the U.S.A. and Canada:
ELSEVIER NORTH-HOLLAND, INC.
52 VANDERBILT AVENUE,
NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data

Main entry under title:

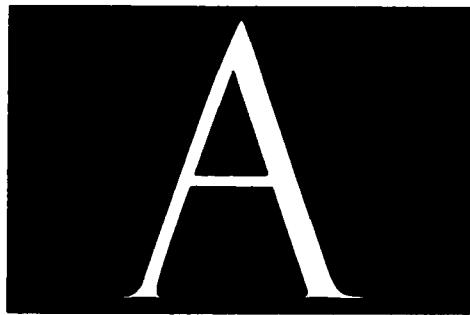
Letter and word perception.

(Advances in psychology ; 4)
Includes bibliographical references and indexes.
1. Reading. 2. Word recognition. 3. English
language--Alphabet. 4. Reading, Psychology of.
I. Massaro, Dominic W. II. Series: Advances in
psychology (Amsterdam) ; 4.
LB1050.L47 428.4 80-14350

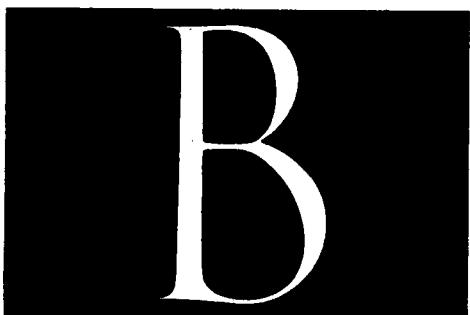
ISBN 0-444-85493-2

PRINTED IN THE NETHERLANDS

'The greatest thing in the world'



is the alphabet



as all wisdom is contained therein - except



the understanding of putting it together

This page intentionally left blank

Preface

...to completely analyze what we do when we read would almost be the acme of a psychologist's achievements, for it would be to describe very many of the most intricate workings of the human mind, as well as to unravel the tangled story of the most remarkable specific performance that civilization has learned in all its history.

These words are no less appropriate today than they were when Edmund Burke Huey confronted the challenge of understanding reading in 1908. And although most literate societies devote valuable financial and human resources to understanding reading, it often seems as if we are no nearer to achieving this goal. However, some progress is apparent. It is generally agreed that reading involves the utilization of a vast range of performance skills along with the active application of the knowledge resources of the reader. The goal for the researcher is to describe each of the skills and to evaluate how knowledge contributes to the skilled performance.

Our research investigates how knowledge of the orthography (rules of spelling) is actively engaged in letter and word perception of skilled readers. As might be expected, our research enterprise had its share of both substantive advances in our understanding of basic psychological processes and distracting glances at some of the mind's bewildering inconsistencies. Our goal in this monograph is to communicate both the progress and some of the distraction that we

experienced. Because of the detailed presentations of various descriptions of orthographic structure, the extended series of experimental studies, and the fine-grained analyses of the contribution of many descriptions of orthographic structure, piecemeal publications of this research effort would have been inappropriate. This monograph provides the challenge of not only communicating a protracted research enterprise but also an illuminating and internally-consistent account of how readers utilize knowledge of orthographic structure in letter and word perception. The success of this work will be evaluated by the researcher, and must stand the test of its applicability to reading instruction.

Given the relatively large number of new experimental studies to be reported, we decided to omit the unimportant details of the method, procedure, and results in the actual discussion of the experiments in Chapter 4. However, since this information is still of interest to many potential readers and researchers, it has been placed in appendices at the end of Chapter 4. Each experimental study is numbered in Chapter 4 and its appendix is given the same number. By placing unnecessary details in appendices, it is easier to maintain the logical continuity in the presentation of the experimental studies. When necessary, the reader can always turn directly to the relevant appendix for additional information.

In addition, Appendix 5.1 gives three performance measures for each of our 200 stimulus items so that the interested investigator can test additional descriptions of orthographic structure with these data. Since sublexical position-sensitive single-letter and bigram counts are important components of orthographic structure, Appendix 5.2 and 5.3 present normative counts based on a sample of roughly one million words of printed text. The only counts currently available in print are based on a sample only 2% as large. These counts should be valuable to researchers and educators interested in the normative structure of English orthography.

The primary objective of this monograph is the presentation of our new empirical and theoretical research. Because of space limitations, we were unable to review much of the other relevant research in letter and word perception. For reviews of this research, the reader is referred to Allport (1979), Baron (1978), Estes (1977, 1978), Gibson and Levin (1975), Henderson (1977), Johnson (1977), Krueger (1975),

Massaro (1975, 1978, 1979b, in press), and Pollatsek and Carr (1979).

This research is the outcome of contributions of five relatively disparate researchers. In time, the need to understand letter and word perception became the driving force and our original biases and beliefs were eventually modified by actual experimentation and theoretical tests. If this contribution benefits reading research half as much as it benefited each of the participants from the collaborative efforts, then the project can be considered to be a success.

This research effort has been supported by funds from the National Institute of Education to the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, Madison, Wisconsin 53706 U.S.A. Jack Wende and Rod McCoy contributed to the development of the computer-controlled research facilities. David Klitzke was responsible for some of the computer programming, Susan Chicone tested some of the subjects. Jennifer Argelander edited and Doris King typed the manuscript using a not always straightforward computer-editing system. Bob Cavey drew the many graphs. Don Anderson contributed the ABC print.

This page intentionally left blank

Contents

PREFACE	vii
1. READING AND INFORMATION PROCESSING	1
An Information Processing Model of Reading	2
2. VISUAL INFORMATION	11
Early Work	11
Feature Analysis	13
Letter Features	14
3. Orthographic Structure	23
Statistical Redundancy	24
Phonologically Based Descriptions	29
Rule-Governed Regularity	30
Phonological Constraints	30
Scribal Constraints	31
Morphemic Features	32
Psychological Assumptions	34
Dichotomy Issue	34
An Initial Algorithm for Rule-Governed Regularity	35
4. EXPERIMENTAL STUDIES	37

Massaro et al. 1979 Studies	38
Study 1: Initial Replications	47
Procedure	47
Orthographic Structure	49
Similarity Effects	51
Target-Catch Differences	55
Information Processing Model	55
Study 2: Precue Versus Postcue	61
Orthographic Structure	63
Similarity Effects	64
Target Versus Catch Trials	65
Information Processing Model	66
Study 3: Mixed-Case	67
Orthographic Structure	69
Similarity Effects	70
Study 4: Limited Viewing Time	71
Orthographic Structure	71
Similarity Effects	72
Target Versus Catch Trials	73
Discussion	73
Study 5: Speeded RT Task	
with Good Visual Information	73
Procedure	76
Orthographic Structure	77
Study 6: Speeded RT Task	
with Poor Visual Information	77
Procedure	78
Orthographic Structure	78
Study 7: High-Accuracy RT Task	
with Poor Visual Information	79
Procedure	79
Orthographic Structure	79
Summary of Accuracy and Reaction	
Time Experiments	80
Overt Judgments	81
Study 8: Typicality Ratings	82
Study 9: Positional-Frequency Ratings	85
Study 10: Regularity Versus	
Positional-Frequency Judgments	87
Summary of Overt Judgment Experiments	90

Appendix 4.1	Details of Method, Procedure, and Results of Study 1: Initial Replication	91
Method	91	
Results	94	
Appendix 4.1A	Rules for the Selection of Orthographically Regular and Irregular Strings	96
Appendix 4.1B	The 200 Stimulus Items Used in Studies 1-10	97
Appendix 4.2	Details of Method, Procedure, and Results of Study 2: Precue Versus Postcue	99
Method	99	
Results	100	
Appendix 4.3	Details of Method, Procedure, and Results of Study 3: Mixed-Case	101
Method	101	
Results	101	
Appendix 4.4	Details of Method, Procedure, and Results of Study 4: Limited Viewing Time	103
Method	103	
Results	103	
Appendix 4.5	Details of Method, Procedure, and Results of Study 5: Speeded RT Task with Good Visual Information	105
Method	105	
Results	106	
Appendix 4.6	Details of Method, Procedure, and Results of Study 6: Speeded RT Task with Poor Visual Information	108
Method	108	
Results	108	
Appendix 4.7	Details of Method, Procedure, and Results of Study 7: High-Accuracy RT Task with Poor Visual Information	111
Method	111	
Results	111	
Appendix 4.8	Details of Method, Procedure, and Results of Study 8: Typicality Ratings	113
Method	113	
Results	114	
Appendix 4.8A	The Pseudowords Used in Study 8	115

Appendix 4.8B Directions for Typicality Ratings Without Words and for Typicality Ratings with Pseudowords	116
Appendix 4.8C Directions for Typicality Ratings with Words	117
Appendix 4.9 Details of Method, Procedure, and Results of Study 9: Positional-Frequency Ratings	118
Method	118
Results	118
Appendix 4.9A Directions for Positional-Frequency Ratings	119
Appendix 4.10 Details of Method, Procedure, and Results of Study 10: Regularity Versus Positional-Frequency Paired-Judgments	120
Method	120
Results	121
Appendix 4.10A Regularity Instructions for Paired-Judgments	122
Appendix 4.10B Positional-Frequency Instructions for Paired-Judgments	123
5. STRUCTURAL DESCRIPTIONS	125
Descriptive Measures	125
Type Versus Token Counts	128
Scale of Frequency	130
Position-Sensitive Versus Position Insensitive Counts	132
The Most Effective Frequency Measure	133
Frequency Versus Regularity	138
Within Class Analysis	140
Word Frequency	141
Multiple Regression	141
Summary	142
Appendix 5.1 Stimulus Items and Results	143
Appendix 5.2 Single-Letter Positional Frequencies	149
Three-letter Words	150
Four-letter Words	152
Five-letter Words	154
Six-letter Words	156
Seven-letter Words	158

Appendix 5.3 Bigram Positional Frequencies	160
Three-letter Words	161
Four-letter Words	172
Five-letter Words	188
Six-letter Words	205
Seven-letter Words	224
6. SUMMARY AND CONCLUSIONS	245
REFERENCES	249
GLOSSARY	263
AUTHOR INDEX	269
SUBJECT INDEX	275

This page intentionally left blank

1 Reading and Information Processing

The processes by which words are recognized during reading have concerned psychologists and educators for almost a century. The basic assumption is that a thorough understanding of the processes involved in word recognition will afford a major advance in the understanding of how people read. Every major treatise on reading from Huey (1908/1968) to Gibson and Levin (1975) has discussed this issue at length, but generally has concluded, as did Huey (1908/1968, p. 102), that "it is very difficult to draw final conclusions concerning visual perception in reading...."

Recently two different panels which recommended funding priorities in basic skills research to the National Institute of Education assigned high priorities to studies on word recognition. Consequently, a high priority currently is being given to obtaining a better understanding of word recognition with stress on the basic stages of visual information processing. Information processing during reading is currently the primary concern of many experimental psychologists, and therefore, an expanding body of experimental data is being brought to bear on these questions. Since word recognition forms a bridge between visual information processing skills and comprehension, it is important for understanding how the two relate to each other.

Research on word recognition processes is seen as having a high potential for improving reading instruction. Good and poor readers have been shown to differ in performance on a variety of general tasks but until recently, little attempt has been made to explore how children with different levels of reading ability might differ in basic word recognition processes. We believe that basic word recognition processes,

once defined for competent adult readers, can be used to assess word recognition abilities of good and poor readers in the elementary grades. If good and poor readers utilize the same processes, but differ in speed or accuracy, then remediation methods based on common approaches to reading instruction would be warranted. However, if fundamental differences in visual recognition processes are found, then totally different remedial methods would be required.

The overall objective of this research project is to assess how the reader's higher-order knowledge of the language interacts with lower-level perceptual analyses during reading. The specific question that is addressed is how the reader's abstract knowledge about orthographic structure is combined with the information derived from visual featural analysis in word recognition. Orthographic structure refers to the spelling constraints in a written language. There is a considerable amount of predictability in English writing, for example, and the reader may actively utilize this information in word recognition. Visual featural analysis refers to the evaluation of component properties of letters in the course of letter and word recognition. Given evidence that both of these sources of information contribute to word recognition, the goal is to define what information the reader actually utilizes (what information is psychologically real) and to define how these sources of information are integrated together in reading. This requires examination of the psychological reality of various descriptions of visual features, of orthographic structure, and of quantitative models that describe how visual features and orthographic structure are integrated during word perception and recognition.

AN INFORMATION PROCESSING MODEL OF READING

Evaluation of the contributions of visual features and orthographic structure to word recognition requires that we describe in some detail the processes involved in reading. The model for describing these processes is part of a more general language processing model which has been developed and tested over the past few years (Massaro, 1975, 1978, 1979a). Figure 1.1 presents a schematic representation of the stages of processing in reading. At each stage of processing, memory and process components are represented. Each memory component (indicated by a rectangle) corresponds to the information available at a particular stage of processing. Each process

component (indicated by a circle) corresponds to the operations applied to the information held by the memory component. The memory components are temporary storages except for long-term memory which is relatively permanent and, in addition, supplements the information at some of the processing stages. This model provides a framework from which hypotheses about word recognition can be derived and tested.

The typical text in reading is a sequence of letters and spaces which conform to orthographic, syntactic, and semantic constraints of the written language. The average English reader begins at the top left-hand corner of the page and reads each line from left to right. The reader's eye movements across a line are not continuous but occur in a series of short jumps called saccades. The fixation time between saccades is roughly ten times longer than the saccade. The typical saccade of 1 to 2 degrees requires 20 to 30 msec, whereas fixation time averages 250 msec (Shebilske, 1975; Woodworth, 1938). Initial processing of the visual stimulus must occur during the fixation time because the stimulus pattern is blurred during a saccade and its duration is too short for sufficient processing to occur. During an eye fixation, the light pattern reflected from letters is transduced by the visual receptors and a process detects and transmits visual features to preperceptual visual storage (see Figure 1.1). In our model, we call this initial process feature detection. The features are described as visual because it is assumed that there is a direct relationship between the stimulus properties of the letters and

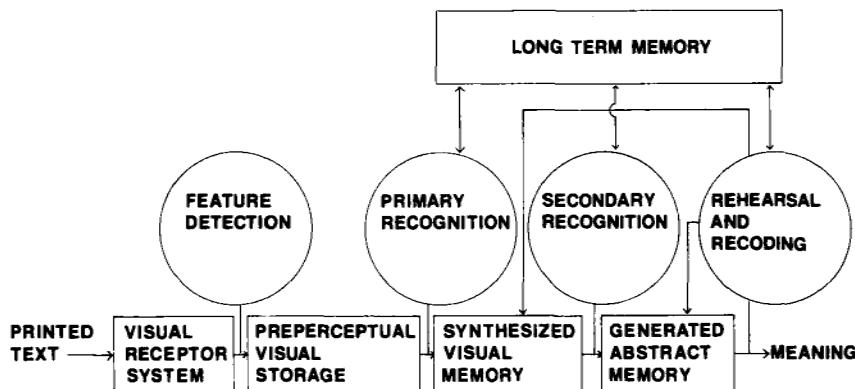


Figure 1.1 A stage model of reading printed text.

the information in preperceptual visual storage. The passive transduction of feature detection contrasts with the active construction of the subsequent processing stages. There is no exact one-to-one relationship between the input and output of the following processing stages since these later stages actively utilize information stored in long-term memory.

Given the set of visual features in preperceptual visual storage, the primary recognition process attempts to transform these isolated features into a sequence of letters and spaces in synthesized visual memory. To do this, the primary recognition process can utilize information held in long-term memory which for the accomplished reader includes a list of features for each letter of the alphabet along with information about the orthographic structure of the language. The primary recognition process utilizes both the visual features in preperceptual storage and knowledge of orthographic structure in its synthesis of the letter strings. The time-honored belief that recognition of letter strings becomes progressively easier as the strings more closely approximate words implies that readers capitalize on the presence of orthographic structure. Recognition usually means the process of discriminating, identifying, and correctly ordering the letters of a letter string (Smith, 1971). Given this interpretation of recognition, the facilitating effect of orthographic structure on letter-string recognition must almost be true by definition. Recognition involves accessing knowledge built up through experience. It is only natural that a reader's experience with words will generalize more to wordlike than nonwordlike strings. Our goals in this research endeavor are to provide a better understanding of the recognition process and to evaluate which aspects of orthographic structure the reader knows and utilizes.

The central assumption addressed here is that orthographic structure influences perceptual recognition (primary recognition in our model), in addition to other influences at later stages of processing. We believe that orthographic structure also facilitates processing from a perceptual to a conceptual level (secondary recognition in our model). By focusing on the facilitating effects of orthographic structure on visual perception, we have aligned ourselves with a tradition over 100 years old concerned with how knowledge influences perception. James (1890/1950, p. 444) captured the spirit of the work begun by Wundt with the often quoted statement, "...the only things which we commonly see are those which we preperceive."

In our model, the primary recognition process evaluates the features in preperceptual visual storage and compares or matches these features to descriptions of perceptual units in long-term memory. A perceptual unit or prototype represents each of the letters in long-term memory as it would be ideally represented in preperceptual visual storage. The primary recognition process seeks for each letter position the perceptual prototype that provides the best match to the featural information. It is at primary recognition that orthographic structure reduces uncertainty or contributes information. The outcome of primary recognition is, therefore, the joint product of featural information in preperceptual visual storage and knowledge of letter constraints in long-term memory.

Since there are a limited number of ways that sequences of letters and letter groups can be assembled to form English words, knowledge of these sequences can help the reader to resolve the letters in strings that conform to the language (cf. Massaro, 1975, Chapter 7). It also has been hypothesized that this knowledge can assist the reader in determining the relative positions of letters once they are recognized (Estes, 1975a, 1975b). For example, given the letters *ch*, knowledge of orthographic structure does not allow these letters to be read in the opposite order. However, Geoffrion (1976) found that the contribution of orthographic structure is not dependent on determining the relative positions of letters. In addition, recent research by Massaro (1979a) found a large effect of orthographic structure on letter recognition but no effect on the determination of relative spatial position. In the present research, the experiments are designed such that any facilitating effect of orthographic structure at primary recognition must be due to letter recognition and not position uncertainty.

The primary recognition process operates on a number of letters simultaneously (in parallel). The visual features detected at each spatial location define a set of possible letters for that position. The primary recognition process chooses from this set of candidates the letter alternative which has the best correspondence in terms of visual features. However, the selection of a best correspondence can be facilitated by knowledge of orthographic structure. The primary recognition process, therefore, attempts to utilize both the featural information in preperceptual storage and knowledge about the structure of legal letter strings. A critical concern is with the interaction of these two sources

of information; in our model, it is assumed that the two sources of information make independent contributions to primary recognition (Massaro, 1973, 1975, 1978, 1979b; Thompson & Massaro, 1973).

The present analysis of the facilitation of orthographic structure in letter recognition is identical in spirit to the original analysis provided by Miller, Bruner, and Postman (1954). They observed that the number of letters correctly reported from a tachistoscopic exposure increased with increases in familiarity of the letter pattern. The familiarity of the letter sequences was varied by the degree to which the sequence approximated printed English (Shannon, 1948). Utilizing the theory of information developed by Shannon (1948), however, they observed that the number of letters reported is not a direct index of the amount of stimulus information derived from the exposure itself since the redundancy of contextually constrained sequences is not taken into account.

Miller et al. (1954) calculated the redundancy of their sequences at each order of approximation to English using the counts given by Shannon (1951) and counts of their own. When performance was corrected for redundancy, there was no effect of familiarity (see also Tulving, 1963). This result was interpreted to mean that the amount of stimulus information received from the exposure did not depend on the familiarity of the letter pattern. In terms of our analysis, over two decades later, the results can be interpreted to mean that orthographic structure and visual information contribute independent sources of information in reading.

The current view was developed, in part, on the basis of experiments carried out using the Reicher-Wheeler paradigm in word recognition (Reicher, 1969; Wheeler, 1970; Thompson & Massaro, 1973). In this paradigm, subjects are presented with either a test word or a single test letter for a short duration followed immediately by a masking stimulus and two response alternatives. The response alternatives both spell words in the test word condition; for example, given the test stimulus WORD, the response alternatives ---D and ---K would be presented. In the corresponding single test letter condition, D is presented and followed by the response alternatives D and K. In these studies, recognition of the test word is about 10% better than recognition of the test letter.

Given the two-alternative forced-choice control, it is assumed that the reader utilizes orthographic structure to eliminate possible alternatives during the perception of the

test display (Thompson & Massaro, 1973). As an example, presented with the test stimulus WORD and given recognition of test letters WOR and a curvilinear segment of the final letter, the reader could narrow down the alternatives for the final letter to D, O, and Q. If O and Q are considered to be orthographically illegal in the context WOR-, then D represents an unambiguous choice. The reader will therefore perceive the word WORD given just partial information about the final letter. If the reader recognizes the same curvilinear segment in the corresponding single test letter condition, however, any of the three letters (D, O, and Q) are possible and the perceptual synthesis will result in D only one out of three times. What is critical in this analysis is that the better performance on words compared to single letters is obtained even though the visual featural information available to the primary recognition process is equivalent in the word and letter conditions. The orthographic structure of the word simply provides an additional but independent source of information. The featural information available to the recognition process does not change with changes in orthographic structure. In this view, although orthographic structure facilitates word perception, it does not modify the feature analysis of the printed pattern (Krueger & Shapiro, 1979; Massaro, 1979a).

As we have discussed, one role of orthographic structure is concentrated in the primary recognition process and serves to facilitate the recognition of individual letters. In addition, orthographic structure is functional at later stages of processing, such as its facilitating effect in short-term memory (Massaro, in press). Baddeley (1964), for example, showed that a well-structured string of letters is easier to remember and recall than a random string. Whether orthographic structure facilitates perceptual recognition is a more controversial issue. We assume that it is utilized in the following manner. Upon presentation of a letter string, the primary recognition process begins integrating and synthesizing featural information passed on by feature detection to preperceptual visual storage. Featural information is resolved at different rates and there is some evidence that gross features are available before the more detailed features (Massaro & Schmuller, 1975). The primary recognition process is faced with a succession of partial information states. These partial visual information states are supplemented with knowledge about orthographic structure. Assume, for example, an initial th has been perceived in a

letter string, and the features available for the next letter eliminate all alternatives except c or e. The primary recognition process might synthesize e without waiting for further visual information, since initial the is not acceptable while initial the is.

The primary recognition process transmits a sequence of recognized letters to synthesized visual memory. Figure 1.1 shows how the secondary recognition process transforms this synthesized visual percept into meaningful form in generated abstract memory. We assume secondary recognition attempts to close off the letter string into a word. The secondary recognition process makes this transformation by finding the best match between the letter string and a word in the lexicon in long-term memory. Knowledge of orthographic structure can also contribute to secondary recognition; word recognition can occur without complete recognition of all of the component letters. Given the letters bea and the viable alternatives l and t in final position, only t makes a word, and therefore word identification (lexical access) can be achieved (Massaro, 1977). Each word in the lexicon contains both perceptual and conceptual codes. The word which is recognized is the one whose perceptual code gives the best match and whose conceptual code is most appropriate in that particular context.

Generated abstract memory corresponds to the short-term or working memory of most information processing models. In our model, this memory is common to both speech perception and reading. Recoding and rehearsal processes build and maintain semantic and syntactic structures at the level of generated abstract memory. It is also possible to go from meaning to a visual or auditory percept in our model. The recoding operation can transform the meaning of a concept into its surface structure in an auditory or visual form.

Two issues concerning letter and word recognition in reading will be addressed in the present research. First, what are the visual features of letters utilized by the primary recognition process? Second, how is knowledge of orthographic structure psychologically represented? To address these issues, any study must be concerned with how these sources of information are evaluated, integrated, and combined in perceptual recognition (Massaro, 1979b). The discussion of visual features is presented in Chapter 2 and the discussion of orthographic structure is presented in Chapter 3. The fourth chapter focuses on how the reader utilizes both sources of information in a variety of perceptual recognition and overt

judgment tasks. To sustain the logical continuity of the presentation of the experiments, details of the method, procedure, and results are placed in appendices at the end of the chapter. The fifth chapter presents a detailed evaluation of the psychological reality of many descriptions of orthographic structure. The data base derived from the present experiments is also presented to allow additional tests of other descriptions of orthographic structure. In addition, normative counts of letter and letter cluster occurrences in text are presented. These counts make possible the measurement of a variety of stimulus and test items according to a number of formal descriptions of orthographic structure. The last chapter summarizes the contributions of the present research. The central concern of the empirical and theoretical work presented here is to evaluate the nature of the processes involved in letter and word recognition and to discover how knowledge about orthographic structure is represented and utilized.

This page intentionally left blank

2 Visual Information

Most of the traditional research relevant to visual information in reading has been concerned with the legibility of various type fonts and text formats. Spencer (1969) summarizes this work in an adequate and interesting manner and we will not attempt to duplicate a detailed coverage here. After mentioning a few general principles of the legibility of printed text, we will provide some discussion of the role of feature theories in describing how letters are recognized. Finally, we will evaluate the functional properties of the type fonts used in our experiments since this information will be necessary in our studies of how visual information and orthographic structure contribute to perceptual recognition.

EARLY WORK

Much of the early work revealed that letters are not equally legible. Table 2.1 presents the rank order of the legibility of the lowercase letters in an experiment by Cattell in 1885 (cited in Anderson & Dearborn, 1952) and an experiment in our laboratory some 90 years later. Except for a few of the letters such as q, x, and y, there is an impressive agreement between the two studies. Both studies found that some of the most frequent letters are the least legible. The highly frequent x-height letters a, c, e, o, and s were poorly recognized in both studies. The similar size and overall shape of these letters cause them to be confusable and, therefore, poorly legible. Cattell's research (1885), as well as more

Table 2.1 Rank order of distinctiveness of lowercase letters according to Cattell (1885) and in the present study.

<u>Letter</u>	Present		<u>Letter</u>	Present	
	Cattell	Study		Cattell	Study
a	19	22	n	18	17
b	6	1	o	16	24
c	25	19	p	7	2
d	1	9	q	4	8
e	22	23	r	15	12
f	17	18	s	26	25
g	24	4	t	12	21
h	5	6	u	9	14
i	23	26	v	13	16
j	11	20	w	8	10
k	2	7	x	20	3
l	10	15	y	21	5
m	3	11	z	14	13

recent studies by Bouma (1971) and ourselves, found that the slim letters f, i, j, l, r, and t were similarly confused for one another leading to relatively poor recognition. Finally, Cattell observed that legibility is highly correlated with overall letter size; contemporary studies also support this observation.

Tinker (1965) also noted agreement among studies of letter legibility. He reported correlations between .48 and .88 across studies using a variety of measurement techniques and typefaces. For instance, the letters d, m, p, q, and w were consistently found to be highly legible whereas c, e, i, l, and n were poorly legible. Tinker concluded that overall size, simplicity of outline, and the amount of white space enclosed by a letter are important determinants of legibility. We will provide a more detailed analysis of letter legibility and letter confusions later in this chapter.

Another common observation in early research was that text printed in uppercase letters is more difficult to read than its lowercase equivalent. Lowercase print may be more familiar than uppercase because of differences in frequency of occurrence. However, as Spencer (1969) points out, proportionately-spaced printed text in capitals occupies about 50% more space than it does in lowercase. (Uppercase and

lowercase occupy the same amount of space in typewritten text, as in the present monograph). The reader must maintain a larger span of apprehension in terms of visual angle in order to read uppercase text at the same rate as lowercase. Given that acuity falls off dramatically with extent into the peripheral vision, it is not surprising that lowercase text will show an advantage. It would be interesting to compare upper- and lowercase text when the size of the letters is adjusted to equate for visual angle. However, performance may still be at a disadvantage with uppercase since on the average, the larger letters will be closer together and allow for more lateral interference (the mutual degradation of the perceptibility of adjacent letters because of the proximity of their contours). The simplest explanation is that lowercase letters may be simply more perceptible than uppercase.

Javal (cited in Spencer, 1969) showed that reading is much easier when the top half, rather than the bottom half, of a line of words is exposed. It is not clear where the vertical center was defined, but we might assume that it was in the middle of an x-height letter. It would be valuable to define the perceptual middle, i.e., the point at which reading the top and bottom portions of a line of print is equally easy. Consistent with Javal's findings, Messmer (cited in Huey, 1908/1968) concluded that ascending letters (e.g., b, h) are "dominant" and contribute more to word recognition than x-height (e.g., a, x) or descending (e.g., g, p) letters. Zeitler (cited in Woodworth, 1938) also found that in short presentations, ascending and descending letters were more likely to be perceived (also see Huey, 1908/1968, pp. 79-87). Javal's idea was that ascending letters are more easily perceived because the eye tends to fixate above the middle of the line of print. Huey (1908/1968 p. 99) rejected this proposal by citing some statistics showing that ascending letters are much more frequent than descending letters, and therefore, a word is simply likely to have more information above than below the middle of a line of print.

FEATURE ANALYSIS

Consistent with the tradition of psychophysics, psychologists have been concerned with the visual characteristics of the stimulus that are functional in letter and word perception. The goal of this work is to relate the physical features of letters and words to their perceptual

recognition. The early workers in this area implicitly assumed that letters were the smallest units of visible analysis. In this view, letters are indivisible units that can not be analyzed further to provide a description of the relationship among different letters. In today's terminology, this view would correspond to a template view of letter recognition. Ideally, the reader would have a template corresponding to each of the letters in the language and perceptual recognition would involve finding the best match between the visual information in the letter and the template in memory.

Opposing the view of letters as indivisible units is the idea of letters having a number of distinctive features. This idea was generated by work on the linguistic description of speech sounds about 30 years ago (Jakobson, Fant, & Halle, 1961). Many structural linguists at that time treated phonemes as the smallest units of speech. If phoneme units are indivisible, then the physical or perceptual relationship between two phonemes is not of central interest. Jakobson et al. challenged this point of view by treating phonemes as consisting of a bundle of distinctive features. A relatively small number of distinctive features characterized each phoneme of the language. As an example, Jakobson et al. (1961) were able to use just nine features to distinguish 28 phonemes in English. The idea of distinctive feature analysis also proved productive in developing pattern recognition by machine. Selfridge and Neisser (1960) reviewed template matching schemes and feature analysis schemes for the recognition of letters. These developments in linguistics and artificial intelligence influenced the development of feature models of the visible characteristics of letters.

LETTER FEATURES

A valuable technique for the study of the visual features of letters involves the generation of confusion matrices. A confusion matrix indicates what letters will be seen as other letters when errors are made. To induce errors, either the visual input from the letter must be degraded or the processing time available for recognition of that letter must be limited. To degrade letters, they have been presented at relatively large distances or in peripheral vision. Processing time is not limited but featural information is degraded. To limit processing time, letters have been presented for a short duration at a good figure-ground contrast but closely followed

by a masking stimulus that serves to halt processing. The assumption in limiting processing time is that the reader makes an error because of insufficient time to process completely the featural information. Of course, it is important to analyze confusion matrices generated from both these methods and to determine whether they give similar results.

Prior to Bouma's (1971) study, most contemporary experiments studied the confusions of uppercase letters (e.g., Gibson, Osser, Schiff, & Smith, 1963; Townsend, 1971). Bouma utilized lowercase letters of the IBM Courier 10 typeface shown in Figure 2.1. Bouma presented the letters at a relatively large reading distance in one study and in peripheral vision in another. The confusions were very systematic and allowed Bouma to assign the letters to groups of letters that were confusable within a group but not among groups. The confusions among letters were primarily dependent upon overall letter shape and size. Many of the recognition confusions could be described by just three classes of letters: short or x-height letters, ascending letters (ascenders), and descending letters (descenders). Given an x-height letter, a recognition confusion usually resulted from another x-height letter and so on. The x-height letters and ascenders can also be broken down into smaller groups as indicated in Table 2.2. Bouma argues that overall letter shape is an important determinant of recognition confusions and introduces the concept of envelope. The envelope of a letter is the smallest polygon surrounding a letter without indentation. For example, the envelope for c would be circular, whereas y would be enclosed by an inverted

a c e m n .
o r s u
v w x z
b d f h
i k l t
g j p q y

Figure 2.1 Lowercase letters of the IBM Courier 10 typeface used by Bouma (1971).

Table 2.2 The seven groups of confusable letters of the IBM Courier 10 typeface (after Bouma, 1971).

1. a s z x
2. e o c
3. n m u
4. r v u
5. d h k b
6. t i l f
7. g p j y q

triangle. These envelopes of letters account for some of the confusion errors. As an example, the group of confusable letters (*a*, *s*, *z*, *x*) have roughly rectangular envelopes with inner parts.

In order to provide an independent assessment of the featural information that is functional in the recognition of lowercase letters, we carried out a letter recognition experiment. Figure 2.2 presents a photograph of the letters as they appeared in the experiment. A single letter was presented on each trial at a relatively dim intensity to keep overall

! " \$ ' [] , -
0 1 2 3 4 5 6 7 8 9 : ; ?
A B C D E F G H I J K L M N O
P Q R S T U V W X Y Z
a b c d e f g h i j k l m n o
p q r s t u v w x y z

Figure 2.2 The set of characters currently being used in our laboratory. The characters are plotted as points on an oscilloscope and the average character is composed of approximately 35 points.

accuracy at about 50%. The test letter was presented at the fixation point for a short duration and no masking stimulus was presented. Three subjects contributed a total of 296 observations for each of the 26 letters of the alphabet. Table 2.3 presents the overall accuracy for each of the 26 letters, the false alarm rates, the d' values, and confusion responses that occurred more than 5% of the time to a given test letter.

One possible limitation with these data and other studies of this nature is that the letters differ in overall accuracy, and therefore, there is not an equal opportunity for confusions for each of the letters. On the other hand, it might be argued that the overall accuracy of a given alternative may reflect how easily it is confused with other letters. That is to say, a letter completely dissimilar from the other letters would give good performance since it will not be confused for any of the alternatives. Regardless of which interpretation is more reasonable, it might be worthwhile to repeat the experiment while adjusting the intensity individually for each letter to keep its performance at 50% correct.

The percentage correct identification for each of the 26 letters of the alphabet in Table 2.3 shows that recognition accuracy ranged widely. This result may represent a visual effect or a decision effect. Some letters may provide more visual featural information than other letters under the same stimulus conditions. Alternatively, observers may derive equivalent amounts of featural information from all letters but may have decision biases in the interpretation of the featural information. As an example, observers might be biased to interpret the featural information in terms of letters that occur more frequently in the language. The decision interpretation, however, appears to be inconsistent with the results. For example, the infrequent letters x, y, and z averaged about 65% correct, whereas the frequent letters e, f, o, and s averaged about 35% correct. In contrast to what would be expected from a frequency bias, these least frequent letters are recognized much better than the more frequent letters. This analysis provides support for the idea that differences in accuracy are due to differences in the featural information of the letters. A more direct measure of visual resolution is to take a d' measure as proposed by the theory of signal detectability (Green & Swets, 1966; Massaro & Schmuller, 1975). Table 2.3 shows that these d' values correlate rather highly with the percentage correct values. Given the assumption that the d' values are not influenced by any differences in decision biases, the results

Table 2.3 Proportion of times a letter was recognized correctly (Hit) and was given as an incorrect response to other letters (False Alarm); the d' value; and the confusion responses greater than 5% (from present study).

Letter	Hit	False Alarm	<u>d'</u>	Confusions Greater than 5%
a	.372	.015	1.85	e(14), g (7)
b	.743	.020	2.72	
c	.449	.016	2.05	o(6), e(5)
d	.639	.011	2.64	
e	.365	.017	1.79	o(8), c(6), g(5)
f	.456	.020	1.94	i(14), l(8), t(7)
g	.703	.022	2.56	
h	.662	.020	2.46	n(5), b(5)
i	.277	.023	1.42	l(22), f(11), t(6), j(6)
j	.412	.017	1.92	l(7), i(6)
k	.655	.014	2.61	
l	.530	.022	2.10	i(10), k(5)
m	.628	.010	2.66	n(6)
n	.497	.013	2.24	h(15), r(8), m(5)
o	.331	.014	1.78	c(10), b(8), p(8), a(6)
p	.730	.016	2.78	
q	.649	.010	2.71	a(6)
r	.591	.023	2.24	t(8)
s	.324	.010	1.87	e(7), c(7), g(7), a(5)
t	.382	.020	1.76	f(12), r(12), i(9)
u	.554	.010	2.46	
v	.520	.015	2.16	y(13)
w	.632	.010	2.66	x(5)
x	.730	.010	2.94	
y	.686	.015	2.67	v(11)
z	.571	.010	2.51	x(7)

locate the differences in letter recognition at a visual stage of processing.

Table 2.3 also provides an index of the letters which are confused with other letters in the task. For each letter, all confusions given for that letter greater than 5% of the time are presented. As an example, given the stimulus a, subjects responded 14% of the time with the alternative e and 7% of the time with alternative g. This means that the letter a is often

interpreted as e. The confusions appear to be relatively systematic and therefore suggest which features are functional in letter recognition. Table 2.4 provides analogous data from the peripheral-viewing experiment carried out by Bouma on the IBM Courier 10 typeface. The confusions will be utilized in our later work when the relative similarity between letters becomes an important question in our experiments.

The letter confusions suggest that overall letter shape information may be obtained before complete letter perception (resolution). For example, 48% of the error responses to the stimulus c were e and o; this indicates that subjects saw a circular shape on many of the trials but did not know whether the letter was c, e, or o. This result might be interpreted to mean that the overall envelope of a letter is resolved before details of its component features. Independently of whether the whole is greater than the sum of its parts, a general impression of the whole (although ambiguous) is seen before all of the parts. That a letter's envelope (no pun intended) becomes available before its details is also consistent with fourier-analysis studies. There is evidence that information given by low spatial frequencies is processed faster than high spatial frequency information (Breitmeyer & Ganz, 1976; Broadbent, 1977). Accordingly, low spatial frequencies provide information about overall shape whereas high frequency information reduces uncertainty about letter detail. Either limited processing time or a reduced figure-ground contrast could allow envelope resolution without corresponding resolution of letter detail. In terms of the present analysis, the reader will have to decide which letter best fits the perceived but fuzzy envelope. For example, recognition of a circular envelope without further detail would be sufficient to limit the alternatives to the three letters c, e, and o.

Given that our primary motivation is to study the interaction of lower-level visual information with higher-order knowledge in letter recognition, we are not directly concerned with which letter features are utilized in perceptual recognition. There are now sufficient data consistent with the idea that letter recognition is mediated by features (Massaro & Schmuller, 1975). Features are not limited to small component properties but can be relatively complex relational properties such as the overall circular envelope of the letters c, e, and o. Additionally, in contrast to the common assumption that features are detected in an all-or-none manner, we prefer to view the feature detection process as continuous. Visible features may be fuzzy so that recognition consists of

Table 2.4 Proportion of times a letter was recognized correctly (Hit) and was given as an incorrect response to other letters (False Alarm); the corresponding d' value; and the confusion responses greater than 5% (from Bouma, 1971).

Letter	Hit	False Alarm	d'	Confusions Greater than 5%
a	.620	.053	1.93	e(7), n(6), d(5), m(5)
b	.620	.024	2.29	h(27)
c	.190	.006	1.62	e(29), o(19), a(8), g(6)
d	.830	.019	3.03	q(5)
e	.340	.045	1.29	a(19), m(10), n(7), u(7), o(5)
f	.810	.013	3.11	r(5)
g	.310	.018	1.60	q(31), p(9), e(6)
h	.800	.028	2.76	b(14)
i	.690	.035	2.31	l(10), t(8)
j	.820	.008	3.33	d(5), l(5)
k	.510	.008	2.46	h(17), b(14)
l	.260	.007	1.81	i(50), j(9)
m	.790	.028	2.72	a(9)
n	.560	.026	2.10	m(21), a(6), h(6)
o	.570	.017	2.30	e(11), c(7), a(5)
p	.840	.017	3.12	b(8)
q	.740	.019	2.71	g(6), d(5)
r	.550	.015	2.31	f(13), t(12), e(5)
s	.100	.009	1.08	a(33), e(14), m(8), n(8), u(7)
t	.590	.013	2.46	i(11), k(5), f(5)
u	.560	.018	2.26	n(11), m(9), b(5)
v	.720	.016	2.73	w(17), y(5)
w	.700	.014	2.73	v(22)
x	.320	.003	2.25	a(13), e(9), z(6), n(6), k(5)
y	.570	.003	2.95	
z	.120	.004	1.46	a(19), e(14), r(11), g(8), s(6)

ascertaining the degree to which that feature is present or absent in the stimulus. Oden and Massaro (1978) have developed a model of speech perception based on the idea of continuous acoustic features. A similar analysis would seem to be a valuable contribution in letter recognition, although beyond the scope of the present research. For interested readers, Oden (1979) initiated an empirical and theoretical study of fuzzy featural information in letters, and Massaro (1979a) quantified a model to account for how continuous featural information of a letter is integrated with orthographic context in letter recognition.

This page intentionally left blank

3 Orthographic Structure

In this chapter we describe the fundamental properties of orthographic structure and speculate on how these properties might facilitate letter and word recognition. The goal of this discussion is to derive different metrics for scaling the orthographic structure of letter strings based upon different assumptions about how orthographic structure contributes to recognition processes. The experiments presented in Chapter 4 attempt to evaluate the psychological reality of several of these measures.

Central to the discussion of this chapter is the assumption that skilled readers derive information about legal occurrences of letters or letter sequences from repeated exposures to printed words and that they employ this information to facilitate word recognition in normal reading tasks. Our concern here is with both the description of the abstracted information and its utilization in the recognition process. In the discussion which follows we concentrate on two broad categories of orthographic structure: statistical redundancy and rule-governed regularity. Statistical redundancy applies to those measures which can be derived solely from the frequency with which letters, letter sequences, and words occur in natural texts. Such measures might be based upon type or token counts. The measures might apply to letters, letter sequences, or to any classification of these entities (e.g., vowel, consonant). Rule-governed regularity applies to those properties of letters and letter sequences which can be generalized from either the phonological constraints of English words or the scribal conventions. The latter control the selection of letters for either representing sounds or marking graphemic, graphotactic, or morphological

functions. Rule-governed regularity usually specifies a letter string as regular or irregular. In addition, through the specification of rules for generating legal (i.e., English-like) letter strings, levels of regularity can be defined.

STATISTICAL REDUNDANCY

Ordered approximations to English, based on algorithms proposed by Shannon (1948), were the first descriptions of orthographic structure utilized by psychologists. Shannon (1948) created a zero-order approximation to English by drawing letters with replacement from a pool of 27 unweighted symbols--the 26 letters of the alphabet plus a space. (In this as in most other sequential generation techniques, an initial space symbol is assumed.) If this same procedure is employed with symbols that are weighted according to their frequencies of occurrence in texts, first-order approximations result.

An n-th order approximation is produced by first selecting a letter string of length n-1. This forms the beginning part of the new string. The text is then scanned for the first occurrence of this initial string. When found, the next letter after the letter string sequence is added to the string being generated, and a new search string is constructed by dropping the first letter of the last search string and adding the letter just located. This procedure is then repeated until a space occurs or a desired length reached. Miller, Bruner, and Postman (1954) used this technique to generate eight-letter pseudowords for a full-report task. Their fourth-order approximations included items like mossiant, oneticul, preveral, and favorial, all quite English-like. Wallach (1963), on the other hand, truncated two letters from each string used by Miller et al. thus occasionally producing a non-English spelling (e.g., mossia, loriai). Although Miller et al. (1954) were primarily concerned with information theory, later experimenters (e.g., Lefton, Spragins, & Byrnes, 1973; Mewhort & Beal, 1977) have advocated ordered approximations as psychologically relevant descriptions of orthographic structure.

A major shortcoming of ordered approximation descriptions is their inability to control structural variations within approximation levels. Although any large group of stimuli generated by this technique will approximate the frequencies of occurrence of English letter sequences, there is no metric

incorporated into the method for assessing the relative degree of orthographic structure of strings within order classes. In Gilmore and Egeth (1976), for example, zero-order approximations to English were used to generate what the authors intended to be nonwords, which they contrasted with real English words. If no other criteria were used to evaluate the 30 strings generated for use in their experiments, there was a 10% chance that at least one real word occurred (based on the assumption of 1,500 common four-letter words in English); the probability that pseudowords occurred was considerably higher.

A second shortcoming of a strict psychological application of this approach is its strict reliance on left-to-right generation of letters during word recognition. For example, given th___ and th_ugh as partially perceived six-letter strings, the probabilities of identifying the third letter in each would be the same in that only information to the left but not to the right of any letter position is relevant.

To overcome these difficulties, summed bigram and trigram counts have been proposed as metrics. Underwood and Schulz (1960) tabulated the bigrams and trigrams in 2,090 words sampled from Thorndike and Lorge (1944). Since each sample word was weighted by its frequency of occurrence in texts, token-based bigram and trigram counts resulted. By summing the frequencies of all the bigrams or trigrams within a letter string, a familiarity index can be derived. Such a procedure was used by Anisfeld (1964), for example, in reanalyzing the Gibson, Pick, Osser, and Hammond (1962) study of a full report of visually-presented letter strings varying in pronounceability. According to Anisfeld (1964), the Gibson et al. results could be accounted for as easily by differences in summed bigram counts as they could by the spelling-to-sound units advocated by Gibson et al. Gibson (1964), in a reply to Anisfeld (1964), however, showed that correlations of summed bigram and summed trigram counts with subject responses were low and nonsignificant. In addition, Gibson, Shurcliff, and Yonas (1970) found neither summed bigram nor summed trigram counts to be an adequate predictor of recognition scores for unpronounceable and pronounceable letter strings.

Gibson et al. (1970) did not provide a fair test of the frequency measures, however, because the counts were not adjusted for word length in any reasonable manner. Word length alone accounted for about 70% of the variance in the recall scores; the number of errors naturally increased with increases in the length of the letter string. The significant

correlation of pronounceability could be due to the simple fact that shorter words are more pronounceable than longer words. Supporting this interpretation, Spreen and Schulz (1966) found a negative correlation of .83 between word length and pronounceability for 329 common nouns. Therefore, the results of Gibson et al. (1962, 1970) really do not speak to the issue of pronounceability versus orthographic structure of letter strings when string length is held constant.

The Underwood and Schulz (1960) tables provide a means for differentiating familiar and unfamiliar strings within approximation classes, but suffer from lack of attention to word length or position of letters within a word. A common suffix like -ies would heavily weight a trigram count for a pseudoword whether in initial or final position according to these tables, yet in English words this sequence never occurs initially. If letter position within a word is a psychologically relevant variable, the Underwood and Schulz (1960) tables could not be used for exploring this factor.

This shortcoming has been overcome, however, in the Mayzner and Tresselt (1965) and Mayzner, Tresselt, and Wolin (1965a, 1965b) tables, which give bigram, trigram, and tetragram counts by word position and by word length for words from three to seven characters in length. Mayzner and Tresselt (1965) also give single-letter positional frequencies, which have been used by Mason (1975) to explore differences between good and poor readers in a recognition task. Although Mason (1975) confounds rule-governed orthographic structure with positional frequency, a small but significant positional-frequency effect remains when these variables are separated (Massaro, Venezky, & Taylor, 1979).

The relationships among single-letter positional frequency, bigram positional frequency, and trigram positional frequency have only begun to be explored. McClelland and Johnston (1977) observed that accuracy of report of four-letter items was significantly correlated with summed single-letter positional frequency but not with bigram positional frequency. Bouwhuis (1979) found that single-letter positional frequency made a significant contribution to lexical decisions about three-letter items. Reaction times to words decreased with increases in frequency whereas reaction times to pseudowords increased. Similarly, high single-letter positional frequency increased the number of "word" responses to both words and pseudowords. These significant correlations were considerably diminished when bigram positional frequency was used as the predictor variable. These results imply that the power of the

bigram frequency measure (see Chapter 5) with our six-letter items may not generalize completely to smaller letter-string lengths.

It is possible to generate letter strings while independently varying summed single-letter positional-frequency and summed bigram positional-frequency counts. With some straining, the same can be done for summed bigram and summed trigram counts. For example, nachim has (according to Mayzner, et al. 1965b) legal bigrams but no legal trigrams. Nather, on the other hand, has legal bigrams and legal trigrams. A legal bigram or a legal trigram means that the spelling pattern occurs with frequencies above zero in the Mayzner et al. tables. The Mayzner et al. tables might be based on too small a sample for reliable counts. Consequently, we have devised tables from the Kucera and Francis (1967) million-word list and have utilized them in the present project. The description of the analysis and the single-letter and bigram tables are given in Appendix 5.2 and 5.3.

A further issue that needs to be clarified is the relative importance of token measures and type measures. Token counts represent a summed occurrence for some letter sequence, regardless of the number of unique words which contribute to this sum. Type counts, on the other hand, represent a sum of the unique words in which a letter sequence occurs, regardless of the frequencies of occurrence of the words. In all of the experimental work reported so far that used sequential redundancy measures, token counts have been employed. Solso and his colleagues (Solso, 1979; Solso & King, 1976; Solso, Barbuto, & Juel, 1979; Solso, Topper, & Macey, 1973; Topper, Macey, & Solso, 1973) report some type-based tabulations based on the Kucera and Francis (1967) word lists but no experimental data. However, there is no a priori reason for assuming that word tokens have greater psychological reality than word types nor are there empirical data relevant to this question. We address this issue in Chapter 5.

Type-based measures could be derived from a representation of an internal lexicon, but token measures also require the existence of distributional information in memory. While it is reasonable to assume that some forms of both token and type information might have psychological reality, how the reader might apply raw token measures in word recognition is not completely obvious (cf. Savin, 1963). However, some interesting approaches to using token measures embedded within type measures have been proposed by Solomon and Postman (1952), Savin (1963), and more recently by Landauer and Streeter

(1973). As an example, the viable alternatives could be narrowed down to all words consistent with the featural information and the reader might choose the most frequent word from this set.

But whatever application might be shown for frequency information in word recognition, the specific role of these measures has yet to be determined. If summed frequency (token or type) is psychologically relevant, then the utilization of a measure like bigram frequency would require a maximizing process. Consider the task of perceiving a five-letter string where the visual information allows an acceptable resolution of all letters except one, e.g., L1L2L4L5. Assume that no visual information is available for L3. The use of bigram information requires that L3 be selected to maximize the sum of the bigram frequencies of L2L3 and L3L4, selecting L3 from a pool of all 26 letters. If some visual information is available for L3, then the pool of available letters might be restricted to those that were compatible with the available information but otherwise the strategy would be the same. For a trigram strategy, a similar process could be used but with the maximizing of the combined frequencies for L1L2L3 + L2L3L4 + L3L4L5. The complexity of this strategy does not necessarily make its psychological relevance suspect.

A potential drawback to summed single-letter, bigram, and trigram positional counts, particularly where token counts are utilized, is the dominating role played by a small group of high-frequency words. For the four-letter words, as an example, the 37 most frequent words of this length account for approximately one-half of all tokens of four-letter words. This guarantees, on the average, a high, positive correlation between different statistical measures. However, low summed single-letter positional-frequency strings will usually be low in summed positional bigram and trigram frequency, but strings high in summed single-letter positional frequency are not necessarily high in the other two measures.

Furthermore, since the extremely high frequency words in English are mostly three, four, and (to a lesser degree) five letters in length, the influence of the highest frequency words necessarily decreases with increasing word length. Thus, we would expect a lower correlation between bigram and trigram counts in seven-letter strings than in four-letter strings.

The central limitation in frequency-based schemes, even when position and word lengths are observed, is that they can generate pseudowords which violate English phonological and spelling conventions. For example, thrsm, sthse, and whrst are

unpronounceable but have relatively high summed bigram positional counts. These can be contrasted with bipon, slevy, and dufip, which have relatively low summed bigram positional counts, but are legal from a rule-governed standpoint. Violations can be avoided, for the most part, by using position-sensitive fourth-order approximations which are also high in summed trigram positional counts.

PHONOLOGICALLY-BASED DESCRIPTIONS

An alternative to statistical redundancy has been to use the relationship between spelling and sound (i.e., pronounceability) as a defining metric. Gibson et al. (1962) found that letters of pronounceable pseudowords were more easily recalled in a full report than were letters of unpronounceable nonwords (e.g., dink vs. nkid). The authors hypothesized that the operative unit in word perception was not the letter (or the whole word), but a sequence of letters which had an invariant spelling-to-sound relationship. However, Gibson et al. (1970) obtained the same results with congenitally deaf subjects, thereby demonstrating that the spelling-to-sound relationship could not alone account for the response differences. It is likely that orthographic structure differences (whether frequency-based or rule-governed) were critical.

Letter-sound correspondences have also been utilized in word recognition models by Rubenstein, Lewis, and Rubenstein (1971) and by Smith and Spoehr (1974). However, both the paradigm employed and the assumed use of letter-sound relationships in the latter two studies differed from that in Gibson et al. (1962). Both Rubenstein et al. (1971) and Smith and Spoehr (1974) assumed that letter recognition occurs on the basis of letter features; then sounds are attached to letters or letter sequences and lexical access made on the basis of a phonological unit. Gibson et al. (1962) assumed that word recognition occurred on the basis of higher-order visual units which the reader formed through experience with letter-sound translation. The reliability of any of these hypotheses, when viewed in relation to the complexities of letter-sound correspondences is questionable (cf. Venezky, 1970).

Furthermore, some empirical evidence indicates that phonological encoding is not necessary for lexical access (Kleiman, 1975; Levy, 1978; Massaro, 1978, 1979b; Tzeng, Hung, & Wang, 1977). However, even if phonological translation were shown to play a role in word recognition, pronounceability

alone would still be inadequate for describing orthographic structure. English spelling is constrained by phonological conventions first, but also by scribal conventions, which outlaw certain possible but pronounceable spellings in selected contexts.

RULE-GOVERNED REGULARITY

Rule-governed regularity describes the predictable structure of the orthography in terms of phonological and scribal constraints. However, a critical feature of this description is that the predictable structure of letter occurrences is utilized without any mediation of the spoken language. In contrast to the spelling-to-sound description of Gibson et al. (1962), Rubenstein et al. (1971), and Smith and Spoehr (1974), utilization of orthographic structure in perceptual recognition is assumed to occur before the printed pattern is translated to sound. Although much of the structure of a written alphabetic language necessarily follows its phonology, access to the phonology is not necessary for utilization of the orthographic structure. Direct evidence for this assumption comes from the utilization of orthographic structure by deaf readers (Gibson et al., 1970). In addition, Singer (in press) found that a phonological rendering of words presented in an artificial alphabet was not necessary for learning and utilizing orthographic structure. In the sections which follow, both the phonological and the scribal constraints of English spellings are outlined.

Phonological Constraints

The primary constraints on English orthography derive from the allowable sequences of sounds in English words. Thus, /sk/ is an allowable initial consonant cluster and therefore spellings for it exist (e.g., sk, sc, sch); on the other hand, /tl/ and /d1/ do not occur and never did occur initially in English words, and therefore no spelling for these sequences occur in English words (but they do in Tlingit and several other languages in which initial /tl/ and /d1/ are legal). Complicating this relationship between sound and spelling, however, are several factors, including scribal pedantry, which leave us with spellings for (1) sound sequences which once occurred, but have since been dropped from the language (e.g., wr-, kn-, gn-), or for (2) sound sequences that were presumed

to have occurred in the language from which a word was borrowed (e.g., psychology, ptarmigan, debt). For psychology, this assumption is supportable; for ptarmigan it is clearly based on mistaken etymology; and for debt (and various others), it is only partially true, in that /b/ and /t/ occurred in the Latin ancestor of debt (debitum), but not as a final cluster.

A further complication is the deviation in serial order of several spellings from the order of sounds which they represent. Chief among these is wh, a spelling which at least some English speakers have retained as /hw/, and -le as in bottle, etc., which spells (in deliberate speech) /əl/ (wh- and -le are scribal reversals of earlier hw- and -el).

But even with these deviations from expected practice, most English sounds are represented in a moderately rational fashion in the orthography (Venezky, 1970). However, since English possesses more than one potential spelling for almost every sound, a second set of constraints, here called scribal constraints, enter into rule-governed orthographic structure.

Scribal Constraints

To explain the rudiments of these constraints, some new terminology is helpful.

Vowel spellings are classed as primary (a, e, i, o, u, y) or secondary (ai, ay, ea, oa, etc.).

Consonant spellings are either functionally simple (b, c, ch, d, etc.) or functionally compound (x, tch, dg, wh, and ck).

Graphemic constraints regulate primarily the distribution of (1) geminated (double) letters, (2) u-y-w, and (3) y-i. First, unlike Spanish, English resists doubled letters in word-initial position, especially consonants. Only a handful of exceptions to this exists, primarily involving the form oo (e.g., oodles, ooze). In other positions only 15 letters can double; those that don't (or rarely do) are a, h, i, j, k, g, u, v, w, x, y. But there are further restrictions. With few exceptions, doubled consonants do not occur after secondary vowels, and in final position only ff, ll, and ss are common. Finally, the digraph units are not allowed to double. For one, ch, a pseudogeminate (tch) was adopted in early modern English, at about the same time that dg was adopted to represent doubled g when it represented /j/. Similarly, ck was invented for double k which now occurs only in inflected forms of trek (e.g., trekking).

A second set of graphemic constraints regulate the distribution of u-y-w and i-y. From the Middle English period

up to about the middle of the 17th century, u and y were used indiscriminately to represent /u/ and /v/, although y tended to be used initially and u elsewhere. (According to Pyles (1964), y was also preferred in the vicinity of m and n). Digraphs like ou, when written before a vowel might be mistaken for ov; therefore u was doubled, and eventually written w. In time, u in final position as the second part of a secondary vowel spelling was replaced by w, giving the present-day alternations au/aw, eu/ew, and ou/ow. In general, the w variant occurs before vowels and in final position; the u variant occurs elsewhere. Where u would occur as a primary vowel spelling in final position, a final e is added (e.g., blue).

For reasons not altogether clear, but related to the ambiguity of u-y spellings, y was also exorcised from final position. But since no simple replacement was available as was for u, an e was added after the letter, thus giving present-day spellings like dove, love, and have, where the final e does not mark a free (long) vowel pronunciation.

Similarly, i-y now alternate exactly as u and w do, as second elements of a digraph vowel, thus, maid/may, oil/boy, and either/grey.

Three additional features of English spelling which have limited utility in sound-to-spelling rules, but which are important for the recognition of real words, are discussed briefly in the following section.

Morphemic Features

Morpheme Identity. The retention of morphemic identity as in pairs like cone/conic is a natural consequence of conditioned sound change unaccompanied by a corresponding spelling change and not the result of overt scribal intention. For the most part, however, retention of morphemic identity is limited to vowels and those consonant letters which are silent in one form but pronounced in a related one. In the first category are alternations of the form sane/sanity and telegraph/telegraphy; in the latter are pairs akin to sign/signal, bomb/bombard, and hymn/hymnal. Where no single spelling can stand for a pair of sounds, consistent spelling of allomorphs which differ by those sounds is not possible. Thus, collide/collision must utilize different spellings for /d/ and /z/.

In bound morphemes, similar processes can be seen as exemplified by the noun plural (and third-person singular) -s and the past tense marker, -ed.

Separation of homophones. Partly by accident and partly by overt attention, English spelling has tended to differentiate homophones wherever orthographic license would allow. Pairs like sea/see, blue/blew, and right/write belong to the accident category, but foul/fowl, plain/plane, and sun/son result from overt scribal practice (Vachek, 1959, 1973; on the orthographic distinctiveness of homophones, see Olson & Kausler, 1971).

The two-letter word ban. Since at least the time of Noah Webster and probably before, English lexicographers have avoided extending the repertoire of two-letter words beyond a small group of function words (e.g., in, of, on, if). Two mechanisms were adopted for lengthening two-letter words which didn't qualify for inclusion in this group: doubling the final consonant, as in ebb, add, odd, egg, and inn, or adding a final e, (e.g., doe, hoe, dye, and lye). The former mechanism has created unique final consonant clusters in that b, d, g, and n do not double in final position in any other English words.

These graphemic constraints have evolved independently of sound, for unlike the orthographies of other major languages, modern English spelling preserves clear evidence of nearly 1,300 years of sound change, scribal tampering, partial reforms, and foreign intrusions. There is an underlying pattern to this orthography, as described in Venezky (1970), but there is also a substantial marginal mess which can only in part be organized into subpatterns or justified by appeal to such principles as the orthographic separation of homophones (e.g., bell/belle). Nevertheless, permeating the entire spelling system from organized core to the most disjoint outer fringes is an interaction between phonological and visual forces which ranges from productive counterpoint as in such spellings as cone/conic to the dissonance of such mots noirs as women.

Since no English language academy ever imposed a grand design on English spelling, what general principles there are resulted from the confluence of various scribal traditions with the periodic ravages of sound change unaccompanied by spelling change. Yet a principle of sorts does emerge from the current system, viz., "Accede to all visual interests up to, but not beyond, the point where passage from spelling to sound is impaired."

Psychological Assumptions

Rule-governed descriptions of orthographic structure differ from other definitions of orthographic structure in two aspects: First, in the utilization of word types rather than word tokens, and second, in classifying as regular certain spellings which do not occur (yet) in English words and classifying as irregular some that do. The first difference is not essential to the definition of orthographic structure, but it is probably more a reflection of the differing epistemologies of linguists and psychologists. The second difference, however, is central to the concept of orthographic structure and requires further explication.

The primary assumption behind the orthographic rules presented here is that readers abstract from their experiences with printed words those spellings that are productive in present-day English and use these to facilitate recognition. For example, silent initial letters, regardless of the frequency of occurrence of the words in which they occur, are classified as irregular since it is unlikely that new words brought into the English language would ever be spelled with them. (Spelling analogies might present an exception to this assumption and cannot be ignored. However, it is unlikely that silent initial spellings would ever be widely extended.)

On the other hand, the initial sequences voa- and woa- and the final sequences -eng, -erl, and -ofe do not occur in the first 20,000 words in the Thorndike and Lorge (1944) list, yet they are classified as orthographically regular since they do not violate any phonological or scribal pattern. Thus pseudowords like voam, woach, treng, derl, and strofe are pronounceable and structurally regular, even though they contain trigrams that do not occur in English words. The perceptual recognition of such strings relative to similar pseudowords with higher summed trigram counts is an issue that remains to be investigated.

Dichotomy Issue

Another issue which merits empirical investigation concerns the assumed dichotomous determinants of orthographic regularity. Given the two different constraints on orthographic structure discussed above, we should question whether or not the scribal and phonological elements are separated psychologically. Are, for example, irregular strings like clav and cklib processed differently from an illegal

string like tprif? All are illegal and non-occurring, yet they differ in type of illegality. Tprif contains an initial consonant cluster that is phonologically illegal and therefore scribally non-occurring. Clav and cklib are both phonologically acceptable, but contain scribally outlawed features, viz., final y in clav and initial ck in cklib. For example, does the probability of correctly identifying a target letter depend upon whether it is in a phonologically-illegal or a scribally-illegal test string?

AN INITIAL ALGORITHM FOR RULE-GOVERNED REGULARITY

The approach that we propose, but have not yet fully developed, for representing the disparate patterns just described is to construct a set of generative rules which begin by selecting a consonant-vowel-consonant structure for a phonological word. Replacement rules are applied to expand left to right of each vowel and each consonant unit until a phonologically legal string results. A second set of rules then applies sound-to-spelling rules to generate a legal pseudoword (or word). By relaxing rules at various levels of generation, we can then generate letter strings with controlled irregularities. For example, by removing the place restriction on final clusters of the form NASAL + STOP, we generate unpronounceable final clusters like -np and -mg. On the other hand, allowing final /v/ to be represented by v rather than ve allows scribally illegal forms like mov and slaiv. So far, a preliminary set of rules for generating legal monosyllables has been developed, drawing on a variety of sources (e.g., Whorf, 1956; Kurath, 1964; Chomsky & Halle, 1968; Venezky, 1970). A sample derivation of a legal pseudoword from these rules is shown in Table 3.1.

In summary, we have presented two general approaches to describing orthographic structure: statistical redundancy and rule-governed regularity. For the former, class metrics are derived from functions on the frequencies with which various size units occur (e.g., summed bigram counts). For the latter class we have suggested that a comparable metric might be developed through a set of generative rules which could be violated at various levels to produce nonwords with controlled deviations from rule-governed regularity.

Table 3.1 Derivation of a legal pseudoword (a).

A. Apply Base Generation Rules

<u>Step</u>	<u>Rule</u>	<u>Result</u>
1	Initialize	#W#
2	W -----> CVC	#C+V+C#
3	C/# -----> Cc	#Cc+V+C#
4	Cc/# -----> SCL	#SCL+V+C#
5	SCL -----> <u>s</u> +NASAL	# <u>s</u> +NASAL+V+C#
6	NASAL -----> <u>n</u>	# <u>s</u> +n+V+C#
7	V/else -----> Vf	# <u>s</u> +n+Vf+C#
8	Vf/else -----> Vfr	# <u>s</u> +n+Vfr+C#
9	Vfr -----> <u>i</u>	# <u>s</u> +n+i+C#
10	C/-# -----> Cs	# <u>s</u> +n+i+Cs#
11	Cs/Vf-# -----> STOPg	# <u>s</u> +n+i+STOPg#
12	STOPg -----> STOPvg	# <u>s</u> +n+i+STOPvg#
13	STOPvg -----> <u>b</u>	# <u>s</u> +n+i+b#

B. Apply Base-to-Spelling Rules

<u>Step</u>	<u>Rule</u>	<u>Result</u>
14	<u>s</u> /else -----> <s>	#<s>nib#
15	<u>n</u> -----> <n>	#<n>ib#
16	<u>i</u> /-Cs# -----> <e...e#>	#<sne>b<e>#
17	<u>b</u> -----> 	#<snebe>#

 (a) Terminal phonological symbols are underlined; terminal spelling symbols occur in wedges, e.g., <a>. A rule of the form X/# -----> Y is read "Rewrite X as Y when X occurs after juncture (#)."

4 Experimental Studies

The research reported in this book evolved from a commonality of interests among the authors. These interests primarily centered around furthering the understanding of the perceptual processes involved in word recognition and especially delineating the role that orthographic structure plays in these processes. An extension of an interesting study by Mason (1975) appeared to be a productive way to attack the problems of orthographic structure and perceptual recognition of words. Mason evaluated the contribution of summed single-letter positional frequency in a target search task. Good and poor readers indicated whether or not a six-letter string contained a predetermined target letter. Good readers averaged 63 msec faster for strings high than for strings low in positional frequency. We were concerned with two primary issues with respect to Mason's results. First, we questioned whether a frequency measure of orthographic structure was sufficient to describe the facilitating effects of structure on perceptual recognition. As discussed in Chapter 3, a rule-governed description of orthographic structure may be a more appropriate representation of the reader's knowledge.

Second, Mason's results appeared to be highly variable and we questioned whether the target search task using reaction time as a dependent measure reveals large effects of orthographic structure. There were a number of reasons to suspect the robustness of the effect in Mason's study. Subjects had only 5 practice trials and just 72 experimental trials. Each subject was tested on 24 words, 24 strings high in positional frequency, and 24 strings low in positional frequency. Given 12 good readers, there were 288 observations minus errors contributing to each experimental

condition. Mason found that 100 practice trials in a followup experiment reduced the advantage of performance on strings high in summed positional frequency to an average of about 22 msec. Since practice trials reduced the advantage by about two-thirds we felt that a replication of Mason's study with improved methodology was necessary to determine the reliability of her results.

MASSARO ET AL. 1979 STUDIES

Massaro, Venezky, and Taylor (1979) replicated Mason's experiment using the target search task with sixth-grade subjects. To obtain greater reliability, both the number of observations and the number of test stimuli were increased. Furthermore, the letter strings were not only high and low in positional frequency but also were contrasted along another dimension which we called orthographic regularity. We had not yet developed a systematic rule-based description of English orthography, but we were able to choose strings that appeared to be regular or irregular. Therefore, although we had no systematic description of regularity, we felt that the decisions we made to choose strings that were regular and irregular would provide a test of whether the facilitating effects of orthographic structure could be completely accounted for by the positional frequency hypothesis proposed by Mason.

To create the test stimuli, Massaro et al. (1979) selected four anagrams from each of 40 six-letter words such that the anagrams were either high or low in single-letter positional frequency and either regular or irregular. The resulting 2×2 factorial design provides a direct assessment of the contributions of frequency and regularity on letter-string perception. Each of the sixth graders began by taking a written comprehension test. The target search task was then introduced by showing the subjects the visual display and the response panel. To acquaint the subjects with the response procedure, they were first given 50 trials of responding "yes" and "no" to the words YES and NO presented on the visual display. Each subject pushed the "yes" button with his or her preferred hand. Next, the subjects received 50 practice trials with random letter strings. The subjects were asked to indicate whether or not a target letter was present in the six-letter string. Each trial began with the presentation of the target letter followed 800 msec later by the test letter string. The test letter string remained on until the subject

responded. Subjects were instructed to respond "yes" if the target letter appeared at least once in the letter string and "no" if it did not appear. The target letter appeared in the letter string on half of the trials (target trials) and did not appear on the other half (catch trials). They were told to respond as accurately as possible. The subjects were tested on each of the unique 160 stimuli, once on target trials and once on catch trials, giving roughly 80 observations per subject at each of the four experimental conditions (less error trials).

Figure 4.1 gives the results from this initial experiment. Although the reaction times were about 12 msec faster for

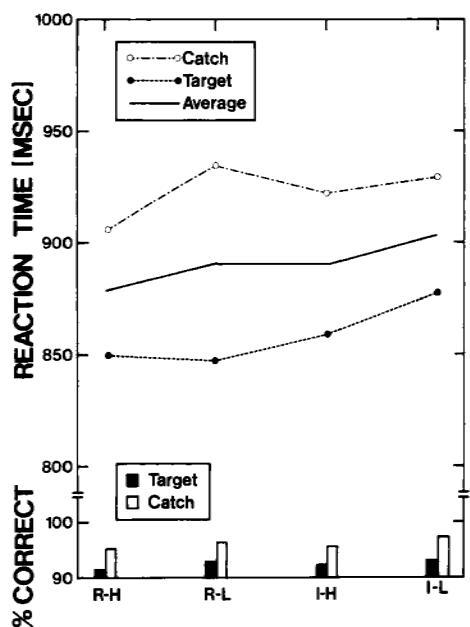


Figure 4.1 Reaction time and percentage correct for sixth graders as a function of orthographic regularity and summed single-letter positional frequency for target and catch trials (from Massaro et al., 1979) (key: R = orthographically regular; I = orthographically irregular; H = high summed single-letter positional frequency; L = low summed single-letter positional frequency).

regular than for irregular strings and about 12 msec faster for high than for low positional frequency, neither result was statistically significant. These results imply that neither positional frequency nor regularity contribute significantly to perceptual recognition in the target search task. One reason could be that our manipulation of positional frequency and regularity did not capture the knowledge that subjects utilize in perceptual recognition. A second possibility is that our manipulation of orthographic structure did mirror differences in the psychological reality of these structures but that the target search task using reaction times is insensitive to utilization of structure.

We suspected that the small effects in the first experiment were primarily due to the target search task rather than to the strings utilized in the experiments. Therefore, we carried out two more experiments to contrast the target search task utilizing reaction time (RT) as a dependent measure with a target search task using accuracy as a dependent measure. We employed college students rather than sixth graders as subjects since they were more readily available for testing. One experiment was a replication of our RT experiment. The other, an accuracy experiment, involved a slight modification of a more common paradigm used to study the utilization of orthographic structure. In variants of a Reicher-Wheeler postcue task (see Chapter 1), the subject is shown a word or a nonword for a brief duration followed by a masking stimulus. Two alternative letters are then given and the subject chooses which letter he or she thinks was presented on that trial. In order to make this accuracy task comparable to the RT task, we modified it by presenting just one alternative letter after the masking stimulus. In this case the six-letter item would be presented followed by a masking stimulus and then the target letter. The target letter would be present in the test item on half of the trials and not present on the other half. The subjects would simply indicate whether or not the target letter occurred in the test letter string. Performance, in terms of reaction time in the first experiment and overall accuracy in the second experiment, reflects how well the test strings were perceived.

The results of these two experiments are shown in Figures 4.2 and 4.3. In the reaction time task, reaction times averaged 16 msec faster for high than for low positional-frequency strings. Responses were 9 msec faster for regular than for irregular strings. The accuracy task showed that response accuracy was 5% higher for regular than for

irregular strings and 4% higher for high than for low positional-frequency strings. All of these results were statistically significant. Therefore, although the magnitude of the effects on reaction time for the adults was of the same order of magnitude as that for the sixth graders, lower variance among the adults made the effects statistically significant. One limitation at this stage of the research was to know whether or not sixth graders would give statistically significant accuracy results in the accuracy task. However, since this latter information was not available, we were faced with the problem of comparing the magnitude of the reaction time result with the magnitude of the accuracy result and realized that it resembled that of comparing apples and olives.

If it is accepted that the reaction time result is small,

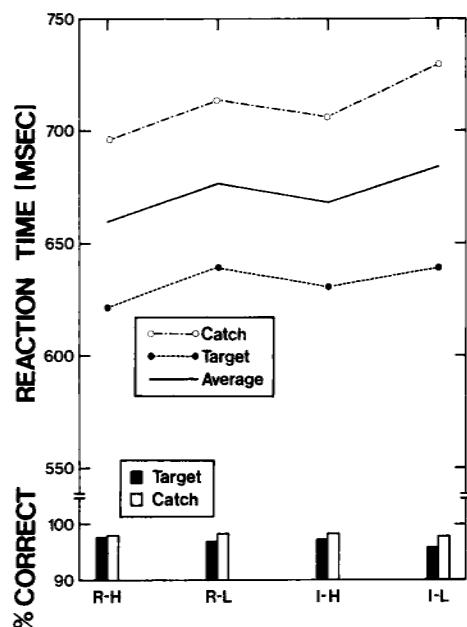


Figure 4.2 Reaction time and percentage correct for college students as a function of orthographic regularity and summed single-letter positional frequency for target and catch trials (from Massaro et al., 1979).

then one explanation for the small effect could take the following form. In the reaction time task the subject was given the target letter before the test string was presented. Since the subject simply was required to indicate whether or not the target was contained in the test string, he or she may not have had to resolve completely all of the letters in the string before a correct decision was made. For example, suppose a subject is looking for the target letter *b* in the letter strings *cose*, *deom*, and *deih* respectively. In all cases the appropriate response is no, but the subject may be able to decide on this faster with some of the strings than with others. Consider the following argument. There is some evidence that overall letter shape can be resolved and made available to later stages of processing before the letter is

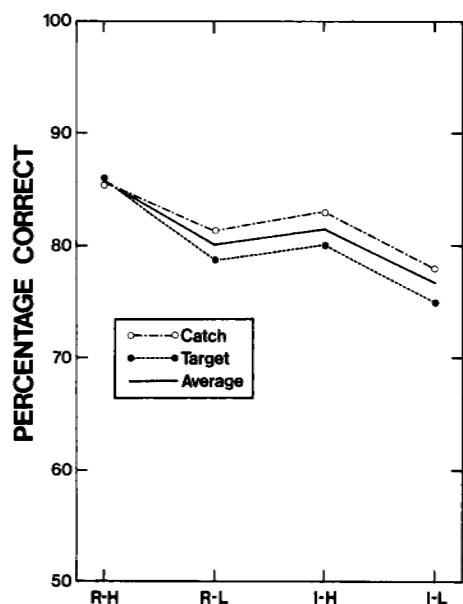


Figure 4.3 Percentage correct for college students as a function of orthographic regularity and summed single-letter positional frequency for target and catch trials (from Massaro et al., 1979).

Table 4.1 The six sets of letters based on the similarity results of Bouma (1971) used by Massaro et al. (1979).

1. m, n, r, u, v, w*
2. a, s, x, z
3. c, e, o
4. b, d, h, k
5. f, i, j, l, t
6. g, p, q, y

* Although Bouma's analysis indicated that the first set could be further divided into two sets consisting of m, n, u, and r, v, w, these two sets were combined because of their relative similarity and to provide more reliable data by increasing the number of observations within a set.

completely recognized (Bouma, 1971; Massaro & Schmuller, 1975). A target search task can be used to study the features necessary for recognition. Neisser (1963, 1967) for example, showed that subjects searched faster for an uppercase Z in a list of curved letters than in a list of linear letters. This result indicates that the subjects were able to reject the curved letters as foils more quickly than the linear letters. Applying this analysis to the present example, subjects should be able to respond "no" quickly to the extent that none of the letters in the test string has the same overall ascender shape as the target letter b. Therefore, it will take the subject longer to reject the string deih than to reject the string deom. The string cose should be rejected fastest of all. To the extent the subject is able to utilize this feature detection strategy, the contribution of orthographic structure might be attenuated. Therefore, it was of interest to see the extent to which subjects appeared to be using this strategy in

the reaction time and accuracy tasks. We hypothesized that this strategy would be utilized in the reaction time but not the accuracy task. In the accuracy task, the subject will resolve the letter string to the deepest level possible. In the reaction time task, having the target specified in advance and having the test letter string visible during the processing interval will allow the subject to terminate early on some catch trials.

To evaluate this hypothesis, the lowercase letters of the alphabet were grouped into six groups based on the similarity results of Bouma (1971). This grouping is given in Table 4.1. The relationship between each target letter and test string was defined in terms of the number of nontarget letters in the string that were members of the same set as the target letter.

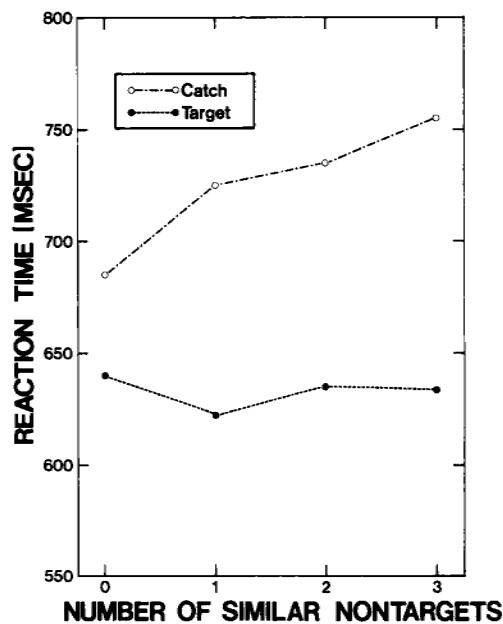


Figure 4.4 Reaction time in the target search task as a function of the number of similar nontarget letters in the test string (from Massaro et al., 1979).

For example, the test string lisver would have two letters (l and i) from the same set as the letter t, two (v and r) from the same set as u, and none from the same set as s. Figures 4.4 and 4.5 give this similarity analysis for reaction time and accuracy tasks respectively. The results show highly significant effects on catch trials in the reaction time task, supporting the idea of a feature detection strategy in the reaction time task.

The number of letters in the test string having the same overall shape as the target should not be important on target trials in the reaction time tasks. The target letter must be completely resolved before a "yes" response can be made. If the subjects simply responded "yes" whenever the overall shape of the target was perceived, recognition errors would be highly

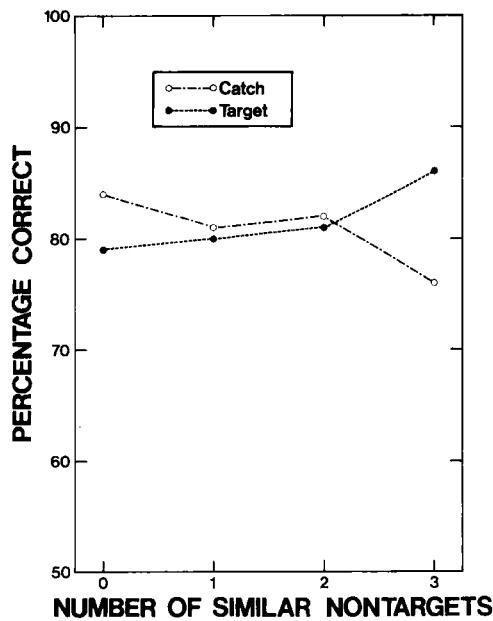


Figure 4.5 Percentage correct in the target search task as a function of the number of similar nontarget letters in the test string (from Massaro et al., 1979).

probable. Since a high accuracy criterion was enforced, we would not expect the number of similarly shaped letters to be a critical variable on target trials. As can be seen in Figure 4.4, the similarity variable had no significant effect on target trials in the reaction time experiment.

In the accuracy task, performance on catch trials tended to decrease with increases in the number of similar nontarget letters. In contrast, performance improved somewhat on target trials with increases in the number of similar nontarget letters. Although these results were not statistically significant, they show that similarity may also play a role in the accuracy task.

Although it is difficult to compare directly the magnitude of the effects of accuracy with those of reaction time, the similarity analysis makes such a comparison more meaningful. The reaction time study indicated small effects of orthographic structure relative to the effects of similarity. The accuracy study revealed large effects of orthographic structure relative to small effects of similarity. Contrasting these two variables in each task might allow a more reliable assessment of the contribution of orthographic structure. From this analysis it appears that orthographic structure plays a larger role in the accuracy study than it does in the reaction time study. One of the primary motivations of the present set of experiments was to explore the differences between these two tasks. Of course, it is not the tasks themselves that are interesting but rather, what they reveal about the perceptual recognition of letter strings.

The following set of experiments was also deemed necessary because of the limitations in the manipulation of orthographic structure in our previous studies. Although we attempted to choose regular and irregular strings for each level of high or low positional frequency, we were not confident that we had been completely successful. It was difficult to select regular-low positional-frequency items and irregular-high positional-frequency items. This followed from the fact that positional frequency and our intuitive impression of orthographic regularity might have been highly correlated. Therefore, we may have inadvertently confounded regularity with positional frequency in that the high positional-frequency items were more regular than the low positional-frequency items. It remained a possibility that the effect of positional frequency would be eliminated if we could find items that were equally regular for high and low positional frequencies. Therefore, we refined our description

of orthographic regularity and chose a new set of items on the basis of these rules for the present experiments. The rules are presented in Appendix 4.1A.

In addition to selecting new items we also employed word stimuli in the experiments. Krueger (1979) speculated that a much larger facilitation in the target search task would be found with words than with pseudowords. By including with our four categories of meaningless anagrams the words from which they were derived, we could provide a direct evaluation of the facilitating effects of word structure relative to positional frequency and regularity. We also saw the potential for testing a wider range of formal descriptions of orthographic structure in the present experiments. As indicated earlier, the qualitative test provided by the factorial design is limited. A good description of orthographic structure must also describe performance for individual letter strings. Therefore, we planned to analyze performance as a function of individual letter strings in the present studies. These analyses would go significantly beyond the qualitative comparisons between word, pseudoword, and nonword items of previous studies. A second motivation for including word items was that any description of orthographic structure must not be limited to nonword items but must also be functional for word items. If in some manner lexical (i.e., word) identity overrules orthographic structure, then the latter would not be psychologically relevant in reading.

Therefore the first two experiments of the present set of experiments can be considered to be a direct replication of the Massaro et al. (1979) experiments which have just been described. However, the present experiments included a new set of items along with the corresponding word stimuli. In this sense, the experiments conform to the experimental framework given by Sidman (1960) in that we are both replicating and extending previous experiments. Figure 4.6 illustrates the five types of items used in all of the current experiments.

STUDY 1: INITIAL REPLICATION

Procedure

Two experiments were carried out with 14 subjects in each experiment. Figure 4.7 gives a schematic representation of the procedures for the accuracy and RT experiments. In the accuracy experiment the test letter string was presented for a

short duration followed immediately by a masking stimulus. The duration of the test item was adjusted for each subject in order to obtain an average of 75% correct across all conditions in the task. The target letter followed the 208 msec masking stimulus. The subject's task was to indicate by pressing the appropriate button whether or not the target letter was contained in the test letter string. In the RT experiment the target letter was presented for 500 msec followed by a 1 sec blank period. The test string was then presented and remained on until the subject responded "yes" or "no" with respect to whether or not the target letter occurred in the test string. A reasonable amount of practice was given on both tasks. The test items were six-letter strings consisting of 40 words and 160 anagrams created by the factorial combination of high or low positional-frequency and regular or irregular

		Positional Frequency	
		High	Low
Orthographic Regularity	Regular	winter charge turned (1293)	trewin greach tredun (540)
	Irregular	wrntei ahcger rdnuet (1421)	rntewi hreagc edtrnu (526)

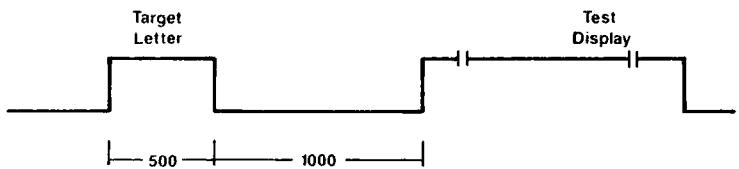
Figure 4.6 The five types of items and examples of each type used in the experiments. The number in each cell is the average summed single-letter positional frequency for that item type (based on the Mayzner & Tresselt, 1965, counts).

structure. On any trial, any of the five types of strings could be presented as either a catch or target trial. The test strings are presented in Appendix 4.1B. The details of the method, procedure, and results of the two tasks are presented in Appendix 4.1.

Orthographic Structure

Figures 4.8 and 4.9 present the results from the accuracy and RT tasks as a function of the five classes of items. Performance in the accuracy task depended greatly on the structure of the test string. Words were recognized 12% more often than the best regular high items. The regular strings were judged 8% more accurately than the irregular strings. Accuracy was 5% higher for high positional-frequency

PRECUE RT



POSTCUE ACCURACY



Figure 4.7 Schematic representations of the precue reaction time and postcue accuracy experiments.

strings than for low positional-frequency strings. All of these results were statistically significant. The detailed results and the statistical analyses are presented in Appendix 4.1.

In the RT task, reaction times to words were only 7 msec faster than reaction times to regular-high anagrams. Regular items were judged 23 msec faster than irregular items, whereas high positional-frequency items were responded to only 4 msec faster than low positional-frequency items. Only the effect of regularity was statistically significant. The detailed results and statistical analyses are presented in Appendix 4.1.

The highly significant effects of orthographic structure on response accuracy and the relatively minor, and for the most part, nonsignificant effects of orthographic structure on reaction time provide a direct replication of our previous

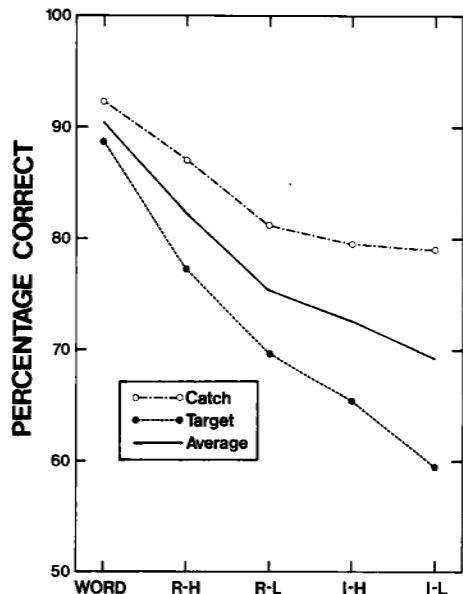


Figure 4.8 Percentage correct as a function of display type for target and catch trials in the postcue accuracy task of Study 1.

work. These results underscore the reliability of the regularity and positional-frequency effects on accuracy by demonstrating a generalization to both a new sample of subjects and a new sample of stimulus items (Clark, 1973). Neither regularity nor positional frequency, however, appears to be sufficient to account for the strong accuracy advantage found for words. The issue of whether or not orthographic structure rather than lexical identity can account for the word advantage will be delayed until a more detailed discussion and evaluation of orthographic structure is presented.

Similarity Effects

To test for the effects of similarity between the target letter sought in a display and the letters actually occurring

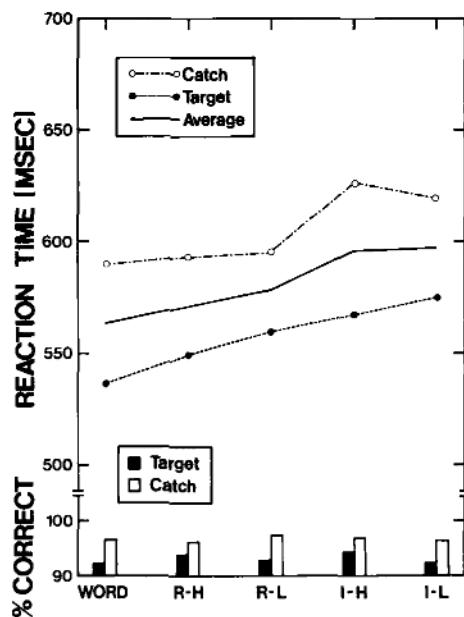


Figure 4.9 Reaction time and percentage correct as a function of display type for target and catch trials in the precue reaction time task of Study 1.

Table 4.2 For each of the 26 letters is listed the other letters in the alphabet that were erroneously identified as that letter more than 5% of the time (from Bouma, 1971).

Similar		Similar	
Target	Nontarget Letters	Target	Nontarget Letters
a	c, e, m, n, o, s, x, z	n	a, e, s, u, x
b	h, k, p, u	o	c, e
c	o	p	g
d	a, j, q	q	d, g
e	a, c, g, o, r, s, x, z	r	f, z
f	r, t	s	z
g	c, q, z	t	i, r
h	b, k, n	u	e, s
i	l, t	v	w
j	l	w	v
k	t, x	x	
l	i, j	y	v
m	a, e, n, s, u	z	x

in the display, the data were pooled across all five display types and repartitioned. The basis for this repartitioning was the classification of each trial's target letter according to which letters are likely to be seen as very similar to the target. The classification was derived from the lowercase single-letter confusability study of Bouma (1971) shown in table 2.4. We counted a nontarget letter as similar to the target letter if the nontarget letter was erroneously identified as the target more than 5% of the time in Bouma's study. Table 4.2 presents the similarity classification according to this analysis. For example, given the target

letter b, the similar nontarget letters are h, k, p, and u. This means that the stimulus letters h, k, p, and u were identified as b more than 5% of the time in Bouma's study. Each unique trial in our target search task was then described by the number of nontarget letters in the display that were similar to the target according to this classification. Roughly 46, 33, 13, and 7% of each subject's trials fell into the categories zero, one, two, and three similar letters (other than the target letter itself), respectively.

Figures 4.10 and 4.11 present the accuracy and reaction time results as a function of the number of similar nontarget letters in the display. Target versus catch trials is the curve parameter in the figures. In the accuracy task, accuracy increased with increases in the number of similar nontarget

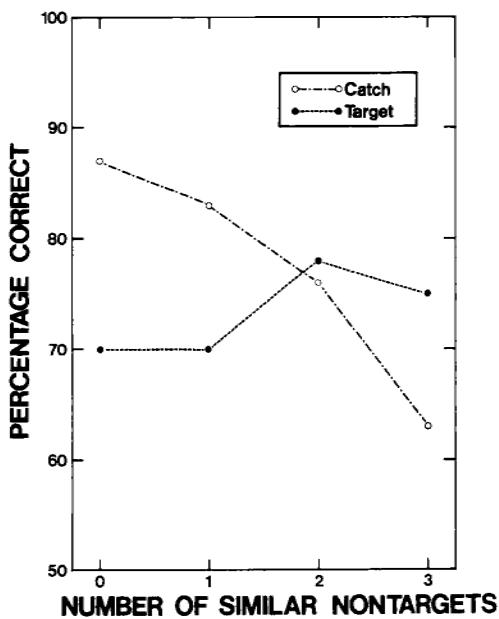


Figure 4.10 Percentage correct as a function of the number of similar nontarget letters in the test strings in the accuracy experiment of Study 1.

letters on target trials and decreased on catch trials. The large similarity effects in the accuracy task contrast with the smaller effects of Massaro et al. shown in Figure 4.5. The present analysis may have provided a more sensitive test of similarity than the Massaro et al. analysis since the confusion sets were computed in a more precise manner. In particular, the present analysis allows for asymmetric confusions. Although c is considered to be a similar nontarget letter given the target letter a, e, g, or o, the only similar nontarget letter for the target letter c is the letter o. In the reaction time task, reaction time increased with increases in the number of similar target letters on catch trials but did not change much on target trials. This result provides a direct replication of the similarity analysis in the Massaro et al. study shown in Figure 4.4.

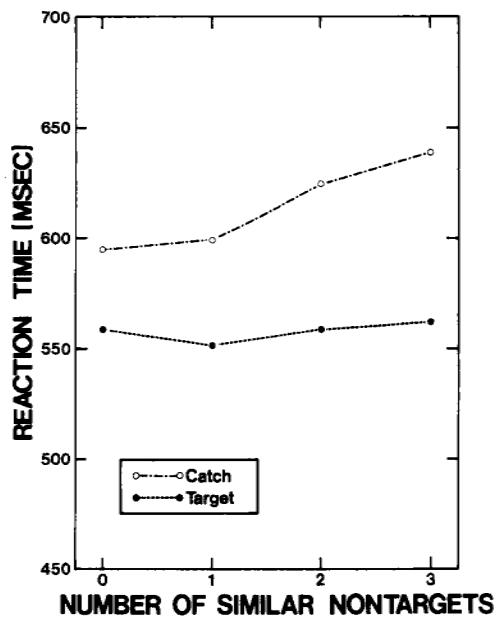


Figure 4.11 Reaction time as a function of the number of similar nontarget letters in the test string in the reaction time experiment of Study 1.

Target-Catch Differences

Accuracy was 12% higher for catch trials than for target trials and this effect interacted with display types (cf. Figure 4.8). Reaction times were 47 msec slower for catch trials than for target trials (cf. Figure 4.9). The results show that subjects in an accuracy task tend to be more accurate on catch trials than on target trials whereas they tend to take longer on catch trials in the reaction time experiment. Given that these results do not come from the same experiment, they cannot be interpreted in terms of a speed accuracy trade-off. However, response accuracy in the RT task was also significantly higher for catch than for target trials (Figure 4.9) and this result might account for the longer reaction times on catch trials in the reaction time task.

A central question raised by this study concerns the nature of the similarities and the differences between accuracy and reaction time paradigms. Results from accuracy and reaction time studies are frequently regarded as convergent indications of common perceptual processing. Systematic differences between the accuracy and RT effects would suggest that perceptual processes may not be the same in both cases. Thus, two important questions are whether or not accuracy and reaction time results can be predicted from a common processing model, and if so, how the different stimulus environments of the accuracy and reaction time tasks cause the observed differences in the magnitude of the effects. The accuracy and the reaction time tasks will be evaluated in the framework of the information processing model. In order to better understand the effects of target and catch trials and to better explicate our interpretation of the processes in the accuracy and reaction time tasks, we will present an information processing analysis of performance in the accuracy and reaction time tasks.

INFORMATION PROCESSING MODEL

With the registration of a letter string as an image upon the retina, sensory transduction begins. Features are extracted simultaneously (in parallel) from each letter position in the display and are entered into preperceptual visual storage. Feature extraction is a temporally extended process so that the contents of preperceptual visual storage

build up over time. Concurrent with the establishment of the new contents of preperceptual visual storage, the primary recognition process transforms the current contents of preperceptual visual storage into synthesized visual memory. The contents of synthesized visual memory are continually monitored by a decision process which, in the case of the precue target search task, endeavors to match the target with the freshly resolved and continuously updated letter string in synthesized visual memory, as in Turvey's (1973) concurrent-contingent model.

To derive specific predictions from the model, it is necessary to make a few assumptions about the primary recognition process and the decision process. Experimental support for these assumptions previously has been demonstrated (Massaro, 1979a). The first assumption is that knowledge of orthographic structure contributes an independent source of knowledge about the letter string. Accordingly, experienced readers will be biased to interpret the letter string as one that is orthographically structured. Therefore, to the extent that the letters being resolved from preperceptual visual storage form well-structured letter sequences, they may be resolved with fewer features or before all features have been extracted. In contrast, a poorly-structured string will require nearly complete featural information for recognition because orthographic knowledge will provide no additional assistance. Thus, the contents of synthesized visual memory are established by primary recognition more quickly and accurately for well-structured than for poorly-structured strings.

The second assumption is related to the decision algorithm in the target search task. When the target letter is known in advance, the decision process monitors the continuously emerging contents of synthesized visual memory seeking a match to the target. All letter positions are monitored independently and in parallel. The decision process maintains two criteria or decision axes, analogous to the model of memory search developed by Atkinson and Juola (1974). Whenever a letter position exceeds the first criterion in resemblance to the target, a match is assumed to have been found, processing terminates, and a positive response is executed. Whenever a letter position is sufficiently dissimilar to the target to exceed the second criterion, the letter is classified as a foil. Letter positions that have not yet been resolved sufficiently to be classified as target or foil continue to be resolved until they can be classified or until further

resolution is not possible. When all letter positions have been classified as foils, a negative response is executed.

If the visual stimulus is impoverished so that full resolution of all letter positions is not possible, a decision is forced when no further resolution is possible. Given that the fully resolved or clearly "seen" letters have been classified as foils, a guess response must be executed based upon the remaining "unseen" letters. The observer's probability of guessing that one of the unresolved letters is a target is a joint function of *a priori* expectations that any given letter will be a target and the number of unresolved letters. These expectations are derived from instructions and past experience.

For the precue RT task, the following predictions also follow directly from the model. Decision time is a direct function of the resolution time of the stimulus string and therefore is influenced by the orthographic redundancies that can be exploited in the stimulus. The time to correctly identify a target in the letter string is a function solely of the time required to resolve the target and will, on the average, be shorter than the time required to exhaustively classify all letters as foils on a catch trial. Identification time for the target letter is influenced by the orthographic constraints among the letters in the catch string. The emphasis on accuracy in the RT task disposes the observer to maintain strict criteria which, for the most part, preclude errors.

In a postcue task, recognition and decision processes analogous to those for the precue task are assumed to operate, except that the contents of synthesized visual memory must be maintained awaiting the arrival and recognition of the target letter. The stimulus letter string is resolved as fully as the combined visual information and orthographic structure allow and is retained awaiting presentation of the target. Once the target is identified, it is compared with however much of the letter string was resolved. If a match is found, a positive response is executed. Following the argument of Thompson and Massaro (1973), it is assumed that the exact featural similarity of the unresolved letters to the target letter is not directly available to the decision process and does not influence the guess probability. Orthographic structure influences accuracy for both target and catch trials because it influences the number of letters that are accurately perceived.

It may seem surprising that the Thompson and Massaro (1973) and the Massaro et al. (1979) studies revealed no

significant effects of letter similarity. Even if featural information is not transmitted to synthesized visual memory we would expect incorrect syntheses to be similar to the test letter. For example, given the test letter o, we would expect that recognition errors will be more likely to consist of c and e than t and f. One factor that may work against observation of this phenomenon may be that the test display was followed by a masking stimulus in the accuracy task and there is now some evidence that features of the mask contribute to perceptual memory of the test (Kallman & Massaro, 1979; Smith, Haviland, Reder, Brownell, & Adams, 1976). Given the contribution of the mask, incorrect responses might be more easily predicted by the features of the mask than the features of the test. A second reason for observing small similarity effects in the accuracy studies is that the accuracy studies with letter strings employ a forced-choice response situation so that those few trials that involve incorrect perceptual syntheses are unlikely to agree with the incorrect response alternative. The systematic letter confusions usually found in single-letter recognition studies could depend on an open-ended set of letter response alternatives. However, it is also possible to observe significant similarity effects as in the present accuracy study (cf. Figure 4.10). The pattern of these results is consistent with the idea that similar letters may be mistakenly identified as the target. Hence, accuracy on target trials increases with the number of similar letters to the target in the test string, but accuracy on catch trials decreases as the number of similar letters to the target is increased.

It is possible to quantify the processing model in order to describe the quantitative results of the target search task. In an accuracy task, it is assumed that the letter string is perceived (i.e., resolved) as fully as possible. Resolution of the display is a direct function of both the featural information available and the orthographic structure of the test string. Since each of the four types of meaningless letter strings was composed by making anagrams of a six-letter word, the same letters were present across the five display types. The five display types were also presented at the same intensity and durations. Therefore, the featural information may be assumed to be constant across the display types. However, the display types differ with respect to orthographic structure and these differences can account for the performance differences across the five display types. Therefore, the number of letters resolved, R , will depend on the display type.

It is necessary to describe a decision rule that allows the subject to select a "yes" or "no" response. It is assumed that if one of the resolved letters is identical to the target, a "yes" response is executed. If the target letter is not contained in the set of resolved letters, a "yes" response is a function of the number of unresolved letters and the reader's expectation, E , that some letter in the six-letter string will be identical to the target. Since E is the expectation for a whole six-letter test string containing the target, then $E/6$ is the expectation that any single letter in the test string is identical to the target. The expectation value E would be some function of instructions, feedback, and how often the target letter is actually seen in the test string. Therefore, the probability of a "yes" response to the target trial is given by Equation 1 for letter strings of length 6.

$$P(\text{Yes}|\text{Target}) = [R/6 \times 1] + [(1 - R/6) \times E/6 \times (6 - R)] \quad (1)$$

The left-hand term of Equation 1 indicates that R -sixths of the time one of the resolved letters will be identical to the target and the observer will respond "yes" with probability one. On the remaining trials, $1 - R/6$, the target was not resolved and the observer guesses with probability $E/6$ weighted by the number of unresolved letters $(6 - R)$.

For catch trials, the observer will never find the target and, therefore, is in an imperfect knowledge state unless he or she resolves all of the letters. Therefore, the likelihood of a "yes" response is simply the observer's expectation that a given test letter is a target letter, $E/6$, weighted by the number of unresolved target letters, $6 - R$:

$$P(\text{Yes}|\text{Catch}) = (6 - R) \times (E/6) \quad (2)$$

Of course, the likelihood of being correct on catch trials is the complement of Equation (2):

$$P(\text{No}|\text{Catch}) = 1 - [(6 - R) \times (E/6)] \quad (3)$$

The model was applied to the results of the accuracy experiment. The model was fit to the 10 average response accuracies (the five display types on target present trials and the five display types on catch or target absent trials). The fit was obtained using a least-squares parameter estimation routine (Chandler, 1969). Six parameters were estimated from the data; one expectation parameter, E , and an estimate of the

number of letters fully resolved, R, for each of the five display types. The obtained fit was quite good; the root mean squared deviation was only 1.3%. Although it is difficult to appreciate the quantitative goodness-of-fit for such a model estimating six parameters for 10 data points, the obtained parameter values nonetheless are revealing. The subjects were clearly informed of the equal number of target and catch trials in the experiment and therefore, they should have adopted an E-value of .5. However, the observed greater accuracy for catch trials would appear to suggest a bias toward "no" responses or an apparent E-value rather close to zero. In fact, the estimated E-value was .46 suggesting that the subjects possessed at most a very slight bias. The strong imbalance between catch accuracy and target accuracy is a logical consequence of the model even when little or no bias is present. Although there are no a priori values of the R parameters to which the estimates may be compared, the obtained estimates for R provide a reasonable scaling for the five types of items. These values are shown in Table 4.3 as Postcue (Study 1).

While the specification of a quantitative model predicting reaction times is considerably more complex, it nonetheless stems directly from the above information processing analysis. The reaction time on a target trial is the sum of the time needed to resolve the target letter in the letter string plus the time needed to initiate and execute the correct response. The resolution time for the target letter is a function of its visual properties and of the orthographic structure of the letter string in which the target occurs. Since orthographic structure affects the time needed to resolve the letter string, orthographic structure is a parameter of the distribution of the reaction times. The expected value of this distribution plus a constant gives average RT for target trials. (Response initiation and execution time is assumed to be constant or at least distributed independently of the effect of primary interest.)

For catch trials, the RT is the sum of a constant time for response execution plus the maximum time needed to classify all letters as foils. Classification time is based upon the joint distribution of completion times for each of the foils on the nontarget trials. These distributions have as parameters the similarity of the foil to the target, other visual properties which affect the overall ease of resolving the foil, the influence of orthographic structure, and finally the subject's criterion for avoiding false alarms. Having outlined the model

Table 4.3 Estimates of perceptual resolution (R) and a priori expectation (E) of the occurrence of a target trial and the root mean squared (RMS) derivations for the accuracy tasks in Studies 1 and 2.

Task	Word	R				E	RMS (in %)
		R-H	R-L	I-H	I-L		
Postcue							
(Study 1)		5.20	4.40	3.72	3.40	2.98	.459 1.27
(Study 2)		5.22	4.38	3.56	3.26	2.86	.550 1.70
Precue							
(Study 2)		4.87	4.29	3.84	3.85	3.42	.880 1.67
Combined Data		5.08	4.33	3.68	3.51	3.08	2.54
Postcue						.514	
Precue						.823	

to this level of detail, we will proceed no further. Unlike the accuracy model where the fit of the model was enlightening about presumed subject bias in the task, the qualitative features of the RT data follow rather directly from the qualitative description of this model. The completion of the quantification of this model might prove an interesting exercise but would be unlikely to be extremely revealing, particularly since the effects of orthographic structure were very small.

STUDY 2: PRECUE VERSUS POSTCUE

Whereas it is not possible to compare directly the magnitude of accuracy and RT effects, the differences in accuracy between words and irregular-low anagrams in the

accuracy task was 21%. The difference between these same conditions in the RT task was only 34 msec or 5.8% of the average RT of 581 msec. Does this represent a real difference in the utilization of orthographic structure in the two tasks? If so, what properties of the tasks allow more of a contribution of orthographic structure in one task than in the other?

The observation of significant advantages for words and pseudowords over nonwords and single letters in the postcue Reicher task (Baron & Thurston, 1973; Reicher, 1969; Thompson & Massaro, 1973; Wheeler, 1970) and the absence of these effects in precue detection tasks (Bjork & Estes, 1973; Estes, Bjork, & Skaar, 1974; Massaro, 1973; Thompson & Massaro, 1973) has led to the suggestion that precuing precludes the contribution of orthographic structure. The subject in a precue task presumably matches the features of the remembered target letter directly to the features of the incoming visual display. This strategy precludes using orthographic structure to assist in resolution of the display. Thus, one hypothesis is that precuing promotes smaller advantages for structured displays. Study 2 tests whether a precue, per se, accounts for diminished effects of orthographic structure. The postcue accuracy task of the Study 1 was replicated in one condition while a comparable precue accuracy task was used in the other condition. If precuing always allows matches to be made at a feature detection level, then the effects of orthographic structure should be severely attenuated in the precue accuracy task. On the other hand, if the display in the accuracy task is resolved to the fullest extent possible and orthographic structure contributes to resolution, then the results for precue and postcue accuracy should be comparable except, perhaps, for differential memory loads in the two conditions.

The test strings, apparatus, and display sequences were identical to those of the accuracy experiment in Study 1. One half of the subjects received the precue task on the first day and the postcue task on the second day; for the remaining subjects this order was reversed. The precue task involved the presentation of the target letter for 500 msec followed by the letter string 1 sec later. The letter string was presented before the target in the postcue task. In both tasks the letter string was exposed for a short duration and masked to give an average performance of 75% correct. Eighteen subjects were tested in the experiment. The details of the method, procedure, and results are presented in Appendix 4.2

Orthographic Structure

Figure 4.12 presents percentage correct target identifications as a function of the five levels of orthographic structure. The precue task versus the postcue task is the curve parameter. As can be seen in the figure, large effects of orthographic structure were found in both the precue and postcue tasks, although slightly larger effects were found in the postcue task. The postcue experiment replicates the previous study. The precue results are very similar to the postcue results, although the attenuation of the orthographic structure effects in the precue task over the postcue task is statistically significant.

The primary focus of this experiment was to determine the relationship of precue accuracy performance to postcue accuracy

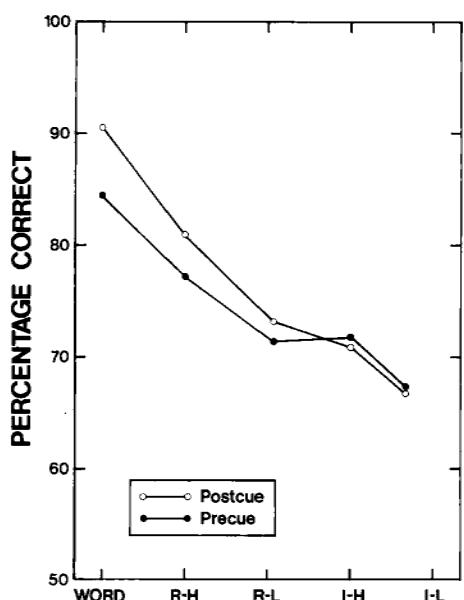


Figure 4.12 Percentage correct as a function of display type in the precue and postcue conditions of Study 2.

performance. Although the interaction of task with display conditions suggests a smaller influence of orthographic structure in the precue than postcue task, the results are sufficient to reject the hypothesis that precuing promotes a drastic reduction in the use of orthographic structure. The effects of structure were very strong for both precue and postcue, and the difference between the two tasks was small by comparison.

The observed interaction of task with display conditions probably reflects different memory requirements for the two tasks. It is reasonable to expect that performance might be poorer in a postcue task because the necessity of retaining the display until the target letter is processed could lead to memory loss. It might further be expected that forgetting would be greater for the more poorly-structured displays (Baddeley, 1964). Any differences in overall accuracy between the postcue than in precue tasks are not meaningful because the adaptive algorithm endeavored to maintain overall accuracy at 75%. In fact, subjects required significantly longer exposures in the postcue than the precue task. The longer durations for the postcue task are consistent with the idea that forgetting occurred to a greater degree in the postcue than in the precue task.

Similarity Effects

Figure 4.13 gives the percentage of correct identifications of the target letter as a function of the number of similar nontargets in the test display. Similar results were observed in both the precue and postcue tasks. The overall accuracy differences on target and catch trials for both tasks are discussed in the next section. Accuracy decreased with increases in the number of similar nontargets on catch trials but remained relatively constant on target trials. Although the similarity effects are smaller in magnitude, the catch trials but not the target trials replicate the results of the accuracy task in Study 1. The decrease in accuracy on catch trials with increases in the number of similar letters appears to be a relatively reliable result. In contrast, any effect of similar nontargets on target trials appears to be smaller in magnitude and less reliable.

Target Versus Catch Trials

Figure 4.14 presents the percentage of correct target identifications for target and catch trials in the precue and postcue accuracy tasks. Replicating the results of the previous study, overall accuracy was much greater on catch trials than on target trials in the postcue task. In contrast, performance on target trials was more accurate than on catch trials in the precue accuracy task. These effects can be described by the decision rule in the model. For accuracy tasks it is assumed that the letter string is resolved as fully as possible with the assistance of orthographic structure. If none of the resolved letters is the target, a "yes" response is a function of the observer's a priori expectation probability that an unresolved letter is a target and the number of

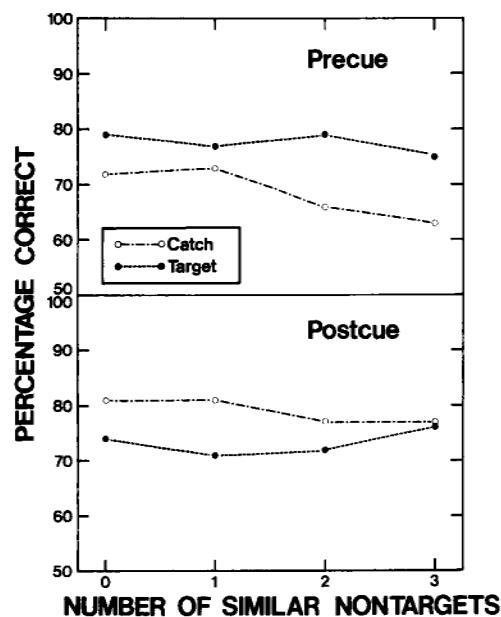


Figure 4.13 Percentage correct for target and catch trials in the precue and postcue conditions of Study 2 as a function of the number of similar nontarget letters in the test string.

unresolved letters. In terms of the present model, the expectation parameter E/6 must have differed for the precue and postcue tasks.

Information Processing Model

The model was tested by fitting it to the data of the present experiment. Six parameters were estimated for each set of 10 means: one E-value, and a separate R-value for each of the five display conditions, e.g., word, regular-high, etc. The parameter estimates and their associated root mean squared deviations are shown in Table 4.3 as Postcue (Study 2) and Precue (Study 2). The model provides a good description of the results of both tasks.

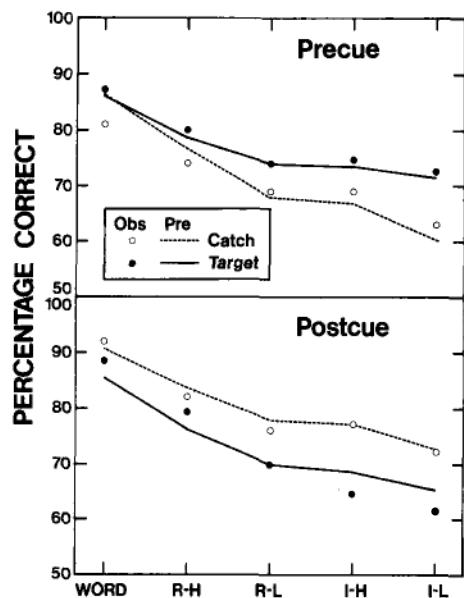


Figure 4.14 Observed (Obs) and predicted (Pre) percentage correct for target and catch trials for the precue and postcue conditions of Study 2 as a function of display type.

Not only does the fit of the model provide support for our conceptualization of the decision mechanism, but the parameter values are meaningful. It is interesting to note that the apparent bias toward "no" responses in the postcue task actually reflects an accurate and unbiased assessment of the equal probability of target and catch trials. The subjects in the precue task, however, do exhibit a marked bias toward expecting the presence of a target. Smith et al. (1976) have reported similar findings in a precued target recognition task and suggest that observers in precue tasks may erroneously incorporate mask features in making their judgments and thus tend to see the target in the mask more often than might be expected. By manipulating the nature of the mask which follows the display they found a more pronounced effect when the display and mask shared features than when they did not.

Perhaps even more interesting than the E-parameters were the R-values which correspond to the number of letters actually resolved for each display type. The parameter values for R as a function of display type change systematically with orthographic structure across both the precue and the postcue experiments. The larger effect of display type for the postcue than for the precue task is reflected by the larger spread of the R-values for the postcue task.

Although the interaction of display conditions with task in the present experiment implies that R-values were somewhat different in the precue and postcue tasks, all 20 data points were also fit with seven parameters by assuming that a single set of five R-values and two E-values, one for precue and one for the postcue, would be sufficient. These parameters and their associated root mean squared deviations are also given in Table 4.3 as Combined Data. The predicted results of the seven-parameter model are shown along with the observed results in Figure 4.14. The reasonably good description of the result shows that the combined fit of the two tasks is quite acceptable. In addition, reducing the number of parameters from 12 to 7 increased the total root mean squared deviation by less than one percent (from 1.6% to 2.5%).

STUDY 3: MIXED-CASE

The results of the comparison between precue and postcue tasks indicates that precuing alone is not a sufficient condition to cause a drastic reduction in the use of

orthographic structure, at least in accuracy tasks. It still might be the case, however, that precuing in the RT task in which unlimited viewing time insures that all of the features of the display can be extracted promotes a feature detection strategy. The next experiment addresses the issue of whether or not discouraging a feature detection strategy will promote the utilization of orthographic structure in the resolution of the test string.

Much of the intuitive appeal of the feature detection strategy is the idea that recognition of the target letter could occur on the basis of only a few features and without the necessity of fully resolving all features of the letter. The effectiveness of this strategy would be reduced if the specific visual configuration of the target, and thus the exact form of the critical features, could not be predicted. One procedure would be to have the letter string appear in several different fonts with the font of the target appearing on a given test trial being unpredictable. The present experiment approximated this method by presenting the displays printed half in uppercase letters and half in lowercase. Thus, the letter case of the target letter was not completely predictable. Whereas it might still be possible to detect features for uppercase and lowercase letters simultaneously, the extra effort should reduce the effectiveness of the feature detection strategy. Therefore, it is predicted that if the feature detection strategy is the cause of the small effect of orthographic structure in a precue RT task, then a significantly larger effect should be observed in the mixed-case condition, where the effectiveness of the feature detection strategy will be diminished.

Taylor, Miller, and Juola (1977) have reported that mixing cases disrupts the utilization of orthographic structure in word perception. The amount of disruption was found to be a monotonically increasing function of the number of transitions between cases in the string to be perceived. To minimize the disruptive influence of case transitions on the utilization of orthographic structure, only a single transition between the third and fourth letters was used in this experiment.

The procedure was the same as the earlier reaction time experiment with the following modifications. The target letter was presented at the beginning of each trial in both its uppercase and lowercase forms. On each trial, the test string was presented so that either the first three letters were in uppercase and the last three letters in lowercase or the

opposite occurred. The details of the method, procedure, and results are presented in Appendix 4.3.

Orthographic Structure

Figure 4.15 presents the mean reaction times in the task as a function of orthographic structure. There was only a 7 msec advantage of words over regular-high pseudowords. The 30 msec advantage of regular over irregular strings was significant as was the 14 msec advantage of strings high in positional frequency relative to strings low in positional frequency.

By mixing the cases of the letters in the present study, the overall effect of orthographic structure increased from 34 msec in the reaction time experiment of Study 1 to 51 msec in

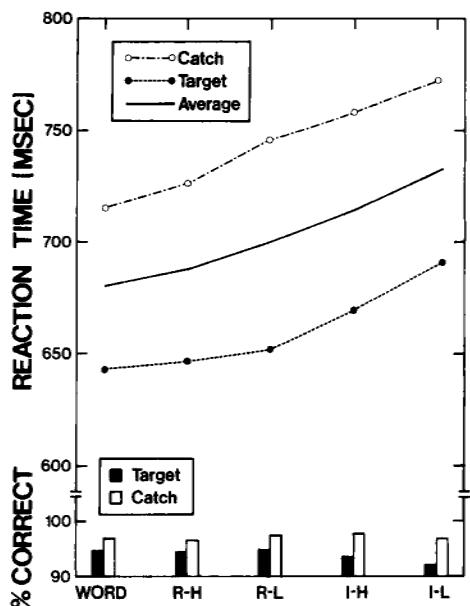


Figure 4.15 Reaction time and percentage correct as a function of display type in Study 3: Mixed-Case.

the present mixed-case reaction time study. Mixing the cases increased the overall reaction time by about 120 msec. It is, therefore, necessary to ask whether the 17 msec increase for the effect of orthographic structure is meaningful. If we can assume that the 120 msec increase in RTs to mixed-case displays represents the additional visual processing required, then the 17 msec increase in orthographic structure effects may provide a meaningful evaluation of the influence of orthographic structure in the RT task. The increase in the effect of orthographic structure relative to the additional visual processing required is 14%. Even 14%, however, may be an underestimate of the size of the effect of orthographic structure since the additional 120 msec of processing required by mixed-case stimuli may include additional time needed for the comparison process when the target can occur in either uppercase or lowercase.

Similarity Effects

If mixing the cases attenuated the feature detection strategy then we would expect a decrease in the effects of the number of similar nontarget letters in the display on catch trials. The similarity analyses are complicated by the mixture of upper- and lowercases in the string. Because the confusion sets were derived from confusions of lowercase letters, application of these confusion sets to the data without regard to letter cases would not be appropriate. Therefore, a restricted similarity analysis was done by examining only the lowercase letters in the display strings. When this was done, there were only two viable similarity categories: zero and one similar letters. For target trials the means were 658 and 658 msec for the zero and one categories respectively. For catch trials these means were 731 and 756 msec. These results replicate the RT task in Study 1; RTs for catch trials but not target trials increase with the number of similar nontargets in the test string. Based on the similarity analyses, mixing the cases did not appear to attenuate the feature detection strategy. The experiment provides mixed evidence for the hypothesis that a feature detection strategy occurs in precue RT tasks. The effect of orthographic structure was increased somewhat by mixing the cases and it might have been concluded that use of a feature detection strategy was precluded and more use made of orthographic context. However, the failure to attenuate similarity effects precludes any firm conclusion about the feature detection strategy. Either the subjects

persisted in the now more difficult task of looking for feature matches in two-letter cases or feature detection is not the correct explanation for the results of either of the two experiments.

STUDY 4: LIMITED VIEWING TIME

In another attempt to discourage the feature detection strategy, subjects were not given unlimited viewing time in the reaction time task of the present experiment. In the RT task (Studies 1 and 3), the subjects were given unlimited time and hence all of the visual information necessary for detailed perceptual resolution of the stimulus displays. In the accuracy task (Studies 1 and 2), viewing time was shortened so that visual information was inadequate to clearly resolve all of the display. Decisions at a feature detection level may depend on the certainty that all the visual information will be resolved. If so, limiting viewing time in the RT task should increase the size of orthographic structure effects. Study 4 tested this hypothesis by giving the subjects the display for a limited time. However, it was not possible to limit viewing time as severely as in the previous accuracy tasks and still maintain error rates low enough to permit meaningful interpretation of RTs. Study 4 was a hybrid; viewing time was shortened and error rates allowed to rise but neither was taken to extremes.

The stimuli and apparatus were the same as in the reaction time experiment of Study 1. Rather than leaving the stimulus display on until the subject responded, the display was presented for only 200 msec and followed immediately by a masking stimulus. On the basis of the pilot work, this duration was chosen because it not only limited viewing time but also gave the subject sufficient visual information to respond reasonably accurately. Sixteen subjects were chosen from the same population as the earlier studies. All other details are given in Appendix 4.4.

Orthographic Structure

Figure 4.16 presents the reaction time for correct responses as a function of orthographic structure. As can be seen in the figure, no effects of orthographic structure were observed. Although orthographic structure did not have a significant effect on reaction time, it did affect

accuracy. The small differences in accuracy as a function of orthographic structure were found to be statistically significant. By limiting the display duration, it appears that the experiment begins to approximate an accuracy experiment, and the smaller effects of orthographic structure can simply be attributed to the overall high level of performance.

Similarity Effects

Although accuracy varied slightly with differences in orthographic structure, there were no significant changes in accuracy as a function of the number of similar nontargets in the test display. In contrast, the RTs continued to reveal the consistent finding that RTs on catch trials increased with increases in the number of similar nontargets whereas RTs on target trials showed no effect. Accordingly, the results show

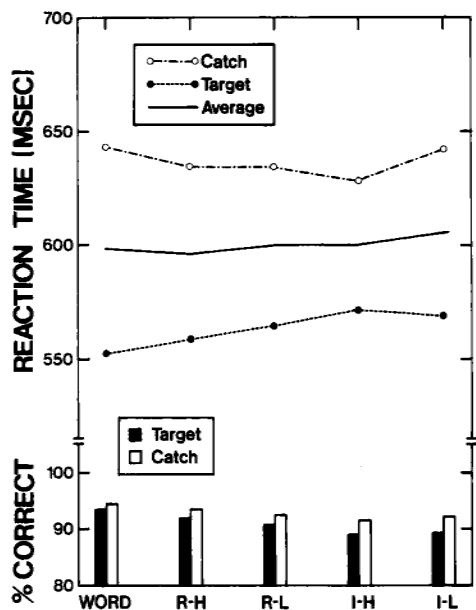


Figure 4.16 Reaction time and percentage correct as a function of display type in Study 4: Limited Viewing Time.

that an orthographic structure effect in terms of response accuracy can coexist with a similarity effect on RT. This result reveals that a featural detection strategy as indexed by significant similarity effects does not preclude a significant contribution of orthographic structure. It appears that orthographic structure effects can be found in both accuracy and RT measures, but that RT effects are not as reliable or sensitive as accuracy effects.

Target Versus Catch Trials

Target trials (563 msec) were responded to more quickly than catch trials (635 msec) whereas there was no significant difference between the two kinds of trials in terms of response accuracy.

Discussion

The results disconfirm the hypothesis that limiting viewing time might increase the influence of orthographic structure on RT. Reaction times were found to be uniform over conditions. However, the pattern of error rates reflected the influence of structure. Limiting visual information clearly has an influence on accuracy, and this influence is modulated by the presence of orthographic structure. As for the unusually uniform RTs, they might reflect the subjects having established and maintained a self-imposed response deadline. Therefore, any effect of orthographic structure would be reflected only on performance accuracy.

STUDY 5: SPEEDED RT TASK WITH GOOD VISUAL INFORMATION

Study 5 tests between two hypotheses of the differences in the sensitivity of the accuracy and RT tasks to orthographic structure. According to the first hypothesis, the difference between the accuracy and RT tasks could be due to a trade-off between the contribution of stimulus information and orthographic structure in the task. There is abundant stimulus information in the RT task, and therefore, there may be very little opportunity for knowledge about orthographic structure to contribute to performance. In the accuracy task, on the other hand, the stimulus information is inadequate and consequently, there is considerable opportunity for

orthographic knowledge to exert an influence. This explanation is clearly different from the feature detection strategy discussed earlier. In the case of a feature detection strategy, there is an opportunity for orthographic structure to exert an influence but a decision about the letter string is made without utilizing the structure. By contrast, in the case of a trade-off between visual information and orthographic structure, the complete visual information precludes a contribution of orthographic structure.

Evidence for the trade-off between the contributions of stimulus information and knowledge comes from a number of RT experiments that have independently varied stimulus quality and knowledge (c.f., Meyer, Schvaneveldt, & Ruddy, 1975). Massaro, Jones, Lipscomb, and Scholz (1978), for example, had subjects name a word as quickly as possible. The word was either presented upright or was rotated 180 degrees in the picture plane. The word was preceded by a word "blank" or by a category prime, which was the superordinate category name of the test word. Priming with the category name is the advance knowledge in the task and it might be expected to facilitate processing of the test word. Therefore, RTs should be faster on primed than unprimed trials. Of course, rotating the test words should slow down the naming response since the stimulus information is degraded; we are not used to seeing and reading rotated text. The results showed that priming had a nonsignificant 7 msec facilitation when the words were presented upright but a significant 121 msec effect when words were rotated 180 degrees. Similar results were found in a lexical decision task. With a good quality test display, priming had essentially no effect; given a poor quality display, advance knowledge facilitated performance. In the present experiments, given a good quality display in the RT task, orthographic structure had a small effect; with a poor quality display in the accuracy task, orthographic structure had a large effect.

The second hypothesis is that the contribution of orthographic structure is similar in the accuracy and RT tasks but the accuracy task allows a more direct measure of the size of this contribution. In the RT task where the visual display is clearly visible and viewing time unlimited, the letter string is resolved very quickly and the time to resolve the letters may be small relative to the times for the remaining processes of search, comparison, decision, and response selection. Thus, the facilitating effect of orthographic structure might be large in terms of the time savings for

visual resolution, but small relative to the overall RT. For example, a 20 msec effect of structure when visual resolution takes 100 msec without any contribution of structure would expand a 500 msec RT by only 4%. Even though this 20% effect on visual resolution is important, it would appear small relative to the overall RT. By contrast, the visual information in the accuracy task is limited and errors are a direct consequence of limitations in visual resolution. Accordingly, a facilitation of visual resolution will be more directly reflected in an improvement in task accuracy. According to this hypothesis, even though orthographic structure contributes roughly the same additional information in both the RT and accuracy tasks, the size of RT effect will appear small relative to the accuracy effect.

In order to test between these two hypotheses, an experiment was carried out in which the reader would have good quality visual information but accuracy rather than RT would be the primary dependent variable. If the contribution of orthographic structure is inversely related to the quality of the visual information as predicted by the trade-off hypothesis, then a very small effect of structure should be apparent on accuracy. However, if the contribution of structure is independent of the quality of the visual information and accuracy is simply a more direct measure, then the typical large effects of structure on accuracy should be observed. Given that accuracy is now tied to 75% correct, it should be a more direct index of the contribution of structure. Therefore, the present experiment provides a test between the two hypotheses without having to compare accuracy effects to RT effects. A modification of a speed-accuracy trade-off paradigm which requires a speeded reaction in synchronization with a signal event (Reed, 1973) was used to keep average performance at 75% correct. The subjects were given the target letter followed by the display and were required to respond as nearly in synchrony with the offset of the display as possible. Using an adaptive algorithm, a display duration was calculated to keep the subject's overall accuracy at 75%. Because the duration of the display averaged around 500 msec, it seems safe to assume that the subjects have good quality stimulus information. Therefore, the trade-off hypothesis predicts no effect of structure whereas the hypothesis that accuracy is a more direct measure than RT predicts a large effect of structure.

Procedure

Eleven students, selected from the same population as those in the earlier experiments, were tested on a single day. A target letter was presented for 500 msec and followed by the test display after an interval of 250 msec. The test display duration was modified every 20 trials throughout the experiment to keep overall performance at 75% correct. Subjects were instructed to respond as close to the stimulus offset as possible and they were given feedback on this task after each block of 20 trials. Essentially, subjects simply learned to maintain a particular response time, not to respond to the offset of the display itself. None of the subjects had any difficulty with this task. Details of the method, procedure, and results are given in Appendix 4.5.

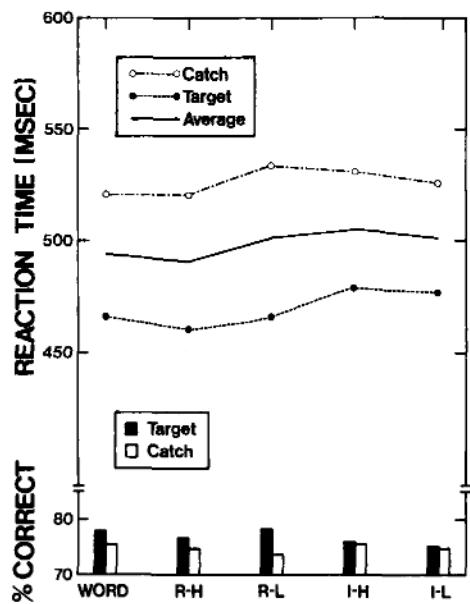


Figure 4.17 Reaction time and percentage correct as a function of display type in Study 5: Speeded RT Task with Good Visual Information.

Orthographic Structure

Figure 4.17 presents the correct reaction times and response accuracy as a function of orthographic structure; target and catch trials are plotted separately. There was a 15 msec effect of orthographic structure on correct reaction times and a 1.9% effect on accuracy. Therefore, no strong effect of orthographic structure on either RT or accuracy was present. This result is consistent with the trade-off hypothesis whereby the contribution of orthographic structure is inversely related to the quality of the visual information in the test string.

STUDY 6: SPEEDED RT TASK WITH POOR VISUAL INFORMATION

The speeded RT task in Study 5 provided evidence that the accuracy task may give larger effects than does the RT task because of the poorer quality of the visual information in the accuracy task. Accordingly, giving good quality information in the speeded RT task shows no effect of orthographic structure on response accuracy. Before accepting this conclusion, it is important to verify whether the speed-accuracy paradigm that was used is itself sensitive to orthographic structure. It could be the case that the perceptual stage goes to completion in all conditions and that errors are simply a function of later stages such as response selection. Any small differences in completion times of the perceptual stage as a function of orthographic structure may not significantly influence the accuracy of later stages of processing, and therefore, the differences in the perceptual stage would not be reflected in either response accuracy or reaction time. To test the sensitivity of the speeded RT task, it is necessary to replicate the task with poor quality information. If the task is sensitive and the effect of structure is inversely related to stimulus quality, then significant effects should be found. If the task is insensitive to the perceptual stage, however, no effect of orthographic structure should be observed.

As discussed previously, Massaro et al. (1978) degraded the quality of test strings by rotating them 180 degrees in the picture plane. Following this approach, the speeded RT task was replicated with the test strings rotated in the same manner. Given that rotation lowers the quality of the visual

information, an effect of orthographic structure should be observed if the speeded RT task is sensitive to perceptual differences due to the orthographic structure. If the task itself is insensitive then no effects of structure should be observed.

Procedure

The speeded RT task in Study 5 was replicated exactly with a new group of students except that the test strings were rotated 180 degrees in the picture plane rather than presented upright. The target letter was still presented upright. The details of the method, procedure, and results are presented in Appendix 4.6.

Orthographic Structure

Figure 4.18 presents the correct reaction times and accuracy performance. There was a 7 msec effect of structure

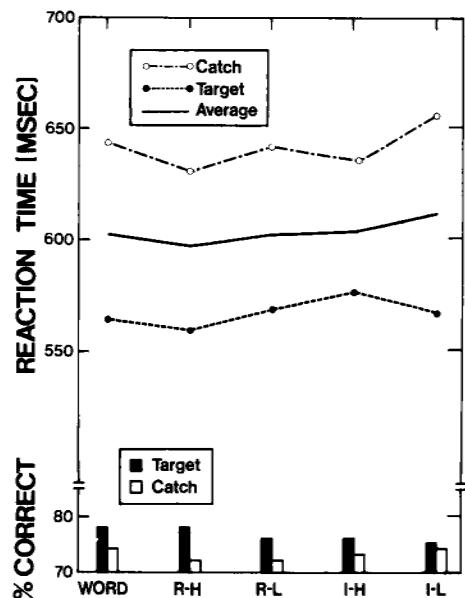


Figure 4.18 Reaction time and percentage correct as a function of display type in Study 6: Speeded RT Task with Poor Visual Information.

on correct reaction times and a 1.7% effect on response accuracy. The results of this experiment are a direct replication of those in the previous speeded RT task and therefore provide strong evidence for the idea that the speeded RT task is insensitive to perceptual differences due to the orthographic structure. Before this conclusion is accepted, however, it will be necessary to verify that orthographic structure effects do occur with rotated test strings in other tasks. In order to have a directly analogous situation to the speeded RT tasks, the task must keep the test string in view during the processing period. Therefore, high accuracy is to be expected and RT must be the dependent measure.

STUDY 7: HIGH-ACCURACY RT TASK WITH POOR VISUAL INFORMATION

Procedure

To verify that orthographic structure effects do occur with rotated test strings, the high-accuracy reaction time experiment of Studies 1 and 3 was replicated but using rotated rather than upright test strings. An upright target letter was presented followed by a rotated test string which remained in view until the last of up to four subjects responded. Subjects performed as quickly as possible while maintaining high accuracy. The details of the method, procedure, and results are presented in Appendix 4.7.

Orthographic Structure

Reaction times generally increased from words to the irregular-low strings. Figure 4.19 shows that overall effect was 32 msec, which is highly similar to the effects of structure on normally presented strings (cf. Figure 4.9). This result shows that structure effects do occur with rotated stimuli but that they are not larger than the structure effects observed with upright stimuli. The idea that a trade-off exists between stimulus quality and orthographic structure is not supported by these results. It appears that both accuracy tasks and high-accuracy RT tasks can reveal orthographic structure effects but that the speeded RT task does not.

SUMMARY OF ACCURACY AND REACTION TIME EXPERIMENTS

These experiments tend to support the idea that perceptual resolution of a letter string is facilitated to the degree the letters in the string are well-structured. Both accuracy tasks with limited stimulus information and high accuracy RT tasks with complete stimulus information appear to be sensitive to the effects of orthographic structure. However, the magnitude of the RT effect appears smaller because of the relatively large overall RT due to the processing time required for other than visual processing stages in the RT task. In contrast, average performance in the accuracy task is titrated to remain at 75% correct and this percentage measure is influenced almost completely by the visual processing stages. The post hoc evaluation of various measures of orthographic structure in

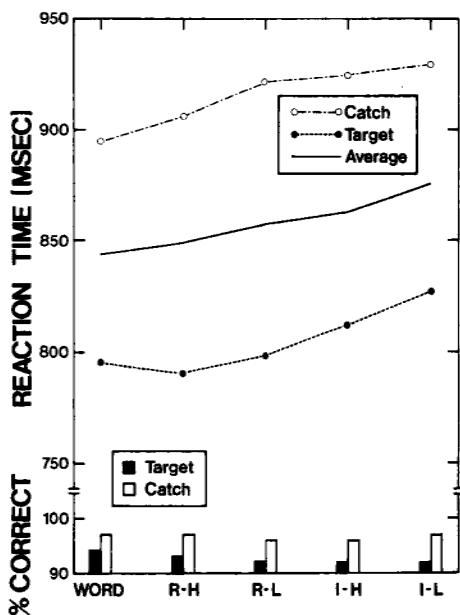


Figure 4.19 Reaction time and percentage correct as a function of display type in Study 7: High-accuracy RT Task with Poor Visual Information.

Chapter 5 will allow a more direct assessment of the sensitivity of performance in the RT and accuracy tasks.

A second important contribution to performance is the featural information contained in the visual display. Performance in both the accuracy and RT task is sensitive to the visual similarity of the test letters to the target letter. Performance on catch trials is disrupted to the extent that the test display contains nontarget letters that are similar to the target letter. Performance on target trials is not significantly affected since processing is primarily a function of target letter processing time. However, there was some hint that similar nontargets may be incorrectly confused for the target on target trials and, therefore, lead to a correct response even though the target itself was not recognized on that trial.

In conclusion, both the accuracy and high accuracy RT tasks appear to offer good measures of the utilization of orthographic structure in visual processing. The speeded RT task does not reveal effects of orthographic structure. Before evaluating specific descriptions of orthographic structure, some rating experiments were carried out to assess to what extent knowledge of orthographic structure is consciously available and can be reported.

OVERT JUDGMENTS

Whereas the accuracy and RT experiments imply that visual recognition is sensitive to the presence of orthographic structure, it is also of interest to determine to what extent subjects are aware of this knowledge. For this purpose, several overt judgment experiments were conducted. The same 200 stimulus items in Appendix 4.1B were presented to native English speaking college students and judgments in the form of ratings and paired judgments were obtained. Answers to the following questions were sought. Could the raters reliably assign ratings of orthographic structure to the items? Would the dimensions utilized by the raters discriminate the items along our regularity dimension, by positional frequency, or both? Would the ratings assigned to the items show any meaningful relationship to their perceptual accuracy and reaction time results found in the previous experiments?

STUDY 8: TYPICALITY RATINGS

Three sets of ratings were obtained. In the first one only the 160 meaningless anagrams were rated. In the second, the words were transformed into the most regular pseudowords possible by the substitution of a single letter and these pseudowords were rated along with the anagrams. In the third, the words themselves were included along with the anagrams. These three procedures were used to obtain the most reliable ratings possible and to determine how alterations in the total set of items rated affected the ratings of the anagrams. All three ratings were obtained under nearly identical instructions which asked the raters to base their judgments on the similarity of the items to English words where similarity was defined as typicality of English spelling. Beyond these general instructions, no specific attempts were made to illustrate either specific regularity rules or positional-frequency constraints.

Subjects were asked to rate how much the letter strings look like English words. The rating scale went from "most like English" to "least like English" on a scale from 10 to 1. College students rated the items in one of three conditions. Seventy subjects rated only the 160 anagrams. Sixty-five subjects rated the 160 anagrams plus 40 pseudowords which were derived by misspelling one letter in each of the 40 word stimuli. Sixty-three subjects rated the 160 anagrams and the 40 words. Subjects were urged not to adopt strategies of trying to see the items as similar to specific words. Those subjects that received actual words were told to disregard the semantic content of the words and simply judge them on the basis of their similarity to English spelling. The method, procedure, and detailed results are presented in Appendix 4.8.

Figure 4.20 shows the average ratings for the items for each of the three groups of subjects. The ratings assigned to the 160 anagrams were very stable across the three rating contexts. The six correlations between all possible pairs of ratings across the three contexts were all .97. This result demonstrates that the ratings of the anagrams were not influenced by the presence of actual words in the sample of items to be rated. The figure also shows that the ratings were primarily determined by the regularity of the items rather than by positional frequency. Although the effect of positional frequency was statistically significant, the results show that

regularity produced about seven times the effect that positional frequency did.

The results indicate that when subjects are asked to rate items on the basis of the similarity to English spelling, they tend to evaluate the items in terms of notions of regularity rather than in terms of the positional frequency of the items. As can be seen in the figure, the words were rated as more typical than even the regular-high anagrams. Even the pseudowords created by misspelling the words by a single letter produced higher ratings than the regular anagrams. This result means that the word stimuli and their misspellings were either more typical according to the subject's notion of English spelling or that the lexical content of the words and possibly their misspellings could not be ignored in making the judgment.

If the ratings and the target search experiments are measuring the same knowledge of orthographic structure, then a

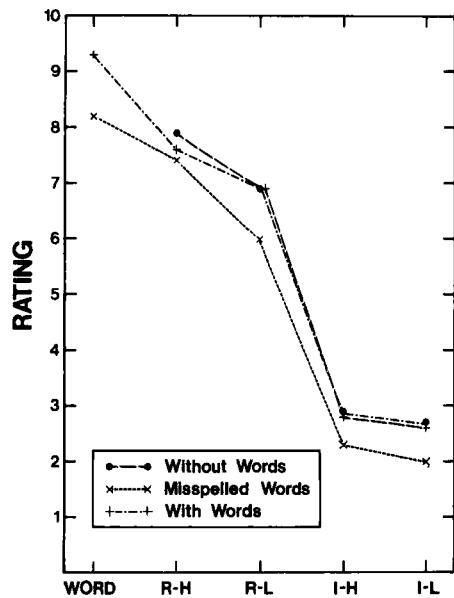


Figure 4.20 Ratings from 10 (most like English) to 1 (least like English) as a function of display type in Study 8: Typicality Ratings.

close relationship between the ratings and the RT and accuracy measures of the previous experiments should be observed. Accordingly, we examined the relationship between these ratings and the perceptual performance data collected in Studies 1, 2, and 3. Average RTs were obtained for each of the 200 stimulus items by averaging the correct response RTs for target and catch trials from the reaction time experiment of Study 1 and the mixed-case experiment of Study 3. This provided an average RT for each item based on approximately 204 observations. Similarly, an average accuracy was found for each item based on the target and catch trials for the subjects in the accuracy experiment of Study 1 and the precue versus postcue experiment of Study 2. The average accuracy of each item was based on 150 observations.

Correlations between the three ratings, average RT, and average accuracy are shown in Table 4.4. Two things may be observed in Table 4.4. First, the correlation between the accuracies and RTs of all items was .56. Therefore, despite the magnitude of the observed differences between RT and accuracy measures, both experimental procedures appear to reveal a common processing component. This will become even more evident in the correlational analyses (to be discussed in Chapter 5) which demonstrate that both item accuracy and RT tend to show similar correlations with various measures of orthographic structure.

The correlations between item accuracy, RT, and the ratings are illuminating. Correlations between the two perceptual measures and the ratings are quite high. The perceptual measures and the ratings share 31% to 53% common variance. Whereas the knowledge structures tapped by the rating procedures need not necessarily relate to the knowledge structure inferred from the perceptual tasks, it is encouraging to observe that there does appear to be a lawful correspondence. It would appear that what one knows about orthographic structure not only influences perception but also is amenable to interrogation by conscious processes. A second encouraging aspect of this observation is that rating tasks and perceptual tasks may be regarded as converging measures of orthographic structure. If both tap a common data base of orthographic structure knowledge, then we can refine our understanding of this knowledge source by means of rating tasks as well as perceptual tasks. For the correlational analyses in Chapter 5, item accuracies, RTs, and ratings are each regarded

Table 4.4 Correlations among average accuracy, average RT, and the three regularity ratings for the 200 stimulus items.

	(1)	(2)	(3)	(4)	(5)
(1) Average accuracy	--				
(2) Average RT	-.56	--			
(3) Rating without words	.58*	-.56*	--		
(4) Rating with pseudowords	.72	-.61	.97	--	
(5) Rating with words	.73	-.60	.97	.97	--

* Correlation based on the 160 anagrams only, all other correlations based on all 200 stimulus items.

as related collateral dependent measures potentially revealing a common orthographic structure component.

STUDY 9: POSITIONAL-FREQUENCY RATINGS

The ratings of Study 8 indicate that subjects, when given relatively neutral instructions, base their ratings more on regularity than on positional frequency. Before we can conclude that the reader's conscious knowledge is best represented by regularity, it is necessary to determine whether subjects can be instructed to respond on the basis of positional-frequency information. Accordingly, the rating experiment with only 160 anagrams was replicated but now the subjects were given specific positional-frequency instructions. The instructions explained the concept of positional frequency along with relevant examples and told the subjects to rate the items explicitly on this basis. The exact instructions,

method, and procedure are given in the Appendix 4.9. Except for the instructions, the procedure replicated the previous rating experiment that gave instructions with respect to typicality of English spelling. The results and detailed analyses are given in Appendix 4.9.

Figure 4.21 gives the results of this experiment stressing positional frequencies along with the results of the previous experiment stressing typicality. The results show that the rating responses were very similar with both sets of instructions. However, positional frequency did make more of a contribution with positional-frequency instructions than with typicality instructions. Overall, subjects tended to rate high positional-frequency items as slightly higher than low positional-frequency items to a greater degree with positional-frequency instructions than with typicality instructions. Also, the effect of regularity on the judgments

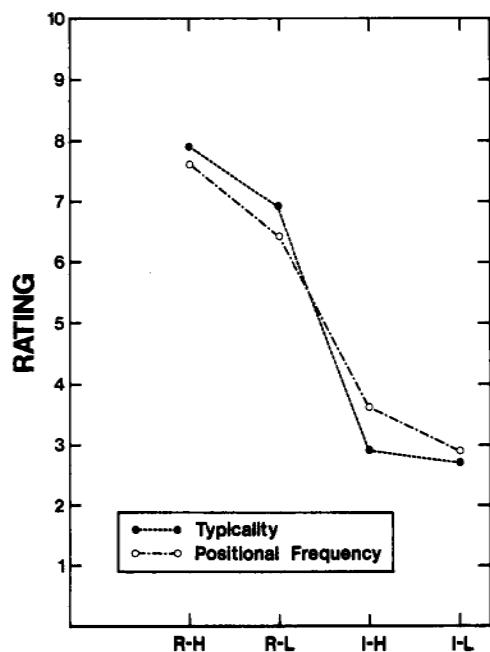


Figure 4.21 Ratings from 10 (most like English) to 1 (least like English) as a function of display type for typicality instructions (Study 8) and positional-frequency instructions (Study 9).

was slightly attenuated with positional-frequency instructions relative to typicality instructions.

The rating experiments indicate that orthographic structure in terms of regularity appears to be more consciously available than structure in terms of positional frequency. Subjects tend to rate regular items as most like English and irregular items as least like English regardless of whether they were given typicality instructions or positional-frequency instructions. One limitation in the rating experiments may be that subjects tend to rate the items on the basis of regularity regardless of the instructions that are given. Therefore, a more direct index of the psychological reality of these two concepts as descriptions of orthographic structure should be obtained.

STUDY 10: REGULARITY VERSUS POSITIONAL-FREQUENCY JUDGMENTS

In order to obtain a more direct index of orthographic structure, precise regularity instructions were contrasted with precise positional-frequency instructions. Items from each of the four types of 160 anagrams were paired with each other such that items of one type were paired with items of the same type or with items of the other types. Some subjects were given pairs of items and asked to pick the item from each pair that was most regular. Other subjects were given the same pairs and asked to pick the item that was highest in terms of positional frequency. The instructions, method, procedure, and results are given in Appendix 4.10. This experiment allowed a direct assessment of the relative weight carried by orthographic regularity and positional frequency in the reader's evaluation of orthographic structure. The judgments of these two groups of subjects were tabulated as the proportion of times they chose an item from one category when it was paired with an item from another category. These results are presented as Table 4.5 where the cells of the table give the proportion of times an item from the column category was selected when paired with an item from the row category.

Before discussing the actual results in Table 4.5, it is instructive to hypothesize how this table should have turned out if the subjects in the two conditions had a complete knowledge of regularity and positional frequency and had applied this knowledge in accordance with the instructions. Whenever the two items differed on the dimension relevant for

Table 4.5 Paired-judgment preferences for the 200 stimuli under regularity and positional-frequency instructions. Cells represent proportion of times the column item was chosen over the row item. The numbers in parentheses give the hypothetical outcomes assuming that the subject had knowledge of both dimensions and followed instructions.

Regularity Instructions				
	R-H	R-L	I-H	I-L
R-H	.475 (.5)			
R-L		.693 (.5)	.445 (.5)	
I-H		.910 (1.0)	.877 (1.0)	.523 (.5)
I-L		.938 (1.0)	.908 (1.0)	.509 (.5) .518 (.5)

Positional Frequency Instructions				
	R-H	R-L	I-H	I-L
R-H	.527 (.5)			
R-L		.704 (1.0)	.488 (.5)	
I-H		.866 (.5)	.740 (.0)	.525 (.5)
I-L		.953 (1.0)	.895 (.5)	.687 (1.0) .494 (.5)

Table 4.6 Chi-square for the two predicted outcomes against the observed outcomes for the two instruction conditions in study 10.

Instructions	Predicted Outcome	
	Regularity	Positional Frequency
Regularity	93.63	1561.26
Positional Frequency	258.90	1367.50

the judgment, the subjects should have always chosen the "correct item," on that dimension, i.e., the one that was more regular or the one that was higher in positional frequency. Whenever the two items were at equivalent levels on the relevant dimension, the subjects should have chosen indifferently between them. These "perfect" hypothetical outcomes are also shown in Table 4.5.

Comparing the two panels of Table 4.5 we observe that for both sets of instructions, subjects made choices more consistent with the application of regularity than with the application of positional frequency. In order to assess how well the results conformed to what would be expected from perfect knowledge and utilization of regularity or positional frequency, the two predicted outcomes were compared with the actual outcomes under both instruction conditions. Table 4.6 gives the Chi-square values for comparisons of the two predicted outcomes against the observed outcomes for both instruction conditions. The values support the interpretation that the subjects' performance is more adequately described by regularity than positional frequency. However, though subjects appear to be using mostly regularity as a basis of their ratings, in neither case did the ratings indicate that only regularity was used. If we compare the two panels of Table 4.5 to each other, we see that there was an additional contribution of positional frequency when subjects were explicitly instructed to base their judgments on frequency. Evidently, subjects can modify their judgments slightly when given explicit instructions to attend to positional frequency.

SUMMARY OF OVERT JUDGMENT EXPERIMENTS

The experiments support the idea that knowledge of orthographic structure is consciously available and capable of report. Subjects are relatively good at assigning ratings on the basis of rule-governed regularity but relatively poor at rating items on the basis of statistical redundancy. Similarly, choosing which of a pair of items is more regular is relatively easy whereas choosing which is more frequent is nearly impossible. The post hoc correlations in Chapter 5 will allow a more direct and broader assessment of the nature of the knowledge utilized in overt judgments of letter strings.

Appendix 4.1

Details of Method, Procedure, and Results of
Study 1: Initial ReplicationMethod

Subjects. One group of 14 Introductory Psychology students participated in the postcue accuracy task for an hour a day on each of 2 consecutive days. The data from three additional subjects were lost due to computer system failures. Six other subjects were tested but their data were not included because of inadmissible overall accuracy. Another group of 14 Introductory Psychology students participated for an hour each day on 2 consecutive days in the precue RT task. All subjects in both tasks passed a preliminary vision test establishing at least 20/40 acuity for both eyes. All subjects were awarded research participation credit toward their psychology grade.

Stimuli. A listing of the 100 highest and 100 lowest summed single-letter positional-frequency anagrams were obtained for several hundred common six-letter words selected from the Kucera and Francis (1967) word list. The listings of the anagrams were produced by a computer program which generated all 720 possible permutations for each six-letter word. The positional frequency for each permutation was computed from the Mayzner and Tresselt (1965) single-letter counts for six-letter words by summing the position-dependent frequency for each letter. Words with repeated letters were not used. From the 720 anagrams for each six-letter word, two were chosen with high-positional frequencies and two were chosen with low-positional frequencies. Within each frequency pair, one member was orthographically regular and the other orthographically irregular. The rules used for judging regularity are listed in Appendix 4.1A. Listed in Appendix 4.1B are the 200 stimulus items--the 40 words and the 4 anagrams of each word. One-hundred twenty practice stimuli were generated from 24 additional words. These items were not as evenly matched for regularity and positional frequency as the experimental items and a few contained repeated letters.

Two occurrences of each of the 120 practice stimuli were randomized for each subject's practice sequence and two occurrences of each of the 200 experimental stimuli were randomized for each subject's experimental sequence. One occurrence was tested as a target trial and one as a catch

trial. On target trials, the letter was selected randomly with replacement from the six letters in the test string. For catch trials, a target was selected randomly from the set of 26 letters weighted by their probability of occurrence in the stimulus set. If the selected letter was present in the display string, additional drawings with replacement were made until an appropriate target letter was selected. The letters j, k, g, x, and z did not occur in the experimental stimulus strings and therefore were never tested.

Apparatus. The displays were presented on a Beehive model video computer terminal under the control of a Harris DC6024/5 computer. Hardware modifications permitted the terminal's video to be turned on and off with program-generated control signals. The display strings were loaded into the memory buffer of the terminal with the video off and then the video was turned on for the appropriate exposure duration. The Beehive employs a P4, blue-white phosphor which decays to .1% of maximum luminance within 32 msec. The experiment was conducted in a partially darkened room to enhance image contrast.

All single-letter target and letter-string test displays were presented in lowercase letters. The Beehive uses a 5 x 7 dot matrix, 2.5 mm wide x 5 mm high. At the average subject's viewing distance of approximately 38 cm, the six-letter strings subtended a horizontal visual angle of 2.25 degrees and a vertical visual angle of .75 degree. The displays were presented in the upper center of the screen. The single-letter target was positioned to appear three character positions to the left of the string display, a horizontal visual angle separation of .75 degree. To achieve a rapid and controlled presentation rate, the entire display sequence for an item was displayed beneath a cardboard mask affixed to the face of the cathode ray tube (CRT). A rectangular window in the mask served as a fixation box and permitted only the desired portion of the display to be observed.

Procedure. The subjects in the accuracy task were informed that the exposure duration would be brief and would be adjusted during the session to maintain an accuracy of 75%. A trial began with the presentation of the six-letter stimulus string for a brief duration. After the appropriate exposure duration, the display was scrolled upward on the CRT replacing the string with a masking stimulus consisting of six uppercase Xs. The masking stimulus was presented for 208 msec then scrolled again to present the target letter. The target letter remained in view until the subject responded, at which time the

display was terminated and a 500 msec intertrial interval begun.

The exposure duration of the test string was individually determined for each subject during the practice trials. Because of the 60 Hz refresh rate of the CRT one refresh cycle (16.7 msec) was chosen as the basic time unit. The number of cycles needed to achieve an overall accuracy level of 75% was determined on-line using a version of the adaptive PEST algorithm (Taylor & Creelman, 1967). Exposure duration was not changed during the experimental trials. The initial exposure duration was 20 cycles (334 msec) for all subjects. The range of exposures needed by the 14 subjects for 75% accuracy was 1 to 12 cycles with a median exposure duration of 3 cycles (50 msec).

The subjects in the RT task were instructed to respond as rapidly and as accurately as possible. All subjects were also told that one-half of the trials would be target trials and one-half would be catch trials. The display sequence was altered to accommodate the precuing. A trial began with the presentation of the target letter at the left of the fixation window for 500 msec. The target letter was then terminated and 1 sec later the stimulus string was displayed. The stimulus remained in view until the subject responded, at which time the display was terminated and a 500 msec interval begun. A masking field was not used.

To acquaint subjects in both the accuracy and RT tasks with the response procedure and visual display, each day's session began with 10 trials consisting of only the word YES or NO instead of the display string. No target letter was presented. The subjects pressed the correspondingly labeled response buttons. The YES response was assigned to each subject's preferred hand (right hand for 9 of the 14 subjects in the accuracy task; 13 of the 14 subjects in the RT task).

The "yes-no" trials were followed by the 240 practice trials and the 400 experimental trials. The practice trials were divided into four 60 trial blocks and the experimental trials were divided into five 80-trial blocks. At the end of each block the subject was given an opportunity to rest and received feedback in the form of a percent correct score or RT and error rate, as appropriate, for the preceding block. All subjects were tested in two such sessions scheduled on consecutive days.

Results

Orthographic structure. Analyses of variance were carried out on the average percentage correct for each subject for each of the 5 display types and target versus catch trials. For the accuracy task, the main effect of display type was significant, $F(4, 52) = 82.56, p < .001$, and the differences among these conditions were tested with specific contrasts. Accuracy for words was 12% greater than for the best pseudoword items (regular-high), $F(1, 52) = 40.54, p < .001$. Three orthogonal contrasts analogous to a 2 x 2 analysis of variance for the four nonword display types revealed that the regular strings were judged 8% more accurately than irregular strings, $F(1, 52) = 7.05, p < .01$. High positional-frequency strings were judged 5% more accurately than low positional-frequency strings, $F(1, 52) = 7.05, p < .01$. The interaction of positional frequency and regularity was not significant, $F(1, 52) = 3.20, p > .05$.

For the RT analysis, RTs for incorrect trials and correct trial RTs which exceeded 2 sec (less than 2% of the data) were excluded from the analyses leaving approximately 76 of 80 possible observations per cell. For the RT data, there was also a significant main effect of display type, $F(4, 52) = 17.16, p < .001$. This main effect was examined with the same set of specific contrasts. The 7 msec advantage for words over regular-high pseudowords was not significant, $F(1, 52) = 2.01, p > .10$. Regular items were judged 23 msec faster than irregular items, a significant difference, $F(1, 52) = 9.54, p < .005$. Neither the 4 msec difference between high and low positional frequency nor the interaction of regularity and positional frequency were significant, both F s < 1 . Error percentages were analyzed similarly to the RTs. The main effect of display type was not significant, $F(4, 52) = 1.13$.

Similarity analyses. To test for the effects of similarity between the target letter and the letters in the display, the data were first pooled across display types. Each trial's target letter was classified according to which letters in the display were likely to be seen as very similar to the target. The classification was derived from the lowercase single-letter confusability study published by Bouma (1971). A letter was counted as similar to the target letter if the letter was erroneously identified as the target more than 5% of the time. Table 4.2 presents the classification according to this analysis. Each trial was then classified by the number of nontarget letters in the display that were similar to the

target according to this criterion. Roughly 46, 33, 13, and 7% of each subject's trials fell into the categories zero, one, two, and three similar letters (other than the target letter itself), respectively.

For the accuracy task data, the effect of number similar letters was significant. $F(3, 39) = 7.14, p < .001$. Accuracy on target trials increased some with increases in the number of similar letters, whereas accuracy on catch trials decreased significantly with increases in the number of similar letters, $F(3, 39) = 22.28, p < .001$.

For the RT task, increasing the number of similar letters produced a significant increase in RT, $F(3, 39) = 4.41, p < .01$. The increase was clearly larger on catch trials than target trials, $F(3, 39) = 3.13, p < .05$.

Target-catch differences. For the accuracy task, accuracy was greater for catch trials (84%) than for target trials (72%), $F(1, 13) = 19.65, p < .001$, and this effect interacted with display types, $F(4, 52) = 8.14, p < .001$. Orthographic structure had a greater effect on target than catch trials. For the RT task, there was a difference between target and catch trials. Catch trials were 47 msec slower than target trials, $F(1, 13) = 42.55, p < .001$. However, this effect did not interact with display type, $F(4, 52) = 1.49$. Target-catch differences were also reflected in the error rates in the RT task. Target trials experienced more errors (6.8%) than did catch trials (3.3%), $F(1, 13) = 29.88, p < .0001$. There was no interaction with display type, $F(4, 52) = 1.24$.

Appendix 4.1A

Rules for the Selection of Orthographically Regular and Irregular Strings

These rules were derived by extending the rules given in Venezky (1970), Venezky and Massaro (1979), and Massaro et al. (1979). Among all possible anagrams of a word, a letter string was regarded as orthographically irregular if it contained at least one of the following spellings:

- a. phonologically illegal (in English) initial or final cluster (e.g., tlfies, iastlr).
- b. orthographically illegal spelling for an initial or final consonant or consonant cluster (e.g., hreagc, tleray).
- c. an illegal vowel spelling (e.g., prsaed, nmtaou).
- d. a phonologically illegal medial cluster (e.g., ilmpes, irltsa).

Letter strings were regarded as regular if they were phonologically legal and contained common vowel and consonant spellings.

Appendix 4.1B

The 200 Stimulus Items Used in Studies 1-10

The items listed are in columns according to type, where W = word; R-H = regular-low; I-H = irregular-high; R-L = regular-low; and I-L = irregular-low. Each row contains a word and the four anagrams generated from it.

<u>Word</u>	<u>R-H</u>	<u>R-L</u>	<u>I-H</u>	<u>I-L</u>
action	cainot	onicat	aioCNT	icnTOA
amount	mauton	otanum	auotnm	nmtaou
answer	swaner	erswan	wnraes	rnwesa
barely	blayer	byeral	lbraey	rbleya
belong	goblen	glenob	lnboeg	ebgnlo
breath	thaber	therab	bhrtea	rtbeha
bridge	begrid	gredib	ierbgd	ebrgdi
charge	chager	greach	ahcger	hreagc
double	bodule	edolub	leoubd	obdeul
famous	foamus	osafum	uaomfs	ofmsua
finger	fering	grefin	rfnieg	efngri
garden	nagred	gredan	arnged	ednrga
gather	thager	greath	hrtaeg	grheta
glance	caleng	glecan	lcnaeg	eclnga
hardly	hyrald	hydral	yrahld	rdlhya
itself	siflet	flesit	tlfies	eflsti
lawyer	weraly	lyeraw	rwlaey	eyrlwa
longer	logren	gleron	glnoer	egnrlo
master	tasmer	steram	mrtaes	emrtsa
modern	dormen	drenom	oermnd	rmedno

APPENDIX 4.1B (cont.)

<u>Word</u>	<u>R-H</u>	<u>R-L</u>	<u>I-H</u>	<u>I-L</u>
mother	thomer	therom	mroteh	hretmo
namely	manely	lyeman	lmnaey	emnyla
nearly	lareny	nyelar	lrnaey	enyrla
permit	primet	trepim	tmriep	rtpeim
player	rapley	lyerap	plraey	epylra
poetry	petory	tyerop	yptoer	erytpo
purely	yulper	ryelup	lpyuer	erplyu
reason	sarone	erosan	snaoer	rnseoa
remain	ramine	amerin	mraien	rniema
result	surtel	tresul	lrtues	elsrtu
shared	shread	dreash	srdaeh	hsreda
simple	pimsel	emslip	ilmpes	lsepmi
single	snigel	glesin	ilgnes	nglesi
spread	prased	dresap	prsaed	dsepra
stared	sarted	dresat	srtaed	dtsera
stream	trames	tresam	mtraes	esrtma
travel	vartle	treval	avrlet	tlerav
trials	sartil	trisal	iastlr	irltsa
turned	drunet	tredun	rdnuet	edtrnu
winter	triwen	trewin	wrntei	rntewi

Appendix 4.2

Details of Method, Procedure, and Results of Study 2:
Precue versus PostcueMethod

Subjects. Eighteen Introductory Psychology students, 8 male and 10 female, participated for an hour each day on 2 consecutive days and received research participation credit toward their psychology grade. The data from an additional 12 students was lost due to machine malfunctions and inadmissible overall accuracy. All subjects passed a preliminary vision test establishing at least 20/40 acuity for both eyes. None of the subjects had served in any of our previous experiments.

Stimuli and apparatus. The stimuli, apparatus, and display sequence for the postcue task were identical to those of the accuracy task of Study 1 described in Appendix 4.1. For the precue task, a trial began with the presentation of the target letter at the left of the fixation window for 500 msec. The target letter was then terminated and 1 sec later the string was displayed for the appropriate exposure duration. After displaying the stimulus string, the masking stimulus was scrolled into view and remained on the screen until the subject responded. The range of exposure durations used for the precue task was 1 to 8 cycles with a median of 2 cycles. For the postcue task the range was 2 to 11 cycles with a median of 3 cycles.

Procedure. Half the subjects received the precue task the first day and the postcue task the second day. For the remaining subjects this order was reversed. Each day the appropriate task was explained to the subject and performance accuracy was stressed. Subjects were told that one-half of the trials would be target and one-half would be catch trials. As in Study 1, the test sequence consisted of 10 "yes-no" trials, 4 60-trial practice blocks during which exposure was adjusted with the adaptive algorithm, and 5 80-trial experimental blocks during which the exposure remained constant. All subjects responded "yes" to target trials with their preferred hand (right hand for 17 of the 18 subjects).

Results

Orthographic structure. An analysis of variance was computed from the subjects' mean accuracies for the 2 x 2 x 5 combinations of type of task (precue versus postcue), trial type (target versus catch), and display condition (word, regular-high, regular-low, irregular-high, and irregular-low). The main effect of display type was significant, $F(4, 68) = 63.57$, $p < .001$, and planned comparisons were performed as in Study 1. Words were judged 8% more accurately than the regular-high positional-frequency nonwords, $F(1, 68) = 34.88$, $p < .001$. For the anagram display types, regular strings were judged 6% more accurately than irregular strings, $F(1, 68) = 42.01$, $p < .001$. High positional-frequency strings were judged 5% more accurately than low positional-frequency strings, $F(1, 68) = 30.63$, $p < .001$. Regularity and positional frequency did not interact, $F(1, 68) = 1.71$, $p > .15$. The interaction of type of task with display condition was significant, $F(4, 68) = 2.99$, $p < .024$. The overall effects of structure were greater in the postcue than the precue task. Although accuracy for irregular-low positional frequency was equivalent in the two tasks, the trends diverged steadily with increasing structure so that accuracy for words was 6% greater in the postcue than the precue task. No other interactions approached significance.

Similarity analyses. A post hoc similarity analysis was performed as in the previous experiment. The analysis of variance included as factors the type of task, the type of trial (target versus catch), and the number of similar letters. There was a main effect of the number of similar letters, $F(3, 51) = 2.86$, $p < .05$, and a significant interaction with catch and target trials, $F(3, 51) = 3.58$, $p < .025$. Relative to the previous accuracy experiment shown in Figure 4.10, the present results shown in Figure 4.13 revealed much smaller decreases on catch trials and little change on target trials with increases in the number of similar letters. The same results were observed in the precue and postcue tasks.

Target versus catch trials. Neither the main effect of type of task (precue versus postcue) nor trial type (target versus catch) was significant (both F s < 1); however, their interaction was significant, $F(1, 17) = 8.63$, $p < .009$. It is apparent from Figure 4.14 that subjects were more accurate on target than catch trials in the precue condition, whereas this bias was reversed in the postcue task.

Appendix 4.3

Details of the Method, Procedure, and Results
of Study 3: Mixed-CaseMethod

Subjects. Eighteen Introductory Psychology students, 3 male and 15 females, participated for an hour each day on 2 consecutive days and received research participation credit. One additional subject was lost for failure to complete the experiment. All subjects passed a preliminary vision test with at least 20/40 acuity for both eyes and no subject had served in previous experiments.

Stimuli, apparatus, and procedure. The procedure and apparatus were the same as in the RT task of Study 1 with the following modifications. The target letter presented at the beginning of each trial was displayed in both its uppercase and lowercase forms to remind the subject that either form might be present in the stimulus display. These two letters were shown for 500 msec followed by a 1 sec blank interval and were centered at what would be serial positions 3 and 4 in the stimulus display. The letter string was then displayed until the subject responded. For a randomly selected one-half of all trials, the first three letters of the stimulus string were presented in uppercase and the last three letters in lowercase. For the remaining trials this order of cases was reversed.

As in the previous experiments, the subjects served in two 1-hour sessions on consecutive days. Each session consisted of 10 "yes-no" warm-up trials, 240 practice trials, and 400 experimental trials. The instructions to the subjects emphasized speed and accuracy of responding and encouraged subjects to try to "read" each display as a whole as the most effective means of making fast and accurate judgments (Johnston & McClelland, 1974).

Results

Orthographic structure. The subjects' mean RTs for the 10 combinations of type of trial (target and catch) and display type (word, regular-high, regular-low, irregular-high, and irregular-low) were analyzed. Incorrect trial RTs and correct trial RTs exceeding 2 sec were excluded. The error rates varied between 1.0% and 8.9%. To determine if this variability in

error rate was affecting RT trends, the subjects were ranked on error rate and divided into two groups; the nine subjects with the lowest error rates formed one group and the remaining subjects the other. The average RT for the subjects with lower error rates was 647 msec, whereas for the higher error rate subjects, the mean RT was 756 msec, $F(1, 16) = 3.47$, $p < .078$. Grouping did not interact with any other factor in the analysis.

The significant main effect for display type, $F(4, 64) = 65.23$, $p < .001$, was analyzed in terms of the same comparisons examined in the previous experiments. The 7 msec difference between words and regular-high pseudowords was not significant, $F < 1$. The comparisons among the anagrams revealed that the 30 msec difference between regular and irregular anagrams was significant, $F(1, 64) = 33.50$, $p < .001$. Also the 14 msec difference between high and low positional-frequency items was significant, $F(1, 64) = 8.09$, $p < .006$. The interaction of regularity with positional frequency was not significant, $F < 1$.

An analysis of variance was calculated on error rates. The display type main effect, $F(4, 68) = 2.86$, $p < .030$, and the interaction of display type with trial (target versus catch), $F(4, 68) = 2.98$, $p < .025$, were both significant. Although the trend does not seem to be particularly meaningful, the source of these two effects is the somewhat greater error rate for the irregular items on target trials.

Similarity analyses. A restricted similarity analysis was done on only the lowercase letters in the display strings. Reaction times increased 24 msec on catch trials and did not change on target trials with increases from zero to one similar letters, $F(1, 17) = 5.84$, $p < .05$.

Target versus catch trials. As in the previous RT task, target trials were judged more quickly (660 msec) than were catch trials (743 msec), $F(1, 16) = 65.23$, $p < .001$. Type of trial did not interact with type of letter string.

For the error data, catch trials (97.1%) were significantly more accurate than target trials (93.8%) as in the RT data of Study 1, $F(1, 17) = 33.81$, $p < .001$. However, trial type did interact with display type as noted above.

Appendix 4.4

Details of Method, Procedure, and Results of Study 4:
Limited Viewing TimeMethod

Subjects. Sixteen Introductory Psychology students, 8 males and 8 females, were tested for a single 90-minute session and received research participation credit. As in the previous experiments, all subjects were tested for visual acuity of at least 20/40, all responded with their preferred hand (right hand for 15 of the 16), and no subject had participated in the earlier experiments.

Procedure. The stimuli and apparatus were the same as the previous experiments. An experimental trial began with the presentation of the lowercase target letter positioned within the viewing rectangle at display serial position 4 for 500 msec. A 1 sec blank interval followed and then the stimulus string in lowercase was presented for 200 msec. The stimulus was scrolled upward and replaced by six uppercase Xs which served as a mask and remained in view until the subject responded. Following the subject's response there was a 500 msec intertrial interval.

The testing session for each subject consisted of 10 "yes-no" warm-up trials, 240 practice trials, and 2 400-trial experimental sessions. As in all previous experiments, trials were grouped into 80 trial blocks. Between blocks subjects were given a minute or two of rest and feedback in the form of a mean RT error percentage for the block. A 5 minute rest was given between the two experimental sessions.

The subjects were told that half the trials would be target trials and half would be catch trials. They were also informed that the stimulus display would be brief and that they would have to read it carefully. Speed and accuracy of responding were emphasized.

Results

Orthographic structure. As in the previous RT studies, the subjects' mean RT for each cell of the design, excluding incorrect trial RTs and correct RTs exceeding 2 sec, were analyzed. The subjects' error rates varied between 4.5% and 11.9%, which is higher than normally desired for RT tasks. For

this reason, subjects were again ranked by error rate and divided into low and high error rate groups. An analysis of variance indicated no effect of grouping the subjects by error rate (all F s < 1). The analyses of the RT data indicated no effect of display type, $F(4, 56) = 1.27$. Only the display type by trial type interaction was significant. The nature of this interaction may be seen in Figure 4.16. Target and catch RTs appear to trade-off across display types. This pattern is indicative of a shift in bias, or at least confidence, across the display types. As subjects become more reluctant (slower) to respond "yes" they become more willing (faster) to respond "no." For the most part, this pattern of decreasing confidence coincides with the decline in orthographic structure which is a reasonable reaction to displays that are more difficult to resolve.

An analysis of the error data revealed only a main effect of display type, $F(4, 60) = 6.79$, $p < .001$. This display type main effect was examined with the usual set of contrasts. The difference in error rate between words and regular-high items was too small to be significant, but the 1.8% difference between words and the average of the regular items was significant, $F(1, 60) = 6.42$, $p < .013$. The error rate was 1.7% smaller for regular than irregular items, $F(1, 60) = 8.59$, $p < .005$. Neither the positional-frequency effect (error rate difference of only 0.4%, $F < 1$) nor the interaction of regularity with the positional frequency were significant, $F(1, 60) = 1.90$.

Similarity analyses. For response accuracy, the number of similar nontarget letters had no effect on either target or catch trials, p 's > .1. For RTs, the effect of the number of similar nontarget letters had no effect on target trials and a large effect on catch trials, $F(3, 45) = 14.24$, $p < .001$. RTs increased only 1 msec on target trials and 67 msec on catch trials with increases from zero to three similar letters.

Target-catch differences. As in the previous postcue RT experiments there was a main effect of trial type, $F(1, 14) = 26.62$, $p < .001$, with target trials (563 msec) being responded to more quickly than catch trials (635 msec). There was also the interaction of trial type with display type discussed above. For the error data, there was neither a significant main effect of trial type, $F(1, 15) = 1.95$, nor a significant interaction with display type, $F < 1$.

Appendix 4.5

Details of Method, Procedure, and Results of
Study 5: Speeded RT Task with Good Visual InformationMethod

Subjects. Eleven Introductory Psychology students, 10 female and 1 male, were tested for a single 1-hour session and received research participation credit. Subjects were tested in four groups; three groups with three participants and one with two. All subjects responded with their preferred hand (right hand for 9 of the 11). No subject had participated in an earlier experiment.

Apparatus and stimuli. The stimuli were the same as in the previous experiments. The visual displays were generated by a DEC LSI-11 computer under software control and presented on Tektronix Monitor 604 oscilloscopes (Taylor, Klitzke, & Massaro, 1978). These monitors employ a P31 phosphor with a decay to .1% of stimulated luminance within 32 msec. The alphabet consisted of lowercase nonserifed letters resembling the type font Univers 55. For an observer seated comfortably at an experimental station, the six-letter displays subtended about 1.9 degree of visual angle horizontally and the distance from the top of an ascender to the bottom of a descender was about .4 degree. This computerized laboratory facility permitted the testing of up to four subjects in parallel. Each subject received the same visual display on his or her display monitor, and the computer collected individual subject's responses.

Procedure. The experimental procedure was a response-signal method (Reed, 1973) in which the subject endeavors to respond in synchrony with a signal which, in this case, was the offset of the stimulus display string. The exact procedure was as follows. A trial began with the presentation of a single fixation point in the center of the screen. After 250 msec the fixation point was replaced by the test letter for 500 msec. The fixation point returned for 250 msec followed by the display string. The display was initially presented for 550 msec with this value revised every 20th trial to achieve an overall accuracy of 75% for the group of subjects being tested. The trial terminated when all subjects responded or, in the event that a subject failed to respond, at the end of 4 sec. A 500 msec intertrial interval was given. At the end of every 20

trials, each subject was given feedback about his or her performance in the following manner. At trial termination the word FEEDBACK was presented at the center of the screen for 1 sec. One-half second later two columns of numbers appeared on the screen, a subject number and the subject's average deviation from the display string offset to the response in msec. A negative deviation indicated an average response prior to the display offset; a positive deviation indicated that the average response fell after the display offset. Each subject's number was clearly posted at his or her subject station to facilitate remembering which feedback score was appropriate. The feedback scores were visible for 2.5 sec after which came the normal half second intertrial interval. The subjects' instructions acquainted them with the task and the general types of stimulus items. It was stressed that the subjects were to be as accurate as possible and yet respond as nearly in synchrony with the signal as possible. They were told how they would receive feedback and again it was stressed that they should endeavor to maintain a zero feedback score. All subjects received 120 practice trials followed by two experimental sessions of 400 trials each. A short rest break was given between the two experimental sessions.

All subjects became adept at the task by the end of the practice trials or early in the first experimental session so that the display offset times were reasonably stable throughout the two sessions. Average display offset times for the four groups of subjects were as follows: 412, 634, 485, and 505 msec.

Results

Separate analyses were made of the mean accuracies and the mean RTs for both correct and incorrect trials. For the accuracy data there were no significant effects (all F s < 1).

The analysis of variance for the RTs included trial type, display condition, and correctness of the response as factors. There were significant effects of type of trial, $F(1, 10) = 8.04$, $p < .017$; correctness of response, $F(1, 10) = 19.99$, $p < .001$; and the interaction of these two factors, $F(1, 10) = 34.32$, $p < .001$. These effects were a result of the significantly slower correct catch trial responses seen in Figure 4.17. Although we observed the normal RT advantage for correct target trials over correct catch trials, the incorrect target and incorrect catch trials nearly coincided with the

correct target trial responses. The main effect of display type, $F(4, 40) = 3.17$, $p < .023$, and the interaction of display type with correctness of response and type of trial, $F(4, 40) = 3.25$, $p < .021$, were significant. Mean RT increased from 477 msec for words to 492 msec for irregular-low nonwords.

Appendix 4.6

Details of Method, Procedure, and Results of Study 6:
Speeded RT Task with Poor Visual InformationMethod

Subjects. Thirteen Introductory Psychology student volunteers participated as subjects; the data of two were lost due to a computer malfunction. All subjects were native English speakers, had normal or corrected-to-normal vision, had not participated in any of the other experiments, were right-handed, and received some credit toward their course grade for participating.

Materials, apparatus, and procedure. The stimuli and apparatus were the same as in Study 5 and are described in Appendix 4.5. The important difference in the present experiment is that the complete test letter string was rotated 180 degrees in the picture plane. Rotating the complete test string rather than the individual letters separately creates a display much easier to read (Kolers, 1970). The initial display duration was lengthened to 700 msec. The target letter was still presented upright. Subjects responded "yes" with their dominant hand.

Results

The means among letter-string types indicated a 1.1% advantage of words over the best pseudowords (regular-high) and a 1.6% advantage of words over all types of nonwords. Accuracy for regular nonwords was 0.2% better than for irregular nonwords. Nonwords of high positional frequency had a 0.5% advantage over nonwords of low positional frequency. As might be expected from differences this small, the analysis of variance indicated no effect of letter-string types, $F < 1$. Furthermore, there was no difference between target and catch trials, $F < 1$, and no interaction, $F < 1$.

The accuracy results of the present experiment can be compared to those of Study 5 with upright letter strings. This analysis indicated no differences of any kind between the two experiments (for upright-rotated, $F(1, 21) = 0.14$; for target-catch, $F(1, 21) = 1.18$; for the upright-rotated \times target-catch interaction, $F(1, 21) = 0.07$; for the letter-string types, $F(4, 84) = 0.92$; for the upright-rotated \times letter-string interaction, $F(4, 84) = 0.09$; for the

target-catch x letter-string interaction, $F(4, 84) = 1.27$; and for the three-way interaction, $F(4, 84) = 0.23$.

An examination of the reaction times in the rotated experiment revealed results similar to those of the accuracy data. Means were calculated on correct reaction times that did not exceed 2,000 msec (0.25% exceeded). Words were actually responded to 9 msec slower than regular-high nonwords. Reaction times were 9 msec longer for irregular nonwords than for regular nonwords and 8 msec longer for nonwords of low rather than of high positional frequency.

The analysis of variance indicated significantly longer reaction times for catch trials as compared to target trials, $F(1, 11) = 46.73$, $p < .001$, but no effect of letter-string types, $F(4, 44) = 2.43$, $p < .05$, and no interaction, $F(4, 44) = 1.60$, $p < .1$.

The reaction times for the rotated letter strings also were compared to those for the upright letter strings. The average reaction time for rotated letter strings was 104 msec longer than for upright letter strings, $F(1, 21) = 26.94$, $p < .001$. Catch trials took 66 msec longer than target trials, $F(1, 21) = 75.84$, $p < .001$, but this did not interact with whether the letter strings were rotated, $F(1, 21) = 1.26$. There was an overall difference of letter-string types, $F(4, 84) = 5.77$, $p < .001$. Letter-string types did not interact with the other factors: with rotation, $F(4, 84) = 0.49$; with target-catch trials, $F(4, 84) = 1.18$; and for the three-way interaction, $F(4, 84) = 1.18$.

To examine differences among display durations, the display durations for each of the 40 sets of 20 trials were grouped into 4 blocks of 10 sets each. Each of the consecutive 4 blocks thus represented 10 display durations for 10 sets of 20 trials. An analysis of variance indicated no differences among blocks, $F(3, 33) = 0.67$, but did indicate a difference among the 10 sets within a block, $F(9, 99) = 4.25$, $p < .001$, and a block by set interaction, $F(27, 297) = 2.21$, $p < .005$. These effects reflect the operation of both practice and fatigue. Initially the display durations decreased with experience, but then increased toward the end of the first 200 trials. Following a break between halves of the experiment, display times again decreased until late in the second half of the experiment.

The display durations also were compared between the upright and rotated experiments. Although the average display duration was 498 msec for upright presentation and 556 msec for rotated presentation, the effect was not significant, $F(1, 21) = 2.64$, nor was there an effect of the blocks of 10 display durations, $F(3, 63) = 1.56$, and the two factors also did not interact, $F(3, 63) = 2.53$. There were, however, differences among individual display durations, $F(9, 189) = 2.05$, $p < .05$, and display durations interacted with everything else: with upright-rotated letter strings, $F(9, 189) = 4.65$, $p < .001$; with blocks, $F(27, 567) = 1.63$, $p < .05$; and in a three-way interaction, $F(27, 567) = 1.69$, $p < .05$. These interactions reflect the slight fatigue factor in the rotated experiment and the absence of such in the upright experiment. For the upright stimuli, the display durations continued to decrease throughout the experiment.

Appendix 4.7

Details of Method, Procedure, and Results of Study 7:
High-Accuracy RT Task with Poor Visual InformationMethod

Twelve subjects were used and met the same constraints as in the speeded RT task in Study 6 (Appendix 4.6). The same rotated stimuli also were used with the following procedural differences. A trial consisted of a 250 msec fixation point followed by an upright letter which was presented for 500 msec at the fixation point and then the rotated letter string. The letter string remained on the screen until the last subject (up to four at a time) responded.

Results

Means were calculated for correct reaction times that did not exceed 2,000 msec (1.4% exceeded). Reaction times generally increased from words to the irregular-low nonwords, $F(4, 44) = 6.47$, $p < .001$. Regular-high nonwords averaged 6 msec longer than words, irregular nonwords resulted in reaction times that were 16 msec longer than regular nonwords, and nonwords of low positional frequency resulted in reaction times that were 10 msec longer than nonwords of high positional frequency. The average reaction times for catch trials was 110 msec longer than for target trials, $F(1, 11) = 100.10$, $p < .001$, but this did not interact with letter-string types, $F(4, 44) = 0.79$.

Differences between letter strings were further examined by orthogonal contrasts. The difference between words and all nonwords was significant, $t(44) = 3.18$, $p < .005$, as was the difference between regular and irregular nonwords, $t(44) = 1.683$, $p = .05$. However, the difference between nonwords of high and low positional frequency was not significant, $t(44) = 0.94$. Similarly, a Newman-Kuels test indicated that the 6 msec difference between words and the best pseudowords (regular-high) was not significant, $p > .10$.

A comparison of reaction times between the initial replication experiment using upright stimuli and the present study indicated that reaction times for rotated stimuli averaged 279 msec longer than for upright stimuli, $F(1, 24) = 42.35$, $p < .001$. The difference between catch and target

trials was significant, $F(1, 24) = 153.44$, $p < .001$, as was the trial type \times rotated-upright interaction, $F(1, 24) = 23.68$, $p < .001$. Differences between letter strings were significant, $F(4, 96) = 20.12$, $p < .001$, but letter strings did not interact with anything: with rotated-upright, $F(4, 96) = 0.53$; with target-catch, $F(4, 96) = 0.53$; and for the three-way interaction, $F(4, 96) = 1.60$.

Appendix 4.8

**Details of Method, Procedure, and Results of Study 8:
Typicality Ratings**

The overt judgment experiments consisted of asking subjects to make conscious judgments about the test items. Three sets of rating judgments were obtained under similar instructions and are described here as "Typicality Ratings." A fourth rating was obtained in Study 9 under explicit instructions as to the nature of positional frequency and is described as "Positional-Frequency Ratings." A final set of judgments in Study 10 were obtained by a "Paired-Judgment" procedure.

Method

Rating Forms. Three rating forms were prepared. The first form employed the 160 anagrams; the second form employed the 160 anagrams plus 40 pseudowords derived from the original words by the substitution of a single letter; and the third form used the entire 200 stimulus strings including the original words. (See Appendix 4.1B for the 40 words and their 160 anagrams; see Appendix 4.8A for the one-letter substitution pseudowords). Each list was randomized and printed with 20 letter strings per page. The strings were typed in lowercase with five underscores typed to the right of each string as a space for writing the string's rating. The rating scale to be used was reproduced at the top of each page as follows:

"Most like English 10 9 8 7 6 5 4 3 2 1 Least like English."

The four or five pages comprising each list were assembled in random order into booklets. Each booklet was prepared with an appropriate sheet of instructions. These instructions are reproduced in Appendix 4.8B, and 4.8C.

Subjects. One hundred ninety-eight Introductory Psychology students filled out one of three rating forms. The raters participated in groups of 8 to 12 persons and received research participation credit. The first 70 subjects rated the first form which contained only the 160 anagrams. The next 65 subjects filled out the second form which included the 160 anagrams and the one-letter substitution pseudowords made from the words. The final 63 subjects rated the third form which

contained the original set of 200 stimulus items (i.e., the 160 anagrams and the 40 words). Each subject worked at his or her own speed, and on the average, took 15 to 25 minutes to rate the items.

Procedure. Each session began with the experimenter reading the instructions while the subject followed along. The formal instructions were supplemented with verbal instructions stressing "similarity to English spelling" as the primary rating criterion. Subjects were urged not to adopt strategies of rating regularity on the basis of being able to transform a string into a word through letter additions, deletions, substitutions, or by making anagrams. The raters who received the list containing actual words were also asked to disregard the semantic content of the words to the fullest extent possible. Finally, the relative nature of the scale was stressed. It was indicated that the psychological scale should span the presented set of items and all 10 scale values should be used.

Results

Average ratings were tabulated both within subjects and across subjects. For the within subject tabulations, the subject's ratings for the 40 stimuli in each category were averaged. For the item analyses, an average for each item was obtained by averaging the ratings assigned by each subject. These data permitted analyses to determine if the ratings' reliability reflected differences in mean ratings for the 4 or 5 categories of stimuli. Six one-way analyses of variance were computed for all pairwise comparisons among the three instruction conditions; three analyses were based on subjects as the sampling variable and three were based on items as the sampling variable. The mean ratings were so reliable that error variances were uniformly small in all six analyses (standard error of estimate is the range of .04 to .08 scale units for all six analyses). With such small error variances, all pairwise differences between category means were reliable at the .05 level.

Appendix 4.8A

The Pseudowords used in Study 8.

<u>Word</u>	<u>Pseudoword</u>	<u>Word</u>	<u>Pseudoword</u>
action	aption	mother	mather
amount	amaunt	namely	nomely
answer	onswer	nearly	mearly
barely	parely	permit	permat
belong	belang	player	plader
breath	breeth	poetry	boetry
bridge	bradge	purely	burely
charge	churge	reason	leason
double	gouble	remain	remoin
famous	vamous	result	resilt
finger	vinger	shared	sharet
garden	gurden	simple	sumple
gather	gither	single	fingle
glance	glunce	spread	spreed
hardly	herdly	stared	stured
itself	itsolf	stream	stroom
lawyer	rawyer	travel	trovel
longer	ronger	trials	treals
master	moster	turned	tarned
modern	madern	winter	wunter

Appendix 4.8B

**Directions for Typicality Ratings Without Words and for
Typicality Ratings With Pseudowords**

On the pages which follow are letter strings made up to look like English words. None of the strings, however, is an English word. Your task is to rate each one on a scale of 1 to 10 according to how much it looks like a real English word. Strings which are extremely close should be given high ratings, while strings which are not very close should be given low ratings. Since no real words are included, the highest possible rating (10) should be given to strings which you feel are most like English words. Similarly, a rating of 1 should be given to the string which are least like English words. The strings have been selected to represent all possible levels of similarity to English with roughly equal numbers at each level.

Skim over the strings on the first page before you begin rating so you have an idea of their range of variation. There are 160(200) strings in all, divided randomly across 4(5) pages, so each page should have roughly the same variation as any other. You should work at a fairly quick and even pace although there is no time limit.

Appendix 4.8C

Directions for Typicality Ratings with Words

On the pages which follow are letter strings made up to be spelled like English words. Some of the strings are, in fact, English words, whereas other strings are not English words. Independent of the meaning of the words, however, your task is to rate each one on a scale of 1 to 10 according to how much it is spelled like typical English spelling. Strings which are extremely close should be given high ratings, while strings which are not very close should be given low ratings. The highest possible rating (10) should be given to strings which you feel are spelled most like English. Similarly, a rating of 1 should be given to the strings which are spelled least like English. The strings have been selected to represent all possible levels of similarity to English spelling with roughly equal numbers at each level.

Skim over the strings on the first pages before you begin rating so you have an idea of their range of variation. There are 200 strings in all, divided randomly across 5 pages, so each page should have roughly the same variation as any other. You should work at a fairly quick and even pace although there is no time limit.

Appendix 4.9

The Method, Procedure, and Results of Study 9: Positional-Frequency Ratings

Method

Rating forms were constructed by generating fresh copies of the form from Study 8 employing the 160 anagrams. Explicit positional-frequency instructions (reproduced in Appendix 4.9A) were substituted for the previous instructions and the scale at the top of each page was modified to read:

"Most frequent 10 9 8 7 6 5 4 3 2 1 Least frequent."

As before, subjects were tested in small groups of 8 to 10; the formal instructions were read followed by verbal emphasis of the salient points of the instructions; and subjects were given unlimited time although they usually completed the form in 15 to 25 minutes. Forty-nine Introductory Psychology students rated this form.

Results

As in Study 8, average ratings were tabulated by letter-string type for each subject and for each stimulus item across subjects. Two analyses of variance were computed on these data and, as in the previous ratings, all possible pairwise comparisons of the four anagram categories were significant.

Appendix 4.9A

Directions for Positional-Frequency Ratings

It is a relatively simple matter to determine the frequency with which the different letters of the alphabet are used at various locations in words. It is another matter to determine if people are aware of this "positional frequency." On the pages which follow are letter strings derived by anagramming common English six-letter words. In some instances the letters have been moved to the positions they most commonly occupy in six-letter words (high frequency positions) and in some instances they have been moved to locations that they very infrequently occupy in six-letter words (low frequency positions). Your task is to rate each string on a scale of 1 to 10 according to how much you believe that the letters in the string occupy their most frequent positions. The highest possible rating (10) should be given to the strings whose letters are in their most frequent positions. The lowest rating (1) should be given to strings whose letters are in their least frequent positions. The strings have been constructed so that some strings have all their letters in their most frequent positions, some strings have their letters in their least frequent positions, and the remaining strings have letters at all various positions of greater or lesser frequency. Thus you should find strings appropriate to all 10 levels of the scale with roughly equal numbers at each level.

Skim over the strings on the first page before you begin rating so you have an idea of their range of variation. There are 160 strings in all, divided randomly across four pages, so each page should have roughly the same variation as any other. You should work at a fairly quick and even pace even though there is no time limit.

Appendix 4.10

**Details of Method, Procedure, and Results of Study 10:
Regularity versus Positional-Frequency Paired-Judgments**

Method

Response forms. The paired-judgment forms were prepared by first partitioning each of the four categories of anagrams into four randomly selected subsets of 10 items each. Eighty pairs were produced by pairing the 160 anagrams as follows: One 10-item subset from each category was matched to a 10-item subset from another category and the members of the subset paired off producing between category pairs. This left one 10-item subset within each category whose members were paired with each other to form five within category pairs. A total list of 320 pairs was formed by performing this matching procedure four times, rotating the mating of the subsets within each category to achieve a completely counterbalanced list where every item was paired once with a member of every other category and once with a member of its own category. Eight such totally different lists were constructed and item balancing (right-hand pair member versus left-hand pair member) was randomized. The total effect was to ensure that every item had an equal opportunity to be paired with any other item.

Once constructed, the lists were printed with 80 pairs per page. Each booklet was prefaced with a set of instructions which either gave a detailed description of regularity or positional frequency. In both cases, appropriate examples were included. The "regularity" instructions group chose the most regular of each pair whereas the "positional-frequency" instructions group chose the member with the highest positional frequency.

Subjects. Forty-eight Introductory Psychology students completed the paired-judgment task. Of these raters, 22 received "regularity" instructions and 26 received "positional-frequency" instructions. All subjects received research participation credit as recompense.

Procedure. As in the previous rating studies, subjects were tested in groups of 10 to 15 and given unlimited time to complete the form. Due to the somewhat lengthier form, subjects took 20 to 30 minutes to complete the form. Approximately one-half of the groups received explicit regularity instructions (see Appendix 4.10A), which described

letter sequence regularity, and required that the more regular member of each pair be chosen. The remaining groups received explicit positional-frequency instructions (see Appendix 4.10B) which described single-letter positional frequency and required that the member of each pair with the higher overall positional frequency be chosen. All subjects within a group received the same instructions.

Results

The proportions of times that items from an anagram category were chosen over items from the same category or different categories are shown in Table 4.5. Table 4.5 gives the proportion of times an item from the column category was selected when paired with an item from the row category. Comparing the two panels of Table 4.5 we observe that subjects in both instructed conditions made choices more consistent with the application of regularity than with the application of positional frequency although both observed tables differ from the hypothesized "perfect" outcome. Comparing the two panels of Table 4.5 to each other, we see evidence again that subjects can modify their judgments slightly when given explicit instructions to attend to positional frequency.

Appendix 4.10A

Regularity Instructions for Paired-Judgments

Our research is directed at discovering what it is we know about English words that allows us to look at meaningless strings of letters and judge how much they resemble words. In trying to understand the structure present in words, we have focused on an important component property--regularity of letter sequencing. For example, there are many consonant sequences which begin words (e.g., sh, sl, ch, th, str, etc.) and which end words (e.g., ng, st, tch, ck, etc.). There are common vowel sequences as well (e.g., oo, ee, ea, ou, etc.). Finally, there is regularity in how these consonants and vowel groupings are themselves sequenced within words.

On the pages which follow are 320 pairs of meaningless six-letter strings which have been constructed to vary along this dimension of letter sequence regularity. That is, some strings preserve normal letter sequencing while others violate normal sequencing to a lesser or greater degree. The object of this experiment is to determine the extent of your knowledge by having you select the more regular string in each pair. You are to evaluate both members of each pair and choose the one that is more permissible in terms of regularity of letter sequencing.

Before you begin, skim over the pairs on the first page to get a feel for the task. Once you begin, try to work at a fairly steady pace and not dwell too long on the pairs you find difficult. You must make a choice for each pair, guessing if necessary. Have you any questions?

Appendix 4.10B

Positional-Frequency Instructions for Paired-Judgments

Our research is directed at discovering what it is we know about English words that allows us to look at meaningless strings of letters and judge how much they resemble words. In trying to understand the structure present in words, we have focused on an important component property--letter frequency. Individual letters and groups of letters occur with varying frequency at different positions within words. For example, you can probably think of words that end in d and words that end in w, but d is 10 times more frequent than w in final position. Likewise, th and ch both occur at the beginning of words, but th is 50 times as frequent in initial position. In general, consonants are more frequent than vowels at word initial and word final positions with the exception of final e, and so on.

On the pages which follow are 320 pairs of meaningless six-letter strings which have been constructed to vary along the dimension of letter frequency by position. That is, some strings have been created by placing each letter in its most frequent position in six-letter English words. Others have their letters placed in less frequent positions and still others have their letters in their least frequent positions.

The object of this experiment is to determine how well you can judge when letters are in their most frequent positions. You are to evaluate both members of each pair and choose the more "frequent" of the two, e.g., the string which appears to have more of its letters in their more frequent positions.

Before you begin, skim over the pairs on the first page to get a feel for the task. Once you begin, try to work at a fairly steady pace and not to dwell too long on the pairs you find difficult. You must make a choice for each pair, guessing if necessary. Have you any questions?

This page intentionally left blank

5 Structural Descriptions

To provide a more detailed analysis of orthographic structure effects, we conducted post hoc correlational analyses of the perceptual recognition and overt judgment data reported in Chapter 4. The independent variables were based upon various frequency and regularity measures. Even though firm conclusions might not be reached, we felt that our sample of items was sufficiently large and sufficiently representative that these correlational analyses could provide valid indications of which measures might warrant further theoretical and empirical examination.

Three dependent measures were derived from the experiments reported in Chapter 4 for each of the 200 stimulus items, i.e., the 40 words and their 4 anagrams. The first measure was the mean accuracy for each item averaged across all trial occurrences in the postcue accuracy experiment of Study 1: Initial Replication and the precue and postcue experiments of Study 2: Precue versus Postcue. A corresponding reaction time (RT) measure was derived by averaging the RTs for all trial occurrences of each item in the reaction time experiments of Study 1: Initial Replication and Study 3: Mixed-Case. The third measure was the average ratings from one of the rating experiments in Study 8: Typicality Ratings. The experiment was the one in which the words themselves were rated along with the 160 anagrams. Each of the 200 stimulus items along with its respective accuracy, reaction time, and rating is presented in Appendix 5.1.

As discussed in Chapter 3, a large number of measures of orthographic structure based upon counts of letter, *n*-gram, and word frequencies may be computed. Quantitative measures based upon orthographic rules are less readily constructed. Although

we devised and tested one measure of regularity based upon the rules outlined in Chapter 3, we concentrated upon frequency measures and questions related to their construction. The specific issues were as follows. Should frequency counts be based upon counts of word types or word tokens? What frequency scale gives the most predictive index of performance, raw frequency or some transformed frequency scale such as log frequency? Is positional dependency, i.e., whether the counts preserve position of the n -gram in the sampled word, an important factor? What n -gram unit gives the most powerful measure: single-letter, bigram, or trigram counts? What is the contribution of lexical status and word frequency to performance? And finally, what method of combining the frequencies of the components of a letter string provide the best overall index of its structure? These questions will be discussed in the following sections.

DESCRIPTIVE MEASURES

The source of all our frequency measures is a word corpus sampled by Kucera and Francis (1967). This corpus consisted of 500 samples of approximately 2,000 words each selected from 15 categories. A description of the corpus, its selection, and its processing are presented by Kucera and Francis. We obtained a magnetic tape of the word count produced from the corpus (i.e., the "Rank List" in the Kucera and Francis monograph). The words were sorted into 10 lists consisting of 1- to 10-letter words respectively. Words longer than 10 characters were deleted as were items containing numbers, punctuation, or special codings for capitalizations, foreign alphabets, and unusual graphic features or symbols. This resulted in 10 lists of words, one for each letter length. These word lists formed the basis for counts of single letters, bigrams, and trigrams.

Tables were prepared by counting the occurrence of each n -gram at the position it occurred in words of a given length. Two counts were obtained. One was a type count based upon the number of word types that contained a given n -gram, and the other was a token count based upon the total number of occurrences of the word containing the n -gram. A position-insensitive count (but still word length dependent) was also obtained for each n -gram by summing across the

position-dependent counts. Because Kucera and Francis maintained a faithful count of the actual graphic patterns found in the corpus, their list contains rare words, typographic errors, foreign person and place names, and other ideo-syncretic items. To limit the impact of such items on our tabulations, cut-off limits were established for both word frequency and number of samples. The cut-offs were a minimum count of at least one occurrence in three of the 500 samples. Thus, unusual words and usages, regardless of their frequency, were ignored unless they occurred in three or more separate samples. In this way, words with the minimum frequency of three were counted only if they occurred once in each of three samples. Although this limit was arbitrary, inspection of the word list in the low frequency range indicated that these were reasonable cut-offs. The single-letter token counts for word lengths 3 through 7 and the bigram token counts for word lengths 3 through 7 are presented as Appendices 5.2 and 5.3, respectively.

All of the frequency measures that will be discussed were derived from these tables with two exceptions, word frequency and Englishness. The frequencies of the 40 words were taken directly from the Kucera and Francis count. Englishness is a measure proposed by Travers and Olivier (1978) and was computed from tables supplied to us by the authors. Englishness is a continued product of n -gram probabilities estimated to the level of third-order transitional probabilities. The formula for six-letter words or letter strings is as follows:

$$\text{Englishness } (\#L1L2L3L4L5L6\#) = \\ - \log [P(L1|\#) * P(L2|L1) * P(L3|L1L2) * \\ P(L4|L2L3) * P(L5|L3L4) * P(L6|L4L5)] \quad (4)$$

where # represents a space and L_i is the letter at position i . The term $P(L_k|L_{(k-2)}L_{(k-1)})$ is estimated by the ratio of the trigram frequency of $L_{(k-2)}L_{(k-1)}L_k$ to the bigram frequency $L_{(k-2)}L_{(k-1)}$. It should be noted that these frequency estimates are not position sensitive except in as much as they are tied to initial position occurrences, i.e., the initial trigram #LL2 and initial bigram #L1. The counts are also not word-length specific. The initial term $P(L_1|\#)$ in the formula is actually the frequency of the initial letter as it occurs in the first position in words of any length. The rationale for examining this measure is to provide a comparison to the other frequency measures we will examine. Travers and Olivier (1978) provided some evidence that the measure had success in

describing performance in a full report of short presentations of letter strings.

Our regularity measure consisted of a count of the number of irregularities contained in each of the 200 stimulus items. The exact rules for the irregularity count are given in Table 5.1. An irregularity was counted for each violation of certain scribal constraints, for each impermissible vowel cluster and for each impossible consonant cluster when considered as part of a monosyllable. The items were treated as monosyllables because an appropriate syllabication is difficult to impose on nonsense strings. Without rigid criteria, judgments for some anagrams would be unreasonably arbitrary. Thus, we chose to treat all strings as monosyllabic even though this required classification of some two syllable words as containing irregularities at syllable boundaries. The counts of irregularities for each of the items of each of the five stimulus categories are shown in Appendix 5.1.

TYPE VERSUS TOKEN COUNTS

The first general issue in deriving a frequency measure is whether type counts or token counts are most appropriate. Certainly, the two methods of counting will produce frequency measures that have some degree of correlation. Even though it is probably possible to find occurrences of n -grams occurring in only a small number of extremely frequent words and other n -grams occurring in many infrequent word types, on the average the greater the type count the greater the token count. Only by examining situations where the two measures are relatively uncorrelated can the importance of the two approaches be measured. Thus, if our items exhibit a strong correlation between type- and token-based measures little can be determined from our data to support one over the other. At the same time, if our sample is representative of a wide range of letter strings and the two approaches are highly correlated, the question may be of no practical significance. In any situation where such a measure would be useful, either might be used with essentially identical effect.

We found that the correlations between comparable type measures and token measures for our sample were very high. Measures based on single letters, bigrams, and trigrams,

Table 5.1 The rules for the irregularity count of the 200 stimulus items

1. Segment string into vowel and consonant substrings. Treat final -le as if it were -el. Treat h between vowels as a (legal) consonant.
 2. For each consonant string, determine minimal number of vowels which must be inserted to make the string pronounceable. Initial consonant clusters must be legal in initial position. Final consonant clusters must be legal in final position, including those followed by final e. Medial consonant clusters must be legal in initial position.
 3. Rate each resulting consonant substring for position-sensitive scribal regularity (count one for each irregular substring).
 4. For each vowel substring, determine minimal number of consonants which must be inserted to create scribally regular sequences. Mark as irregular illegal initial and final vowel substrings.
 5. Count number of inserted vowels and consonants, plus number of scribally irregular consonant and vowel substrings. This yields an irregularity index.
 6. The vowel strings ao, ae, oe, and ye (among more obvious cases) would be illegal vowel strings. y would be illegal as a vowel in initial position and i, u, a, oa, and o would be illegal in final position. ue is legal as is y as a single, non-initial vowel.
 7. h is not allowed in final position unless preceded by c, g, or s.
 8. y and w between vowels are to be counted as consonants.
-

both position sensitive and position insensitive, correlated between .84 and .99. With such high correlations no meaningful discrimination can be made between the two approaches. Figure 5.1 shows some n -gram correlations based on type and token counts; the two sets of counts give highly similar correlations. For this reason, unless otherwise stated, we will discuss only measures based on token counts in the remainder of this chapter.

SCALE OF FREQUENCY

This analysis concerns the frequency scale to be chosen in computing frequency measures. While the effects of frequency

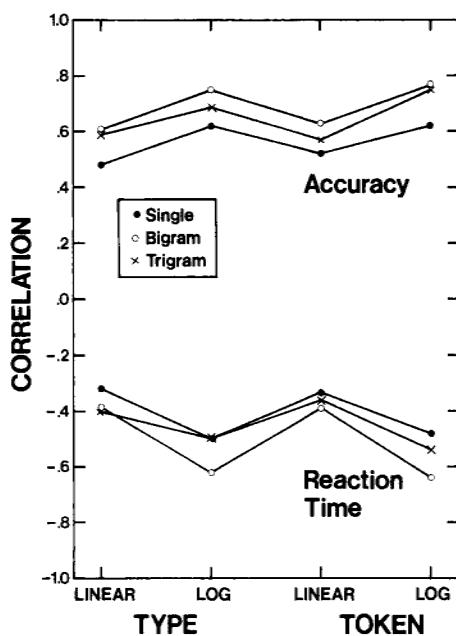


Figure 5.1 Correlations of accuracy and reaction times as a function of linear or log and type or token position-sensitive single-letter, bigram, and trigram counts.

seem to be psychologically real, it is not necessary that the mental representations of frequency directly reflect the frequency of objective counts. One alternative scale that has been successful in other research is a logarithmic (base 10) scale. Not only are there some data to suggest the possibility of a logarithmic representation (Solomon & Postman, 1952; Taylor, 1977; Travers & Olivier, 1978), but also a logarithmic representation is consistent with recent studies of number representation (Shepard, 1978) and with many other psychological scales. Therefore, we computed all of our frequency measures based upon both regular linear frequencies and log frequencies. For these measures, and all transformed scales (to be discussed), linear-frequency tables for the position-dependent *n*-grams were not rescaled before measures were computed for any item. Therefore, the two sets of measures being correlated were sums of position-dependent single letters, bigrams, and trigrams derived from either linear-frequency or log-frequency tables. In nearly every instance the log-frequency measure was more highly correlated with the performance measure than was linear frequency (see, for example, Figure 5.1).

Although log frequencies, for purposes of describing the effects of orthographic structure on recognition performance, were better than linear frequencies, we were also interested in comparing the log scale with other transformations of the linear frequencies. For this investigation, we chose our best frequency measure, summed token position-sensitive bigrams, and applied a range of transformations to it. A power function of the form

$$y = a^x \quad (5)$$

gave various scales depending on the value of the exponent *x*. Exponents of .08, .07, .05, .02, and .01 were selected. The .02 exponent approximates the log function very closely in the range of frequency values represented by our stimuli. The larger exponents yielded values intermediate between those of the linear and log scales while the .01 exponent produced values more extreme than the log function. The correlation between these bigram measures and three performance measures grew monotonically from the linear scale measure to the log and .02 exponent power function where correlations peaked and then fell for the more extreme .01 exponent power function. The correlations between average accuracy and these bigram measures were .63 for the linear scale; .77 for the log scale; and .68,

.70, .74, .77, .75 for exponents of 0.8, 0.7, 0.5, 0.2, and 0.1 respectively. These data indicate that if frequency measures are utilized by the reader, they appear to be computed on log frequencies, and so only log-frequency measures will be discussed further.

POSITION-SENSITIVE VERSUS POSITION-INSENSITIVE COUNTS

There have been many attempts to derive a measure of perceptability based upon position-insensitive frequency counts (Gibson 1964; Gibson, et al., 1970; Postman & Conger, 1954; Spoehr & Smith, 1975). These attempts have typically shown that counts that accumulate frequencies without regard to position give inadequate measures. On the other hand, position-sensitive measures have received some support as yielding better measures (Gibson et al., 1970; Mason, 1975, McClelland & Johnston, 1977). We compared the two types of measures against our data. Table 5.2 shows the correlations between the three performance measures with single-letter, bigram, and trigram counts derived from both position-sensitive and position-insensitive counts. The striking difference in this table is between the two different single-letter measures. As noted by earlier researchers (e.g., Postman & Conger, 1954), overall letter frequency without regard to position is not a very powerful predictor of recognition performance. By contrast, a position-dependent single-letter measure is almost as powerful as position-insensitive bigrams and trigrams for describing accuracy and RT. It should be noted that position-insensitive counts do better as the *n*-gram's length increases. For *n*-grams of a given length, position dependence seems to lead to a better index of structure although the advantage is essentially eliminated for trigram counts. The apparent superiority of position-sensitive bigrams over position-sensitive trigrams will be discussed next. For the remainder of these analyses, position-sensitive counts will be used.

Table 5.2 Correlations between the three dependent measures and summed log frequency counts, both position dependent and position independent.

<u>Type of Frequency Measure</u>	<u>Dependent Measures</u>		
	<u>Accuracy</u>	<u>RT</u>	<u>Rating</u>
Position Independent			
Single Letter	.20	-.36	.04
Bigram	.59	-.55	.73
Trigram	.68	-.57	.84
Position Dependent			
Single Letter	.62	-.48	.45
Bigram	.77	-.64	.83
Trigram	.75	-.54	.80

THE MOST EFFECTIVE FREQUENCY MEASURE

We have noted that summed position-sensitive log bigram frequencies had the highest correlations with our three performance measures. This measure, in fact, had the highest correlations of any of the almost 50 forms of frequency measures that we examined. For example, position-sensitive single letters account for an average of 29% of the variance in our three performance measures, while position-sensitive bigrams account for 57%. Although it might be expected that larger n -grams should naturally capture more orthographic structure than smaller n -grams, we observed that position-sensitive bigram counts of several kinds (discussed below) were better than comparable frequency measures based

upon trigrams. It is possible to hypothesize several reasons why this should be the case. First, the sampling error in the trigram tables may be large enough relative to the sampling error in the bigram tables to account for the effect. The rough equivalence between position-independent and position-dependent trigram counts might have resulted from a trade-off between the advantage of position sensitivity and the larger error variance for position-sensitive counts. Second, trigrams simply may have reached a point of diminishing returns. With bigrams accounting for 57% of the variance, there simply may be no room for noticeable improvements with larger *n*-gram measures. Finally, it could be that bigram counts are more representative of the reader's knowledge and utilization of orthographic structure. We will return to this question in a later section of this chapter.

The measures that have been discussed were based upon the sum of the frequencies for the component *n*-grams in a letter string. Englishness was described as a product of frequency ratios and would appear to provide an interesting comparison in terms of how a frequency measure should be computed from component *n*-gram frequencies. The observation that log frequency is a better scale narrows the distance between Englishness and summed *n*-gram measures since Englishness is computed as the sum of the logarithms of frequency ratios. In actual comparison, Englishness proved to be a less powerful measure than the position-dependent bigram measure. It correlated -.65 with average accuracy, .47 with average RT, and -.82 with the ratings.

It occurred to us that the normalization process of dividing a trigram frequency by the total frequency of all related trigrams might have the potential for producing an extremely powerful measure. That is essentially what Englishness does except that it is based upon position-insensitive counts. This would make particularly good theoretical sense in the case where partial visual information was available and was used to resolve or estimate unknown letters. For example, if some trigram xyz is considered at some point in a letter string, the predictability of z given the preceding letters xy is a function of the frequency of xyz relative to the total frequency of all trigrams xy_, where the final letter ranges over the entire alphabet and takes on all actually occurring values. Of course, constraints need not work only forward, the ratios xyz/x_z and xyz/_yz are also of interest. Analogously, ratios of particular bigrams xy to the general bigram sums x_ and _y can be computed.

Table 5.3 Correlations between the performance measures and summed bigram, Englishness, and constraint-ratio measures all based on log word-token frequencies.

<u>Frequency Measure</u>	<u>Performance Measure</u>		
	<u>Accuracy</u>	<u>RT</u>	<u>Rating</u>
Summed Bigrams	.77	-.64	.85
Englishness	-.65	.47	-.82
$xyz/xy_$.75	-.55	.81
xyz/x_z	.74	-.56	.82
$xyz/_yz$.74	-.54	.80
$xy/x_$.76	-.64	.86
$xy/_y$.75	-.64	.85

These five constraint ratios suggest five potentially interesting measures derived by summing the four trigram constraint ratios or five bigram constraint ratios for our six-letter strings. Again, the measures were computed as ratios of linear frequencies and as ratios of log frequencies. The counts could be based on word-type frequencies or word-token frequencies. In this case, yet another type frequency is possible, letter-type frequency. Letter-type frequency is simply the number of letters that actually occur in the free variable position. Thus, a trigram of the form s_r might consist of a family of eight or nine trigrams while q_i has only a single possible member. Letter-type frequency was taken as the reciprocal number of trigrams in a family or zero if the family has no naturally occurring member, i.e., all illegal trigrams. A similar letter-type measure was constructed for bigrams. In all, there were 30 measures computed from constraint ratios (five kinds of ratios x two frequency scales x three kinds of type and token counts).

The results for these constraint-ratio measures mirrored the results for the simple summed *n*-gram frequencies. Ratios of log frequencies consistently gave higher correlations than ratios of linear frequencies. Word-token frequencies and word-type frequencies were nearly indistinguishable in terms of their correlations with the performance measures; these frequency measures correlated with each other at .90 or higher. Letter-type frequencies correlated slightly less highly with word-type and word-token frequency measures but also did not correlate as highly with the performance measures. Finally, the bigram ratio measures gave slightly higher correlations with the performance measures than did the trigram ratios. A representative sample of these measures is shown in Table 5.3. Correlations among the frequency measures were quite high. The three trigram measures shown in Table 5.3 correlated with each

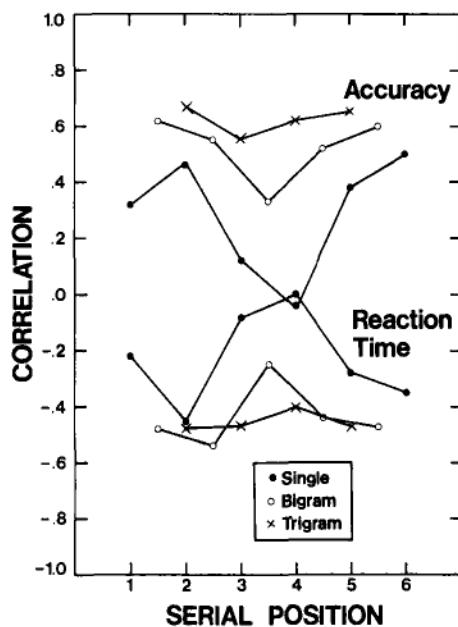


Figure 5.2 Correlations of accuracy and reaction times with single-letter, bigram, and trigram frequencies as a function of their respective serial positions in the six-letter strings.

other .98 or higher. Likewise, the two bigram constraint-ratio measures correlated .99 with each other. Given the uniformly high correlations between all the constraint-ratio measures, it is impossible to discriminate among them. Furthermore, these ratios do not account for any more variance than the computationally simpler summed log bigram measure. For these reasons, the summed log bigram measure is to be preferred over the constraint ratios.

Having found that computing more complex measures was no more effective than summing the component bigrams of a letter string, we investigated a simpler measure. We examined the correlations of the frequencies of each of the component n -grams with the performance measures. Thus, there were six correlations for single-letter frequencies, five for bigram frequencies, and four for trigram frequencies with each

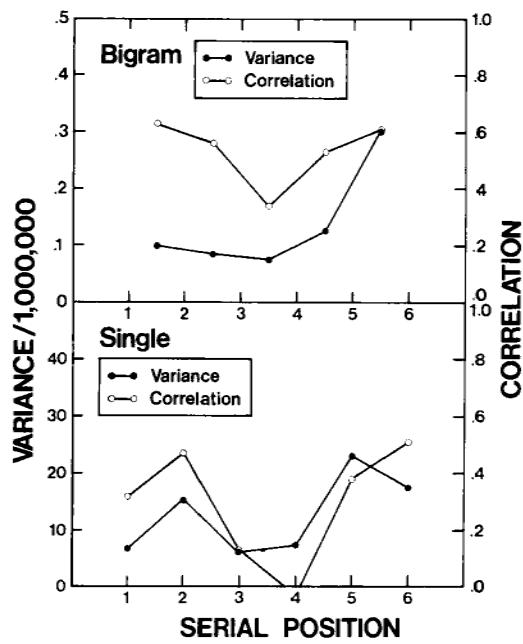


Figure 5.3 Variances of single-letter and bigram occurrences along with their observed correlations with accuracy measures at the respective serial positions in the six-letter strings.

performance measure. Figure 5.2 presents the results of this analysis. Although some of the individual n-grams did nearly as well as the summed measures, no single n-gram unit had as high a correlation as its corresponding summed measure. It is apparent from the figure that the end letter positions were the most important, but as the n-gram length increased the positional differences were diminished.

Figure 5.2 shows that the degree to which a particular n-gram predicts performance is critically dependent on its serial position, especially for the single-letter and to some extent the bigram counts. In order to evaluate whether this effect is due to the informational constraints in the stimuli themselves, we derived a measure of redundancy or predictability for each serial position. The variance of letter occurrences at each serial position was computed based on the table of frequencies given in Appendix 5.2 and 5.3. High variances occur to the extent that some letters occur more often than others. These variance measures for the single-letter and bigram counts are shown in Figure 5.3. As might be expected from phonological and scribal constraints, letters are much more predictable in initial and final positions than in the middle of six-letter words. Also presented in Figure 5.3 are the corresponding correlations with the accuracy measure. The variance and correlation measures are highly correlated for single-letter counts which means that subjects exploit redundancy directly to the degree that it is present. Although this correspondence might be expected to occur necessarily, the bigram results argue the opposite. Although the variance measures are roughly similar for the first three serial positions, the correlations are much higher for the first two positions than for the third. In addition, although the final bigram is much more predictable than the penultimate bigram, the correlations do not differ by much.

FREQUENCY VERSUS REGULARITY

A major concern of the present research has been to compare frequency with rule-governed regularity as indices of structure. We have discussed in Chapter 3 the form of experimentation that will need to be conducted to further differentiate these two indices. We have presented evidence in

Chapter 4 that regularity is at least as important for perceptual recognition and more important for overt judgments as is single-letter positional frequency. In the present chapter, we have shown that single-letter positional frequency is a good measure but that bigram positional frequency is a better measure. We have examined several dimensions of the frequency domain and presented several findings about frequency measures. This investigation of the quantitative aspects of frequency measures was relatively direct and straightforward.

A comparable investigation is not possible for regularity. We have made only a very modest start. As an initial effort to quantify regularity, we computed a simple count of orthographic irregularities based upon our preliminary attempt at a grammar of orthography (see Table 5.1). The result does rather well in correlating with performance measures but is not sufficiently developed to allow for a decisive comparison with bigram frequency. The regularity count correlated .57 with average accuracy, -.53 with average RT, and .85 with the ratings. Although these correlations with the two recognition measures are quite modest in comparison with the frequency measures described in this chapter, the correlation with ratings is comparable with the best frequency measure. While we hesitate to draw strong conclusions, it would appear that there is potential for developing not only refined rule-governed measures but also quantitatively useful structure indices based upon these rules.

One major problem with distinguishing between frequency and regularity measures of orthographic structure is that these two measures are naturally correlated in the written language. It is possible to untangle these measures as we did in the present studies. However, by independently varying single-letter positional frequency and regularity, we actually created letter strings that varied systematically with bigram frequency. Our regular strings were higher in bigram frequency than the corresponding irregular strings. Therefore, regularity and position-sensitive log bigram frequency correlated .759 in our list of 200 test items. In future experiments, we plan to independently vary these two measures of orthographic structure in perceptual recognition and overt judgment experiments.

WITHIN CLASS ANALYSIS

All of the analyses presented so far in this chapter have examined correlations between measures computed for the items across the five classes of stimulus strings. Whereas it is very important to explain this between class variance, it is equally important to examine how well several measures of orthographic structure explain the variance within each class. To do this, the correlational analyses were repeated separately for the words and the regular-high, regular-low, irregular-high, and irregular-low anagram classes. Figure 5.4 presents the correlations for the three position-sensitive log

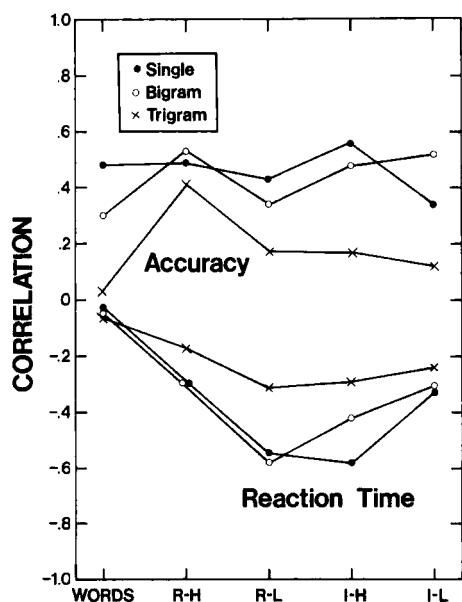


Figure 5.4 Correlations of accuracy and reaction times as a function of single-letter, bigram, and trigram frequencies for each of the five-letter string types.

n-gram counts with the accuracy and RT performance measures. As can be seen in the figure, the most significant result is the relatively poor correlations of performance with trigram frequencies. Single letter and bigram frequencies do about equally well overall. Finally, all correlations for the word RT data were close to zero. This latter result might indicate that lexical status makes a unique contribution in the RT task attenuating the contributions of orthographic structure. Future research will be necessary to address this issue.

WORD FREQUENCY

An additional measure of interest is word frequency. Word frequency has shown to be a good predictor of word recognition (Broadbent, 1967). Log word frequency correlates .64 with average accuracy and .62 with ratings; however, for just the word items the correlations are -.27 and -.11 respectively. This means that word frequency cannot be directly responsible for the effects of orthography since much higher correlations should have been seen in the class of words. In fact, a dummy variable assigning the value 1 to all words and 0 to all nonwords correlates .66 with average accuracy and .63 with ratings. Log word frequency correlates .52 and .78 with position-sensitive log bigram and trigram frequencies respectively in our set of 200 items. Therefore, lexical status itself appears to contribute very little beyond what can be accounted for by sublexical orthographic structure.

MULTIPLE REGRESSION

A series of multiple regression analyses were carried out to see how much of the variance of each of the dependent measures could be accounted for by some combination of descriptive measures. For response accuracy, log bigram frequency, log word frequency, and log single-letter frequency accounted for a total of 68% of the variance. No other descriptive variable improved on the 40% of the RT variance accounted for by log bigram frequency. Finally, 88% of the

variance of the rating judgments was accounted for by regularity count, log trigram frequency, and log word frequency. These results show that the knowledge of orthographic structure revealed by an overt judgment task does not reveal itself to the same degree in perceptual recognition tasks. This result is not unreasonable considering the additional visual processing requirements in the perceptual recognition tasks. In addition, the accuracy task allows a significantly larger contribution of orthographic structure than does the RT task. The reason for this difference appears to be due to the relatively small contribution of visual processing to the RT measure compared to the relatively large contribution of visual processing to the accuracy measure (see Chapter 4).

SUMMARY

The post hoc correlations were reasonably successful in illuminating the role of a number of variables relevant to measures of orthographic structure. Type and token counts were equally good measures for our stimulus items, whereas log counts were significantly better than linear counts. Position sensitivity is critical, especially for smaller n -gram counts. Finally, position-sensitive log bigram frequency appeared to be the most effective measure, although there appears to be some potential for developing an equally good regularity measure.

Appendix 5.1

Stimulus Items and Results

Each of the 200 stimulus items and its respective average percentage accuracy (Acc), average RT, average rating, summed position-sensitive log single-letter frequency, summed position-sensitive log bigram frequency, summed position-sensitive log trigram frequency, log word frequency, and number of irregularities.

WORDS

	Acc %	RT msec	Rating	Single Letter	Bigram	Trigram	Word Freq.	# of Irr.
action	81	644	8.8	22.33	15.26	10.98	2.46	1
amount	77	656	9.3	21.83	14.12	9.67	2.24	0
answer	89	644	8.2	22.10	14.08	8.87	2.18	1
barely	89	635	9.3	23.30	15.63	9.59	1.49	0
belong	89	603	9.5	22.76	15.12	7.99	1.57	0
breath	86	629	9.3	21.48	13.56	8.36	1.72	0
bridge	94	654	8.9	21.94	13.05	8.58	1.99	0
charge	86	642	9.4	22.28	14.51	9.61	2.09	0
double	92	621	8.9	22.13	13.56	7.94	1.75	0
famous	89	639	8.9	22.05	13.34	8.56	1.95	0
finger	92	653	9.5	22.80	15.58	9.78	1.60	0
garden	93	619	9.7	22.75	14.26	9.04	1.78	1
gather	89	673	9.4	22.87	15.91	10.23	1.30	0
glance	87	637	9.3	21.77	13.70	8.94	1.60	0
hardly	89	661	9.4	22.49	14.87	9.06	2.03	2
itself	79	637	9.2	21.03	13.63	10.25	2.48	1
lawyer	86	646	8.7	21.60	12.50	6.99	1.63	0
longer	94	627	9.6	22.96	15.67	10.20	2.29	0
master	93	627	9.7	23.55	16.56	10.54	1.86	0
modern	93	616	9.5	22.33	14.40	9.64	2.30	0
mother	93	610	9.5	23.31	16.47	11.10	2.33	0
namely	90	662	9.4	22.53	14.84	8.22	1.52	0
nearly	89	620	9.3	22.52	14.76	9.47	2.15	1
permit	83	590	9.6	22.52	14.16	9.01	1.89	1
player	95	619	9.6	21.85	14.61	9.47	1.71	0
poetry	93	642	9.2	22.80	12.36	8.09	1.94	1
purely	92	607	9.1	23.03	15.85	9.33	1.48	0
reason	85	605	9.1	22.46	15.60	10.76	2.38	0
remain	85	638	9.5	22.19	14.42	9.15	1.97	0
result	86	636	9.7	23.02	14.56	9.74	2.39	0
shared	93	640	9.4	23.55	15.92	9.74	1.60	0
simple	91	622	9.4	22.94	15.01	10.78	2.21	1
single	91	632	9.2	23.02	14.58	9.30	2.24	0
spread	91	611	9.2	22.60	13.42	8.78	1.92	0
stared	95	672	9.4	23.44	16.32	10.10	1.78	0
stream	89	616	9.5	21.94	14.19	9.85	1.71	0
travel	85	622	9.6	21.87	13.51	8.53	1.79	0
trials	82	623	9.2	22.53	12.64	7.07	1.59	0
turned	85	675	9.3	23.42	15.57	10.74	2.51	1
winter	91	621	9.7	23.11	16.34	10.32	1.92	1

REGULAR-HIGH

	Acc %	RT msec	Rating	Single Letter	Bigram	Trigram	Word Freq.	# of Irr.
cainot	82	645	6.5	22.69	13.74	4.38	0.00	0
mauton	70	620	7.5	22.49	13.54	2.28	0.00	0
swaner	81	615	7.6	22.25	13.40	5.50	0.00	0
blayer	81	666	8.0	21.79	13.78	8.14	0.00	0
goblen	81	638	8.6	21.94	13.35	3.60	0.00	0
thaber	78	637	7.6	22.42	14.95	5.14	0.00	0
begrid	78	640	7.8	22.27	12.89	3.96	0.00	0
chager	86	673	8.3	22.90	15.27	7.54	0.00	0
bodule	77	673	8.0	22.78	13.92	2.63	0.00	0
foamus	74	670	7.3	21.80	10.39	0.00	0.00	0
fering	83	617	7.7	23.22	15.85	9.73	0.00	0
nagred	83	659	7.0	22.66	13.74	4.75	0.00	0
thager	72	612	7.7	22.70	15.29	6.39	0.00	0
caleng	77	659	5.6	22.85	15.50	6.17	0.00	0
hyrald	69	677	6.6	21.84	9.21	3.10	0.00	0
siflet	72	640	7.7	23.06	12.69	5.33	0.00	0
weraly	81	649	5.7	22.87	13.87	3.36	0.00	0
loqren	79	604	7.4	22.53	12.07	3.89	0.00	0
tasmer	77	648	7.3	22.92	13.94	3.87	0.00	0
dormen	87	625	8.2	22.92	14.56	6.00	0.00	1
thomer	87	654	7.3	22.68	15.54	8.41	0.00	0
manely	87	639	8.4	22.96	15.33	8.76	0.00	0
lareny	79	656	7.2	23.15	14.18	5.68	0.00	0
primet	81	629	8.0	22.64	13.97	4.78	0.00	0
raplev	88	634	7.7	22.49	11.86	2.62	0.00	0
petory	89	645	7.3	23.13	14.12	5.36	0.00	0
yulper	74	662	8.0	21.17	10.87	5.09	0.00	1
sarone	83	635	7.8	23.83	14.63	6.14	0.00	0
ramine	82	605	8.2	23.29	14.84	8.42	0.00	0
surtel	82	610	7.3	23.43	14.64	5.23	0.00	1
shread	79	676	9.2	23.19	13.92	7.94	0.00	0
pimsel	81	641	7.7	22.48	11.77	2.36	0.00	1
snigel	81	604	6.8	22.42	12.58	1.34	0.00	0
prased	87	639	8.4	23.18	15.33	6.16	0.00	0
sarted	91	622	8.4	24.44	16.02	6.87	0.00	1
trames	80	582	8.3	22.80	14.11	7.74	0.00	0
vartle	79	638	7.2	22.95	15.01	8.74	0.00	1
sartil	70	672	7.3	22.85	12.89	3.90	0.00	1
drunet	81	636	6.8	22.49	13.05	2.81	0.00	0
triwen	72	672	6.6	21.90	11.54	1.90	0.00	0

REGULAR-LOW

	Acc %	RT msec	Rating	Single Letter	Bigram	Trigram	Word Freq.	# of Irr.
onicat	71	637	6.5	20.84	10.31	0.60	0.00	0
otanum	59	668	6.9	20.12	12.46	2.65	0.00	0
erswan	71	675	6.0	21.01	8.35	0.00	0.00	1
byeral	77	667	4.8	20.60	8.29	1.20	0.00	1
glenob	79	660	7.3	19.23	9.63	1.74	0.00	0
therab	75	666	7.0	19.48	9.68	3.39	0.00	0
gredib	66	672	5.9	18.96	10.13	5.34	0.00	0
greach	68	648	8.7	21.03	13.62	6.69	0.00	0
edolub	68	679	5.3	17.97	4.79	0.00	0.00	0
osafum	53	669	4.4	18.41	5.98	0.00	0.00	0
grefin	81	633	7.0	20.20	12.27	4.00	0.00	0
gredan	85	653	7.0	21.13	13.02	4.13	0.00	0
greath	83	601	8.1	21.00	13.66	8.11	0.00	0
glecan	81	648	6.5	20.67	10.86	0.00	0.00	0
hydral	72	654	7.6	19.92	6.68	1.20	0.00	0
flesit	71	621	7.6	21.38	11.58	1.00	0.00	0
lyeraw	73	690	6.8	19.87	6.87	0.00	0.00	1
gleron	80	621	8.0	21.27	12.32	2.12	0.00	0
steram	79	626	7.6	21.21	11.08	2.84	0.00	0
drenom	75	627	7.1	21.10	12.93	3.93	0.00	0
therom	80	658	7.6	21.07	12.72	3.39	0.00	0
lyeman	66	693	8.1	20.53	9.64	2.16	0.00	1
nyelar	77	656	5.3	20.64	7.45	1.95	0.00	1
trepim	73	645	7.2	20.67	10.67	1.74	0.00	0
lyerap	75	702	5.9	19.08	6.47	0.00	0.00	1
tyerop	72	647	5.3	19.29	7.49	0.00	0.00	1
ryelup	73	684	6.1	18.59	4.56	0.00	0.00	1
erosan	73	621	6.9	21.72	11.75	1.08	0.00	0
amerin	77	638	7.0	20.78	11.53	1.26	0.00	0
tresul	74	636	6.9	20.74	11.28	2.22	0.00	0
dreash	73	686	7.9	21.20	12.29	6.56	0.00	0
emslip	73	638	6.7	18.97	6.56	1.32	0.00	1
glesin	71	632	7.1	20.90	11.50	2.71	0.00	0
dresap	75	652	6.8	19.73	9.52	2.48	0.00	0
dresat	87	647	6.3	21.54	11.21	2.48	0.00	0
tresam	76	635	7.0	20.79	11.26	2.52	0.00	0
treval	84	663	8.1	20.78	11.91	1.74	0.00	0
trisal	71	608	7.1	21.33	13.45	4.13	0.00	0
tredun	72	641	7.7	20.99	9.29	3.55	0.00	0
trewin	79	651	7.3	20.91	10.96	2.34	0.00	0

IRREGULAR-HIGH

	Acc %	RT msec	Rating	Single Letter	Bigram	Trigram	Word Freq.	# of Irr.
aiocnt	63	656	3.1	22.36	7.20	0.00	0.00	2
auotnm	69	652	3.7	22.07	5.49	1.49	0.00	3
wnraes	70	666	2.4	22.93	8.58	1.18	0.00	3
lbraey	81	699	3.0	21.76	7.90	1.18	0.00	2
lnboeg	65	692	2.1	21.87	4.07	0.60	0.00	3
bhrtea	75	664	1.5	22.22	9.43	1.59	0.00	4
ierbgd	66	728	2.3	21.96	5.31	1.32	0.00	3
ahcger	71	674	3.0	22.64	6.86	2.73	0.00	2
leoubd	76	691	2.5	21.38	10.15	0.00	0.00	2
uaomfs	61	698	2.0	20.03	3.61	0.00	0.00	4
rfnieg	70	656	2.2	22.21	5.60	1.83	0.00	2
arnged	84	663	3.8	22.98	14.28	6.08	0.00	1
hrtaeg	70	663	2.3	22.50	5.98	0.00	0.00	3
lcnaeq	67	675	2.2	22.06	3.46	0.00	0.00	3
yrahld	61	681	2.4	21.14	9.17	1.34	0.00	2
tlfies	68	664	2.6	22.89	10.79	3.30	0.00	2
rwlaey	73	688	2.1	21.52	6.14	0.00	0.00	3
glnoer	78	662	2.9	22.42	9.15	0.00	0.00	2
mrtaes	76	657	2.3	23.27	9.56	0.00	0.00	3
oermnd	71	650	2.8	22.62	10.37	2.30	0.00	3
mroteh	82	660	2.8	22.84	8.68	2.53	0.00	2
lmnaey	70	705	2.6	21.74	5.91	0.00	0.00	3
lrnaey	69	648	2.8	22.70	7.39	0.00	0.00	3
tmriep	68	662	2.7	20.72	7.81	2.31	0.00	2
plraey	84	673	2.9	22.96	8.90	1.18	0.00	2
yptoer	70	692	2.6	21.13	8.56	0.70	0.00	3
lpyuer	68	671	3.0	21.70	6.44	0.00	0.00	2
snaoer	75	645	4.3	23.41	8.96	1.56	0.00	2
mraien	75	612	3.4	23.27	11.40	1.97	0.00	2
lrtues	71	656	2.0	23.34	11.26	2.42	0.00	2
srdaeh	67	680	2.1	22.36	4.32	0.00	0.00	4
ilmpes	74	635	3.8	22.08	11.74	3.47	0.00	2
ilgnes	68	668	4.5	22.00	8.80	3.72	0.00	2
prsaed	85	623	3.3	23.25	9.44	0.00	0.00	2
srtaed	86	676	2.2	23.75	9.92	0.00	0.00	3
mtraes	72	647	2.3	23.20	10.13	2.82	0.00	2
avrlet	74	656	4.4	21.71	10.12	3.63	0.00	2
iastlr	81	650	2.6	22.67	9.07	4.33	0.00	3
rdnuet	75	712	2.1	21.74	7.91	0.00	0.00	2
wrntei	63	696	2.4	20.67	11.47	2.76	0.00	3

IRREGULAR-LOW

	Acc %	RT msec	Rating	Single Letter	Bigram	Trigram	Word Freq.	# of Irr.
icntoa	62	683	2.0	20.38	6.29	0.48	0.00	3
nmtaou	55	707	2.1	19.06	2.62	0.00	0.00	4
rnwesa	65	704	2.2	20.26	6.02	1.45	0.00	3
rbleya	67	728	2.4	18.96	2.98	0.00	0.00	2
ebgnlo	69	670	2.7	19.04	5.08	0.00	0.00	4
rtbeha	72	736	2.0	19.78	3.45	0.00	0.00	3
ebrgdi	58	697	2.0	18.09	4.46	0.00	0.00	3
hreagc	68	646	2.4	20.49	7.66	2.37	0.00	2
obdeul	73	692	4.6	19.15	8.15	0.00	0.00	1
ofmsua	62	710	2.5	18.80	6.02	0.00	0.00	3
efnqri	71	697	2.1	18.98	6.80	1.36	0.00	3
ednrga	67	698	2.4	19.25	6.05	0.00	0.00	4
grheta	77	655	3.2	20.07	9.11	0.00	0.00	2
eclnqa	65	738	2.6	20.09	5.10	0.00	0.00	2
rdlhya	53	743	1.8	18.53	0.00	0.00	0.00	4
eflsti	66	707	2.7	18.72	7.60	0.00	0.00	2
eyrlwa	67	707	2.8	18.53	4.09	0.00	0.00	4
egnrlo	67	655	2.9	19.24	6.11	0.00	0.00	4
emrtsa	66	629	2.3	20.19	4.25	0.00	0.00	4
rmedno	71	644	3.1	19.48	3.88	0.00	0.00	3
hretmo	69	633	2.4	19.84	5.25	2.03	0.00	3
emnyla	73	695	4.5	19.05	4.15	0.00	0.00	2
enyrla	78	677	3.0	20.20	8.91	0.00	0.00	2
rtpeim	67	652	2.4	20.65	4.62	0.00	0.00	2
epylra	70	670	3.1	19.56	3.29	0.00	0.00	2
ervtpo	71	659	3.2	19.23	5.59	0.00	0.00	2
erplvu	66	666	2.8	18.37	3.54	0.00	0.00	2
rnseoa	63	648	2.4	21.08	5.96	1.51	0.00	4
rniema	69	656	2.3	20.44	8.26	0.48	0.00	2
elsrtu	71	664	2.2	19.52	4.51	0.00	0.00	4
hsreda	75	656	2.4	19.90	7.92	0.70	0.00	3
lsepmi	69	672	2.8	18.23	4.49	0.00	0.00	3
nglesi	65	662	2.4	18.25	5.28	2.00	0.00	2
dsepra	82	665	3.6	19.81	5.25	0.00	0.00	2
dtsera	77	668	2.4	21.03	10.08	5.92	0.00	3
esrtma	74	681	3.4	20.06	7.24	0.00	0.00	4
tlerav	67	680	2.4	17.70	6.34	1.28	0.00	2
irltsa	70	655	2.8	20.53	3.39	0.00	0.00	4
edtrnu	71	656	2.1	19.12	4.93	0.00	0.00	3

Appendix 5.2

Single-Letter Positional Frequencies (Token Counts)
According to Serial Position and Word Length for Words
3 to 7 Letters in Length.

Instructions for Using Frequency Tables

The words used to obtain the single-letter frequencies came from the Kucera and Francis (1967) word count. Words which had a Kucera and Francis count of at least three and which appeared in at least three different sampled texts were selected. They were then separated for each word length. These counts were multiplied by the frequency of each of the words so that the single-letter frequencies of the tables are based on token rather than type counts (see Chapter 1). This selection procedure yielded the following number of word tokens for each word length: 210990, 153315, 102699, 76295, 68894 for 3 through 7 letter words, respectively.

The number of times each letter of the alphabet occurred in each letter position of the words was counted. These totals, along with their sum, are presented in the tables as the first row of numbers for each letter. Thus, for example, for six-letter words, the letter k occurred in the sixth letter position 231 times and occurred in all positions a total of 3,804 times. Below each of the frequencies is the base 10 logarithm of the frequency count.

It should be noted that for some single letters, the frequency of occurrence is zero in some positions. The logarithm of zero is undefined, but for the purposes of the tables was set equal to zero.

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	LETTER POSITION			
	1ST	2ND	3RD	
			SUM	
A	39519	28085	221	67825
	4.597	4.448	2.344	4.831
B	5902	3	364	6269
	3.771	.477	2.561	3.797
C	2466	379	186	3031
	3.392	2.579	2.270	3.482
D	2301	142	37652	40095
	3.362	2.152	4.576	4.603
E	760	11050	82860	94670
	2.881	4.043	4.918	4.976
F	10988	644	639	12271
	4.041	2.809	2.806	4.089
G	1967	498	790	3255
	3.294	2.697	2.898	3.513
H	21515	75511	16	97042
	4.333	4.878	1.204	4.987
I	2009	13084	7	15100
	3.303	4.117	.845	4.179
J	515	0	0	515
	2.712	0.000	0.000	2.712
K	182	62	150	394
	2.260	1.792	2.176	2.595
L	1604	3766	3192	8562
	3.205	3.576	3.504	3.933
M	4180	15	3108	7303
	3.621	1.176	3.492	3.864

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	LETTER POSITION			
	1ST	2ND	3RD	
			SUM	
N	7865	33973	6169	48007
	3.896	4.531	3.790	4.681
O	8900	23643	4829	37372
	3.949	4.374	3.684	4.573
P	1227	27	474	1728
	3.089	1.431	2.676	3.238
Q	11	0	0	11
	1.041	0.000	0.000	1.041
R	864	5591	16036	22491
	2.937	3.747	4.205	4.352
S	6354	736	22017	29107
	3.803	2.867	4.343	4.464
T	73207	1937	16145	91289
	4.865	3.287	4.208	4.960
U	592	9472	3307	13371
	2.772	3.976	3.519	4.126
V	93	41	73	207
	1.968	1.613	1.863	2.316
W	14079	2199	5380	21658
	4.149	3.342	3.731	4.336
X	0	6	671	677
	0.000	.778	2.827	2.831
Y	3855	126	6704	10685
	3.586	2.100	3.826	4.029
Z	35	0	0	35
	1.544	0.000	0.000	1.544

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	LETTER POSITION				SUM
	1ST	2ND	3RD	4TH	
A	3077	26766	20259	953	51055
	3.488	4.428	4.307	2.979	4.708
B	7266	244	253	301	8064
	3.861	2.387	2.403	2.479	3.907
C	4530	90	6697	54	11371
	3.656	1.954	3.826	1.732	4.056
D	4614	393	2462	10129	17598
	3.664	2.594	3.391	4.006	4.245
E	3714	20802	21789	32881	79186
	3.570	4.318	4.338	4.517	4.899
F	10556	0	1640	528	12724
	4.023	0.000	3.215	2.723	4.105
G	3091	128	1169	1351	5739
	3.490	2.107	3.068	3.131	3.759
H	9784	29988	376	12322	52470
	3.991	4.477	2.575	4.091	4.720
I	2191	22145	9661	35	34032
	3.341	4.345	3.985	1.544	4.532
J	2051	0	0	0	2051
	3.312	0.000	0.000	0.000	3.312
K	2252	85	3066	5572	10975
	3.353	1.929	3.487	3.746	4.040
L	7898	2843	11225	7346	29312
	3.898	3.454	4.050	3.866	4.467
M	10777	50	6902	8373	26102
	4.032	1.699	3.839	3.923	4.417

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	LETTER POSITION				
	1ST	2ND	3RD	4TH	SUM
N	2119	5753	8265	14256	30393
	3.326	3.760	3.917	4.154	4.483
O	4261	24056	10798	3158	42273
	3.630	4.381	4.033	3.499	4.626
P	3518	978	1040	1715	7251
	3.546	2.990	3.017	3.234	3.860
Q	15	0	0	3	18
	1.176	0.000	0.000	.477	1.255
R	3075	6514	13598	5335	28522
	3.488	3.814	4.133	3.727	4.455
S	10815	742	10091	11497	33145
	4.034	2.870	4.004	4.061	4.520
T	30636	563	12251	24894	68344
	4.486	2.751	4.088	4.396	4.835
U	1340	7042	2751	38	11171
	3.127	3.848	3.439	1.580	4.048
V	1308	2843	5723	0	9874
	3.117	3.454	3.758	0.000	3.994
W	22367	555	1526	2163	26611
	4.350	2.744	3.184	3.335	4.425
X	0	62	470	77	609
	0.000	1.792	2.672	1.886	2.785
Y	1996	673	998	10214	13881
	3.300	2.828	2.999	4.009	4.142
Z	64	0	305	120	489
	1.806	0.000	2.484	2.079	2.689

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	LETTER POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
A	7359 3.867	10455 4.019	12306 4.090	5563 3.745	523 2.719	36206 4.559
B	5264 3.721	2151 3.333	826 2.917	438 2.641	36 1.556	8715 3.940
C	6416 3.807	556 2.745	938 2.972	8332 3.921	582 2.765	16824 4.226
D	2379 3.376	625 2.796	2681 3.428	2860 3.456	11845 4.074	20390 4.309
E	2189 3.340	9709 3.987	14845 4.172	18479 4.267	21152 4.325	66374 4.822
F	5614 3.749	1518 3.181	314 2.497	447 2.650	493 2.693	8386 3.924
G	3565 3.552	700 2.845	3796 3.579	1475 3.169	3349 3.525	12885 4.110
H	3348 3.525	18808 4.274	1740 3.241	2465 3.392	7039 3.848	33400 4.524
I	917 2.962	9821 3.992	13572 4.133	6832 3.835	75 1.875	31217 4.494
J	525 2.720	0 0.000	291 2.464	0 0.000	0 0.000	816 2.912
K	618 2.791	85 1.929	1182 3.073	1242 3.094	2195 3.341	5322 3.726
L	4187 3.622	4783 3.680	2801 3.447	11708 4.068	4948 3.694	28427 4.454
M	4220 3.625	1357 3.133	2103 3.323	1641 3.215	404 2.606	9725 3.988

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	LETTER POSITION				
	1ST	2ND	3RD	4TH	5TH
N	2336	2899	4331	8513	5318
	3.368	3.462	3.637	3.930	3.726
O	3064	18464	10612	2682	481
	3.486	4.266	4.026	3.428	2.682
P	4418	1255	1046	879	747
	3.645	3.099	3.020	2.944	2.873
Q	515	115	0	0	0
	2.712	2.061	0.000	0.000	0.000
R	3036	7818	8523	9276	11791
	3.482	3.893	3.931	3.967	4.072
S	11526	959	2106	8108	15072
	4.062	2.982	3.323	3.909	4.178
T	14043	5756	4727	6054	11098
	4.147	3.760	3.675	3.782	4.045
U	2038	3058	9278	3139	8
	3.309	3.485	3.967	3.497	.903
V	938	687	2979	1239	9
	2.972	2.837	3.474	3.093	.954
W	12634	682	755	1121	395
	4.102	2.834	2.878	3.050	2.597
X	0	151	364	7	159
	0.000	2.179	2.561	.845	2.201
Y	1547	287	457	102	4977
	3.189	2.458	2.660	2.009	3.697
Z	3	0	126	97	3
	.477	0.000	2.100	1.987	.477
					229
					2.360

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	LETTER POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
A	5529	11057	6081	4226	3257	449	30599
	3.743	4.044	3.784	3.626	3.513	2.652	4.486
B	5116	284	1102	1286	79	40	7907
	3.709	2.453	3.042	3.109	1.898	1.602	3.898
C	6237	1753	3765	1494	3384	743	17376
	3.795	3.244	3.576	3.174	3.529	2.871	4.240
D	3146	294	1648	2638	1284	13473	22483
	3.498	2.468	3.217	3.421	3.109	4.129	4.352
E	3313	14308	3548	7619	23111	12471	64370
	3.520	4.156	3.550	3.882	4.364	4.096	4.809
F	3953	903	2424	347	57	593	8277
	3.597	2.956	3.385	2.540	1.756	2.773	3.918
G	1712	211	1746	2500	2425	3784	12378
	3.234	2.324	3.242	3.398	3.385	3.578	4.093
H	2310	4498	1292	2784	1096	2531	14511
	3.364	3.653	3.111	3.445	3.040	3.403	4.162
I	1668	7396	4374	10676	2909	45	27068
	3.222	3.869	3.641	4.028	3.464	1.653	4.432
J	541	0	140	0	0	0	681
	2.733	0.000	2.146	0.000	0.000	0.000	2.833
K	225	73	732	2002	541	231	3804
	2.352	1.863	2.865	3.301	2.733	2.364	3.580
L	3918	3028	4417	3448	8721	2771	26303
	3.593	3.481	3.645	3.538	3.941	3.443	4.420
M	4718	548	3912	2143	1297	980	13598
	3.674	2.739	3.592	3.331	3.113	2.991	4.133

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	LETTER POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
N	2043	3306	4535	4451	8005	5286	27626
	3.310	3.519	3.657	3.648	3.903	3.723	4.441
O	1398	10986	6715	5925	5221	172	30417
	3.146	4.041	3.827	3.773	3.718	2.236	4.483
P	5827	1154	1398	2401	786	109	11675
	3.765	3.062	3.146	3.380	2.895	2.037	4.067
Q	70	166	91	64	0	0	391
	1.845	2.220	1.959	1.806	0.000	0.000	2.592
R	4276	4923	9119	3680	6188	8147	36333
	3.631	3.692	3.960	3.566	3.792	3.911	4.560
S	11423	862	4710	2784	2717	10904	33400
	4.058	2.936	3.673	3.445	3.434	4.038	4.524
T	3929	3427	7531	7658	3149	6979	32673
	3.594	3.535	3.877	3.884	3.498	3.844	4.514
U	1079	5252	2800	5928	1027	61	16147
	3.033	3.720	3.447	3.773	3.012	1.785	4.208
V	1073	183	1606	984	364	0	4210
	3.031	2.262	3.206	2.993	2.561	0.000	3.624
W	2645	308	1382	772	110	670	5887
	3.422	2.489	3.141	2.888	2.041	2.826	3.770
X	0	710	154	13	3	31	911
	0.000	2.851	2.188	1.114	.477	1.491	2.960
Y	140	665	999	350	494	5825	8473
	2.146	2.823	3.000	2.544	2.694	3.765	3.928
Z	6	0	74	122	70	0	272
	.778	0.000	1.869	2.086	1.845	0.000	2.435

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	LETTER POSITION							
	1ST	2ND	3RD	4TH	5TH	6TH	7TH	SUM
A	4796 3.681	6979 3.844	7353 3.866	4582 3.661	5253 3.720	5036 3.702	567 2.754	34566 4.539
B	4147 3.618	407 2.610	690 2.839	1357 3.133	611 2.786	85 1.929	19 1.279	7316 3.864
C	7268 3.861	946 2.976	3977 3.600	2921 3.466	1735 3.239	2707 3.432	355 2.550	19909 4.299
D	2846 3.454	372 2.571	1168 3.067	2263 3.355	1202 3.080	2059 3.314	8884 3.949	18794 4.274
E	3068 3.487	15655 4.195	4768 3.678	6532 3.815	10073 4.003	16835 4.226	9106 3.959	66037 4.820
F	3548 3.550	623 2.794	1162 3.065	859 2.934	458 2.661	132 2.121	769 2.886	7551 3.878
G	1982 3.297	663 2.822	1493 3.174	1564 3.194	1803 3.256	2323 3.366	7215 3.858	17043 4.232
H	3232 3.509	3963 3.598	501 2.700	2243 3.351	2874 3.458	1035 3.015	1644 3.216	15492 4.190
I	1268 3.103	6571 3.818	4053 3.608	6031 3.780	15115 4.179	1735 3.239	18 1.255	34791 4.541
J	471 2.673	0 0.000	167 2.223	254 2.405	0 0.000	0 0.000	0 0.000	892 2.950
K	532 2.726	48 1.681	69 1.839	1280 3.107	524 2.719	137 2.137	120 2.079	2710 3.433
L	1990 3.299	2524 3.402	3526 3.547	4545 3.658	3125 3.495	4433 3.647	3830 3.583	23973 4.380
M	3649 3.562	917 2.962	2923 3.466	1730 3.238	983 2.993	656 2.817	1478 3.170	12336 4.091

SINGLE LETTER FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	LETTER POSITION							
	1ST	2ND	3RD	4TH	5TH	6TH	7TH	SUM
N	1628	3150	5166	3785	3401	11623	4650	33403
	3.212	3.498	3.713	3.578	3.532	4.065	3.667	4.524
O	1460	8667	5792	2962	5046	2700	210	26837
	3.164	3.938	3.763	3.472	3.703	3.431	2.322	4.429
P	6067	1211	2154	2655	994	576	138	13795
	3.783	3.083	3.333	3.424	2.997	2.760	2.140	4.140
Q	358	113	199	26	23	0	0	719
	2.554	2.053	2.299	1.415	1.362	0.000	0.000	2.857
R	3726	6991	8487	4046	4791	6193	4948	39182
	3.571	3.845	3.929	3.607	3.680	3.792	3.694	4.593
S	8092	558	4162	4055	1224	4233	12054	34378
	3.908	2.747	3.619	3.608	3.088	3.627	4.081	4.536
T	3731	1976	5538	8755	4215	3466	6324	34005
	3.572	3.296	3.743	3.942	3.625	3.540	3.801	4.532
U	603	4693	2573	3082	3862	1840	3	16656
	2.780	3.671	3.410	3.489	3.587	3.265	.477	4.222
V	1043	386	1421	1442	1106	629	0	6027
	3.018	2.587	3.153	3.159	3.044	2.799	0.000	3.780
W	3289	161	764	1544	269	254	116	6397
	3.517	2.207	2.883	3.189	2.430	2.405	2.064	3.806
X	0	954	314	0	31	38	114	1451
	0.000	2.980	2.497	0.000	1.491	1.580	2.057	3.162
Y	93	366	392	304	116	66	6332	7669
	1.968	2.563	2.593	2.483	2.064	1.820	3.802	3.885
Z	7	0	82	77	60	103	0	329
	.845	0.000	1.914	1.886	1.778	2.013	0.000	2.517

Appendix 5.3

Bigram Positional Frequencies (Token Counts)
According to Serial Position and Word Length for Words
3 to 7 Letters in Length.

Instructions for using frequency tables

These counts were based on the same word tokens as were the single-letter counts in Appendix 5.2. In this case, however, each word was broken down into its bigram components (i.e., battle----> ba, at, tt, tl, le) and the frequency of occurrence of each of the bigrams was tabulated for each word length. For example, for six-letter words, the bigram ap occurred as the initial bigram 185 times, as the second bigram 160 times, as the third bigram 148 times, etc. Notice that aq never occurred as the sixth bigram. Also notice that the bigram ao never occurred at all. A bigram which did not occur at least once for words of a given length was omitted from the listing in the tables.

It should be noted that for some bigrams, the frequency of occurrence is zero in some positions. The logarithm of zero is undefined, but for the purposes of the tables was set equal to zero.

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

BIGRAM POSITION			
	1ST	2ND	SUM
AB	3 .477	12 1.079	15 1.176
AC	298 2.474	0 0.000	298 2.474
AD	98 1.991	5378 3.731	5476 3.738
AF	5 .699	0 0.000	5 .699
AG	473 2.675	75 1.875	548 2.739
AI	424 2.627	0 0.000	424 2.627
AK	0 0.000	15 1.176	15 1.176
AL	3008 3.478	33 1.519	3041 3.483
AM	15 1.176	122 2.086	137 2.137
AN	30232 4.480	3316 3.521	33548 4.526
AO	0 0.000	24 1.380	24 1.380
AP	18 1.255	98 1.991	116 2.064
AR	4736 3.675	1324 3.122	6060 3.782
AS	144 2.158	12364 4.092	12508 4.097
AT	16 1.204	414 2.617	430 2.633
AU	25 1.398	0 0.000	25 1.398
AV	13 1.114	0 0.000	13 1.114
AW	5 .699	713 2.853	718 2.856
AX	6 .778	226 2.354	232 2.365
AY	0 0.000	3971 3.599	3971 3.599
BA	348 2.542	0 0.000	348 2.542

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
BE	194	3	197
	2.288	.477	2.294
BI	492	0	492
	2.692	0.000	2.692
BO	367	0	367
	2.565	0.000	2.565
BU	4501	0	4501
	3.653	0.000	3.653
CA	2108	0	2108
	3.324	0.000	3.324
CE	0	60	60
	0.000	1.778	1.778
CO	73	0	73
	1.863	0.000	1.863
CR	48	0	48
	1.681	0.000	1.681
CT	0	307	307
	0.000	2.487	2.487
CU	237	0	237
	2.375	0.000	2.375
CY	0	12	12
	0.000	1.079	1.079
DA	740	0	740
	2.869	0.000	2.869
DD	0	132	132
	0.000	2.121	2.121
DE	44	0	44
	1.643	0.000	1.643
DI	1152	0	1152
	3.061	0.000	3.061
DO	135	0	135
	2.130	0.000	2.130
DR	73	0	73
	1.863	0.000	1.863
DS	0	10	10
	0.000	1.000	1.000
DU	157	0	157
	2.196	0.000	2.196
EA	90	126	216
	1.954	2.100	2.334
EB	0	26	26
	0.000	1.415	1.415

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

BIGRAM POSITION			
	1ST	2ND	SUM
EC	0	36	36
	0.000	1.556	1.556
ED	0	513	513
	0.000	2.710	2.710
EE	0	845	845
	0.000	2.927	2.927
EG	25	69	94
	1.398	1.839	1.973
EI	0	7	7
	0.000	.845	.845
EL	3	16	19
	.477	1.204	1.279
EM	0	8	8
	0.000	.903	.903
EN	410	1083	1493
	2.613	3.035	3.174
EO	0	7	7
	0.000	.845	.845
EP	0	13	13
	0.000	1.114	1.114
ER	30	3408	3438
	1.477	3.532	3.536
ES	3	155	158
	.477	2.190	2.199
ET	58	2243	2301
	1.763	3.351	3.362
EV	19	33	52
	1.279	1.519	1.716
EW	0	2262	2262
	0.000	3.354	3.354
EX	0	97	97
	0.000	1.987	1.987
EY	122	103	225
	2.086	2.013	2.352
FA	505	0	505
	2.703	0.000	2.703
FE	679	0	679
	2.832	0.000	2.832
FF	0	639	639
	0.000	2.806	2.806
FI	161	0	161
	2.207	0.000	2.207

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
FL	51	0	51
	1.708	0.000	1.708
FO	9535	0	9535
	3.979	0.000	3.979
FT	0	5	5
	0.000	.699	.699
FU	57	0	57
	1.756	0.000	1.756
GA	154	0	154
	2.188	0.000	2.188
GE	781	227	1008
	2.893	2.356	3.003
GG	0	12	12
	0.000	1.079	1.079
GI	27	0	27
	1.431	0.000	1.431
GO	819	259	1078
	2.913	2.413	3.033
GU	186	0	186
	2.270	0.000	2.270
HA	7694	0	7694
	3.886	0.000	3.886
HE	3078	72830	75908
	3.488	4.862	4.880
HI	9747	0	9747
	3.989	0.000	3.989
HO	970	2252	3222
	2.987	3.353	3.508
HU	26	0	26
	1.415	0.000	1.415
HY	0	429	429
	0.000	2.632	2.632
IA	0	48	48
	0.000	1.681	1.681
IC	57	8	65
	1.756	.903	1.813
ID	0	1301	1301
	0.000	3.114	3.114
IE	0	169	169
	0.000	2.228	2.228
IG	0	455	455
	0.000	2.658	2.658

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

BIGRAM POSITION			
	1ST	2ND	SUM
IK	4 .602	0 0.000	4 .602
IL	39 1.591	93 1.968	132 2.121
IM	0 0.000	2741 3.438	2741 3.438
IN	36 1.556	168 2.225	204 2.310
IO	6 .778	4 .602	10 1.000
IP	0 0.000	88 1.944	88 1.944
IR	0 0.000	352 2.547	352 2.547
IS	0 0.000	6997 3.845	6997 3.845
IT	1858 3.269	413 2.616	2271 3.356
IV	9 .954	0 0.000	9 .954
IX	0 0.000	247 2.393	247 2.393
JA	77 1.886	0 0.000	77 1.886
JE	63 1.799	0 0.000	63 1.799
JI	36 1.556	0 0.000	36 1.556
JO	333 2.522	0 0.000	333 2.522
JU	6 .778	0 0.000	6 .778
KA	21 1.322	0 0.000	21 1.322
KE	100 2.000	4 .602	104 2.017
KI	61 1.785	0 0.000	61 1.785
KY	0 0.000	58 1.763	58 1.763
LA	491 2.691	17 1.230	508 2.706

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
LD	0	660	660
	0.000	2.820	2.820
LE	618	0	618
	2.791	0.000	2.791
LI	113	0	113
	2.053	0.000	2.053
LL	0	3040	3040
	0.000	3.483	3.483
LM	0	3	3
	0.000	.477	.477
LO	377	0	377
	2.576	0.000	2.576
LT	5	0	5
	.699	0.000	.699
LU	0	8	8
	0.000	.903	.903
LY	0	38	38
	0.000	1.580	1.580
MA	2690	0	2690
	3.430	0.000	3.430
ME	895	0	895
	2.952	0.000	2.952
MI	13	0	13
	1.114	0.000	1.114
MO	16	0	16
	1.204	0.000	1.204
MR	534	0	534
	2.728	0.000	2.728
MU	32	0	32
	1.505	0.000	1.505
MY	0	15	15
	0.000	1.176	1.176
NA	17	0	17
	1.230	0.000	1.230
NC	0	20	20
	0.000	1.301	1.301
ND	0	29265	29265
	0.000	4.466	4.466
NE	1669	3292	4961
	3.222	3.517	3.696
NI	3	0	3
	.477	0.000	.477

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
NK	0	7	7
	.0000	.845	.845
NN	0	38	38
	.0000	1.580	1.580
NO	6161	0	6161
	3.790	0.000	3.790
NT	0	6	6
	.0000	.778	.778
NU	15	0	15
	1.176	0.000	1.176
NY	0	1345	1345
	.0000	3.129	3.129
OA	15	0	15
	1.176	0.000	1.176
OB	0	307	307
	.0000	2.487	2.487
OC	24	23	47
	1.380	1.362	1.672
OD	44	357	401
	1.643	2.553	2.603
OE	0	81	81
	.0000	1.908	1.908
OF	639	0	639
	2.806	0.000	2.806
OG	0	114	114
	.0000	2.057	2.057
OI	93	0	93
	1.968	0.000	1.968
OL	660	10	670
	2.820	1.000	2.826
OM	0	66	66
	.0000	1.820	1.820
ON	3292	306	3598
	3.517	2.486	3.556
OO	0	844	844
	.0000	2.926	2.926
OP	0	230	230
	.0000	2.362	2.362
OR	3	9684	9687
	.477	3.986	3.986
OS	0	52	52
	.0000	1.716	1.716

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION	1ST	2ND	SUM
OT	Ø	5397	5397	
	Ø.000	3.732	3.732	
OU	3348	3299	6647	
	3.525	3.518	3.823	
OV	Ø	40	40	
	Ø.000	1.602	1.602	
OW	782	2405	3187	
	2.893	3.381	3.503	
OX	Ø	101	101	
	Ø.000	2.004	2.004	
OY	Ø	327	327	
	Ø.000	2.515	2.515	
PA	253	Ø	253	
	2.403	Ø.000	2.403	
PE	400	3	403	
	2.602	.477	2.605	
PI	75	Ø	75	
	1.875	Ø.000	1.875	
PO	40	Ø	40	
	1.602	Ø.000	1.602	
PR	22	Ø	22	
	1.342	Ø.000	1.342	
PT	Ø	15	15	
	Ø.000	1.176	1.176	
PU	437	Ø	437	
	2.640	Ø.000	2.640	
PY	Ø	9	9	
	Ø.000	.954	.954	
QU	11	Ø	11	
	1.041	Ø.000	1.041	
RA	212	30	242	
	2.326	1.477	2.384	
RC	Ø	41	41	
	Ø.000	1.613	1.613	
RE	248	4396	4644	
	2.394	3.643	3.667	
RI	39	Ø	39	
	1.591	Ø.000	1.591	
RM	Ø	94	94	
	Ø.000	1.973	1.973	
RO	121	16	137	
	2.083	1.204	2.137	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
RS	0	539	539
	0.000	2.732	2.732
RT	0	208	208
	0.000	2.318	2.318
RU	240	0	240
	2.380	0.000	2.380
RY	4	267	271
	.602	2.427	2.433
SA	1194	0	1194
	3.077	0.000	3.077
SE	1414	589	2003
	3.150	2.770	3.302
SH	2872	11	2883
	3.458	1.041	3.460
SI	439	0	439
	2.642	0.000	2.642
SK	58	128	186
	1.763	2.107	2.270
SL	5	0	5
	.699	0.000	.699
SO	185	0	185
	2.267	0.000	2.267
SP	9	0	9
	.954	0.000	.954
SS	0	5	5
	0.000	.699	.699
ST	0	3	3
	0.000	.477	.477
SU	178	0	178
	2.250	0.000	2.250
TA	241	0	241
	2.382	0.000	2.382
TC	0	58	58
	0.000	1.763	1.763
TD	0	5	5
	0.000	.699	.699
TE	208	16	224
	2.318	1.204	2.350
TH	69983	0	69983
	4.845	0.000	4.845
TI	82	0	82
	1.914	0.000	1.914

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
TO	1125	0	1125
	3.051	0.000	3.051
TR	140	0	140
	2.146	0.000	2.146
TS	0	1858	1858
	0.000	3.269	3.269
TU	16	0	16
	1.204	0.000	1.204
TW	1412	0	1412
	3.150	0.000	3.150
UB	0	19	19
	0.000	1.279	1.279
UD	0	41	41
	0.000	1.613	1.613
UE	0	166	166
	0.000	2.220	2.220
UG	0	65	65
	0.000	1.813	1.813
UH	0	5	5
	0.000	.699	.699
UM	0	74	74
	0.000	1.869	1.869
UN	3	486	489
	.477	2.687	2.689
UO	0	11	11
	0.000	1.041	1.041
UP	0	45	45
	0.000	1.653	1.653
UR	0	1268	1268
	0.000	3.103	3.103
US	589	37	626
	2.770	1.568	2.797
UT	0	7134	7134
	0.000	3.853	3.853
UY	0	121	121
	0.000	2.083	2.083
VA	32	0	32
	1.505	0.000	1.505
VE	0	32	32
	0.000	1.505	1.505
VI	52	0	52
	1.716	0.000	1.716

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 3

	BIGRAM POSITION		
	1ST	2ND	SUM
VO	9 .954	0 0.000	9 .954
VY	0 0.000	9 .954	9 .954
WA	11203 4.049	0 0.000	11203 4.049
WE	64 1.806	15 1.176	79 1.898
WH	2656 3.424	0 0.000	2656 3.424
WI	75 1.875	0 0.000	75 1.875
WN	0 0.000	772 2.888	772 2.888
WO	76 1.881	1412 3.150	1488 3.173
WR	5 .699	0 0.000	5 .699
XE	0 0.000	6 .778	6 .778
YE	569 2.755	126 2.100	695 2.842
YO	3286 3.517	0 0.000	3286 3.517
ZE	26 1.415	0 0.000	26 1.415
ZO	9 .954	0 0.000	9 .954

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
AB	236	70	33	339
	2.373	1.845	1.519	2.530
AC	65	3085	0	3150
	1.813	3.489	0.000	3.498
AD	54	1210	1191	2455
	1.732	3.083	3.076	3.390
AF	0	82	28	110
	0.000	1.914	1.447	2.041
AG	69	178	39	286
	1.839	2.250	1.591	2.456
AH	0	0	31	31
	0.000	0.000	1.491	1.491
AI	62	3156	0	3218
	1.792	3.499	0.000	3.508
AK	10	1569	55	1634
	1.000	3.196	1.740	3.213
AL	1127	1667	593	3387
	3.052	3.222	2.773	3.530
AM	33	1958	231	2222
	1.519	3.292	2.364	3.347
AN	98	2519	2390	5007
	1.991	3.401	3.378	3.700
AO	0	64	0	64
	0.000	1.806	0.000	1.806
AP	4	124	95	223
	.602	2.093	1.978	2.348
AQ	0	0	3	3
	0.000	0.000	.477	.477
AR	660	2197	1427	4284
	2.820	3.342	3.154	3.632
AS	62	2345	52	2459
	1.792	3.370	1.716	3.391
AT	47	988	13013	14048
	1.672	2.995	4.114	4.148
AU	44	51	0	95
	1.643	1.708	0.000	1.978
AV	0	4415	0	4415
	0.000	3.645	0.000	3.645
AW	461	158	74	693
	2.664	2.199	1.869	2.841
AX	45	16	0	61
	1.653	1.204	0.000	1.785

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
AY	0	763	1004	1767
	0.000	2.883	3.002	3.247
AZ	0	151	0	151
	0.000	2.179	0.000	2.179
BA	1579	0	47	1626
	3.198	0.000	1.672	3.211
BE	3186	28	48	3262
	3.503	1.447	1.681	3.513
BI	214	0	0	214
	2.330	0.000	0.000	2.330
BL	217	216	0	433
	2.336	2.334	0.000	2.636
BO	1803	0	0	1803
	3.256	0.000	0.000	3.256
BR	14	0	0	14
	1.146	0.000	0.000	1.146
BS	0	0	83	83
	0.000	0.000	1.919	1.919
BT	0	0	13	13
	0.000	0.000	1.114	1.114
BU	244	0	0	244
	2.387	0.000	0.000	2.387
BY	9	0	62	71
	.954	0.000	1.792	1.851
CA	1837	15	0	1852
	3.264	1.176	0.000	3.268
CE	223	0	1196	1419
	2.348	0.000	3.078	3.152
CH	73	14	3267	3354
	1.863	1.146	3.514	3.526
CI	400	13	0	413
	2.602	1.114	0.000	2.616
CK	0	0	1706	1706
	0.000	0.000	3.232	3.232
CL	279	0	0	279
	2.446	0.000	0.000	2.446
CO	1514	0	21	1535
	3.180	0.000	1.322	3.186
CR	56	9	0	65
	1.748	.954	0.000	1.813
CT	0	39	458	497
	0.000	1.591	2.661	2.696

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
CU	148	0	0	148
	2.170	0.000	0.000	2.170
CY	0	0	49	49
	0.000	0.000	1.690	1.690
DA	1004	44	0	1048
	3.002	1.643	0.000	3.020
DD	0	24	0	24
	0.000	1.380	0.000	1.380
DE	717	204	1851	2772
	2.856	2.310	3.267	3.443
DG	0	78	3	81
	0.000	1.892	.477	1.908
DI	268	0	0	268
	2.428	0.000	0.000	2.428
DL	0	19	0	19
	0.000	1.279	0.000	1.279
DN	0	3	0	3
	0.000	.477	0.000	.477
DO	2146	21	3	2170
	3.332	1.322	.477	3.336
DR	238	0	0	238
	2.377	0.000	0.000	2.377
DS	0	0	204	204
	0.000	0.000	2.310	2.310
DU	234	0	0	234
	2.369	0.000	0.000	2.369
DY	7	0	401	408
	.845	0.000	2.603	2.611
EA	1296	3589	529	5414
	3.113	3.555	2.723	3.734
EB	0	13	0	13
	0.000	1.114	0.000	1.114
EC	10	119	0	129
	1.000	2.076	0.000	2.111
ED	90	16	1414	1520
	1.954	1.204	3.150	3.182
EE	0	5115	371	5486
	0.000	3.709	2.569	3.739
EF	0	487	41	528
	0.000	2.688	1.613	2.723
EG	35	67	0	102
	1.544	1.826	0.000	2.009

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
EI	0	46	0	46
	0.000	1.663	0.000	1.663
EK	0	0	344	344
	0.000	0.000	2.537	2.537
EL	176	2512	276	2964
	2.246	3.400	2.441	3.472
EM	10	0	2149	2159
	1.000	0.000	3.332	3.334
EN	73	1158	8037	9268
	1.863	3.064	3.905	3.967
EO	0	19	0	19
	0.000	1.279	0.000	1.279
EP	18	238	534	790
	1.255	2.377	2.728	2.898
ER	3	5130	1646	6779
	.477	3.710	3.216	3.831
ES	0	1450	1394	2844
	0.000	3.161	3.144	3.454
ET	0	186	567	753
	0.000	2.270	2.754	2.877
EU	0	0	5	5
	0.000	0.000	.699	.699
EV	1588	10	0	1598
	3.201	1.000	0.000	3.204
EW	0	159	803	962
	0.000	2.201	2.905	2.983
EX	7	454	34	495
	.845	2.657	1.531	2.695
EY	408	34	3645	4087
	2.611	1.531	3.562	3.611
FA	1390	0	6	1396
	3.143	0.000	.778	3.145
FE	1233	0	1021	2254
	3.091	0.000	3.009	3.353
FF	0	0	5	5
	0.000	0.000	.699	.699
FI	1480	0	0	1480
	3.170	0.000	0.000	3.170
FL	242	0	0	242
	2.384	0.000	0.000	2.384
FO	1204	0	0	1204
	3.081	0.000	0.000	3.081

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	SUM
FR	4662	0	0	4662
	3.669	0.000	0.000	3.669
FT	0	0	601	601
	0.000	0.000	2.779	2.779
FU	345	0	0	345
	2.538	0.000	0.000	2.538
FY	0	0	7	7
	0.000	0.000	.845	.845
GA	583	0	13	596
	2.766	0.000	1.114	2.775
GE	103	69	304	476
	2.013	1.839	2.483	2.678
GG	0	35	0	35
	0.000	1.544	0.000	1.544
GH	0	0	517	517
	0.000	0.000	2.713	2.713
GI	647	0	0	647
	2.811	0.000	0.000	2.811
GL	61	21	0	82
	1.785	1.322	0.000	1.914
GN	0	0	94	94
	0.000	0.000	1.973	1.973
GO	1296	3	0	1299
	3.113	.477	0.000	3.114
GR	292	0	0	292
	2.465	0.000	0.000	2.465
GS	0	0	241	241
	0.000	0.000	2.382	2.382
GU	109	0	0	109
	2.037	0.000	0.000	2.037
HA	5393	14317	0	19710
	3.732	4.156	0.000	4.295
HE	2247	9152	4	11403
	3.352	3.962	.602	4.057
HI	658	5500	0	6158
	2.818	3.740	0.000	3.789
HN	0	0	362	362
	0.000	0.000	2.559	2.559
HO	1276	648	10	1934
	3.106	2.812	1.000	3.286
HR	0	10	0	10
	0.000	1.000	0.000	1.000

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
	SUM			
HU	201 2.303	361 2.558	0 0.000	562 2.750
HY	9 .954	0 0.000	0 0.000	9 .954
IA	0 0.000	21 1.322	44 1.643	65 1.813
IB	0 0.000	11 1.041	0 0.000	11 1.041
IC	0 0.000	447 2.650	28 1.447	475 2.677
ID	221 2.344	666 2.823	2285 3.359	3172 3.501
IE	0 0.000	433 2.636	0 0.000	433 2.636
IF	0 0.000	999 3.000	0 0.000	999 3.000
IG	3 .477	631 2.800	0 0.000	634 2.802
IK	0 0.000	1426 3.154	0 0.000	1426 3.154
IL	9 .954	3093 3.490	427 2.630	3529 3.548
IM	0 0.000	1649 3.217	73 1.863	1722 3.236
IN	1831 3.263	2217 3.346	698 2.844	4746 3.676
IO	13 1.114	30 1.477	47 1.672	90 1.954
IP	0 0.000	134 2.127	254 2.405	388 2.589
IR	46 1.663	687 2.837	292 2.465	1025 3.011
IS	5 .699	839 2.924	5199 3.716	6043 3.781
IT	59 1.771	7847 3.895	309 2.490	8215 3.915
IV	4 .602	877 2.943	0 0.000	881 2.945
IZ	0 0.000	138 2.140	5 .699	143 2.155
JA	261 2.417	0 0.000	0 0.000	261 2.417

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	SUM
JE	97	0	0	97
	1.987	0.000	0.000	1.987
JI	7	0	0	7
	.845	0.000	0.000	.845
JO	539	0	0	539
	2.732	0.000	0.000	2.732
JU	1147	0	0	1147
	3.060	0.000	0.000	3.060
KA	48	20	0	68
	1.681	1.301	0.000	1.833
KE	528	0	3048	3576
	2.723	0.000	3.484	3.553
KI	529	62	0	591
	2.723	1.792	0.000	2.772
KL	0	3	0	3
	0.000	.477	0.000	.477
KN	1136	0	0	1136
	3.055	0.000	0.000	3.055
KO	11	0	0	11
	1.041	0.000	0.000	1.041
KS	0	0	18	18
	0.000	0.000	1.255	1.255
LA	1654	663	10	2327
	3.219	2.822	1.000	3.367
LB	0	0	7	7
	0.000	0.000	.845	.845
LD	0	3	1264	1267
	0.000	.477	3.102	3.103
LE	1228	118	866	2212
	3.089	2.072	2.938	3.345
LF	0	0	388	388
	0.000	0.000	2.589	2.589
LG	0	6	0	6
	0.000	.778	0.000	.778
LI	2835	79	0	2914
	3.453	1.898	0.000	3.464
LK	0	0	365	365
	0.000	0.000	2.562	2.562
LL	0	18	5379	5397
	0.000	1.255	3.731	3.732
LM	0	0	157	157
	0.000	0.000	2.196	2.196

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
LO	2032 3.308	254 2.405	11 1.041	2297 3.361
LP	0 0.000	0 0.000	316 2.500	316 2.500
LS	0 0.000	1245 3.095	24 1.380	1269 3.103
LT	0 0.000	4 .602	508 2.706	512 2.709
LU	144 2.158	449 2.652	0 0.000	593 2.773
LV	0 0.000	4 .602	0 0.000	4 .602
LY	5 .699	0 0.000	1930 3.286	1935 3.287
MA	3631 3.560	0 0.000	54 1.732	3685 3.566
MB	0 0.000	0 0.000	82 1.914	82 1.914
ME	504 2.702	19 1.279	6252 3.796	6775 3.831
MI	895 2.952	14 1.146	0 0.000	909 2.959
MM	0 0.000	10 1.000	0 0.000	10 1.000
MN	0 0.000	0 0.000	43 1.633	43 1.633
MO	3751 3.574	0 0.000	0 0.000	3751 3.574
MP	0 0.000	0 0.000	178 2.250	178 2.250
MS	0 0.000	0 0.000	161 2.207	161 2.207
MU	1961 3.292	7 .845	0 0.000	1968 3.294
MY	35 1.544	0 0.000	132 2.121	167 2.223
NA	350 2.544	12 1.079	10 1.000	372 2.571
NC	0 0.000	539 2.732	10 1.000	549 2.740
ND	0 0.000	102 2.009	2140 3.330	2242 3.351

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
	SUM			
NE	1196 3.078	552 2.742	1701 3.231	3449 3.538
NG	0 0.000	0 0.000	1245 3.095	1245 3.095
NI	181 2.258	113 2.053	0 0.000	294 2.468
NK	0 0.000	0 0.000	308 2.489	308 2.489
NL	0 0.000	1747 3.242	0 0.000	1747 3.242
NN	0 0.000	49 1.690	25 1.398	74 1.869
NO	330 2.519	755 2.878	7 .845	1092 3.038
NS	0 0.000	0 0.000	304 2.483	304 2.483
NT	0 0.000	1877 3.273	1349 3.130	3226 3.509
NU	62 1.792	0 0.000	5 .699	67 1.826
NV	0 0.000	7 .845	0 0.000	7 .845
NY	0 0.000	0 0.000	1161 3.065	1161 3.065
OA	13 1.114	608 2.784	0 0.000	621 2.793
OB	8 .903	81 1.908	4 .602	93 1.968
OC	0 0.000	123 2.090	10 1.000	133 2.124
OD	28 1.447	395 2.597	1053 3.022	1476 3.169
OE	0 0.000	746 2.873	14 1.146	760 2.881
OF	0 0.000	67 1.826	63 1.799	130 2.114
OG	0 0.000	78 1.892	0 0.000	78 1.892
OH	38 1.580	362 2.559	0 0.000	400 2.602
OI	25 1.398	189 2.276	0 0.000	214 2.330

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
OK	23 1.362	39 1.591	1070 3.029	1132 3.054
OL	6 .778	1363 3.134	267 2.427	1636 3.214
OM	0 0.000	2934 3.467	4952 3.695	7886 3.897
ON	2422 3.384	1854 3.268	867 2.938	5143 3.711
OO	0 0.000	3703 3.569	0 0.000	3703 3.569
OP	322 2.508	330 2.519	302 2.480	954 2.980
OR	27 1.431	4755 3.677	451 2.654	5233 3.719
OS	0 0.000	2065 3.315	64 1.806	2129 3.328
OT	3 .477	993 2.997	373 2.572	1369 3.136
OU	27 1.431	1582 3.199	14 1.146	1623 3.210
OV	1251 3.097	410 2.613	0 0.000	1661 3.220
OW	58 1.763	1209 3.082	1286 3.109	2553 3.407
OX	10 1.000	0 0.000	5 .699	15 1.176
OY	0 0.000	170 2.230	3 .477	173 2.238
PA	1638 3.214	28 1.447	40 1.602	1706 3.232
PE	104 2.017	332 2.521	558 2.747	994 2.997
PH	65 1.813	0 0.000	0 0.000	65 1.813
PI	269 2.430	34 1.531	0 0.000	303 2.481
PL	560 2.748	0 0.000	0 0.000	560 2.748
PO	666 2.823	552 2.742	0 0.000	1218 3.086
PR	30 1.477	0 0.000	0 0.000	30 1.477

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

BIGRAM POSITION		1ST	2ND	3RD	SUM
PS	0	0	164	164	
	0.000	0.000	2.215	2.215	
PT	0	0	240	240	
	0.000	0.000	2.380	2.380	
PU	186	32	0	218	
	2.270	1.505	0.000	2.338	
PY	0	0	38	38	
	0.000	0.000	1.580	1.580	
QU	15	0	0	15	
	1.176	0.000	0.000	1.176	
RA	561	273	19	853	
	2.749	2.436	1.279	2.931	
RB	0	0	27	27	
	0.000	0.000	1.431	1.431	
RC	0	13	0	13	
	0.000	1.114	0.000	1.114	
RD	0	0	747	747	
	0.000	0.000	2.873	2.873	
RE	706	860	7375	8941	
	2.849	2.934	3.868	3.951	
RF	0	0	3	3	
	0.000	0.000	.477	.477	
RG	0	21	0	21	
	0.000	1.322	0.000	1.322	
RI	424	176	6	606	
	2.627	2.246	.778	2.782	
RK	0	0	1475	1475	
	0.000	0.000	3.169	3.169	
RL	0	0	283	283	
	0.000	0.000	2.452	2.452	
RM	0	253	792	1045	
	0.000	2.403	2.899	3.019	
RN	0	0	555	555	
	0.000	0.000	2.744	2.744	
RO	1162	4586	76	5824	
	3.065	3.661	1.881	3.765	
RP	0	0	13	13	
	0.000	0.000	1.114	1.114	
RR	0	0	22	22	
	0.000	0.000	1.342	1.342	
RS	0	0	288	288	
	0.000	0.000	2.459	2.459	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
RT	0	66	845	911
	0.000	1.820	2.927	2.960
RU	222	266	14	502
	2.346	2.425	1.146	2.701
RX	0	0	8	8
	0.000	0.000	.903	.903
RY	0	0	1050	1050
	0.000	0.000	3.021	3.021
SA	3211	0	0	3211
	3.507	0.000	0.000	3.507
SC	15	0	6	21
	1.176	0.000	.778	1.322
SE	1160	675	1057	2892
	3.064	2.829	3.024	3.461
SH	628	0	342	970
	2.798	0.000	2.534	2.987
SI	837	44	0	881
	2.923	1.643	0.000	2.945
SK	52	18	232	302
	1.716	1.255	2.365	2.480
SL	166	5	0	171
	2.220	.699	0.000	2.233
SM	7	0	0	7
	.845	0.000	0.000	.845
SN	71	0	0	71
	1.851	0.000	0.000	1.851
SO	2387	0	1080	3467
	3.378	0.000	3.033	3.540
SP	139	0	3	142
	2.143	0.000	.477	2.152
SS	0	0	1092	1092
	0.000	0.000	3.038	3.038
ST	448	0	6087	6535
	2.651	0.000	3.784	3.815
SU	1665	0	0	1665
	3.221	0.000	0.000	3.221
SW	29	0	0	29
	1.462	0.000	0.000	1.462
SY	0	0	192	192
	0.000	0.000	2.283	2.283
TA	1022	159	173	1354
	3.009	2.201	2.238	3.132

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
TC	0	5	0	5
	.0000	.699	.0000	.699
TE	693	219	990	1902
	2.841	2.340	2.996	3.279
TH	24777	0	8165	32942
	4.394	.0000	3.912	4.518
TI	1818	10	0	1828
	3.260	1.000	.0000	3.262
TO	1384	163	1903	3450
	3.141	2.212	3.279	3.538
TR	453	0	0	453
	2.656	.0000	.0000	2.656
TS	0	0	536	536
	.0000	.0000	2.729	2.729
TT	0	4	12	16
	.0000	.602	1.079	1.204
TU	282	3	0	285
	2.450	.477	.0000	2.455
TW	7	0	0	7
	.845	.0000	.0000	.845
TY	200	0	472	672
	2.301	.0000	2.674	2.827
UA	0	28	0	28
	.0000	1.447	.0000	1.447
UB	0	78	148	226
	.0000	1.892	2.170	2.354
UC	0	2366	0	2366
	.0000	3.374	.0000	3.374
UD	0	46	23	69
	.0000	1.663	1.362	1.839
UE	0	26	389	415
	.0000	1.415	2.590	2.618
UF	0	5	0	5
	.0000	.699	.0000	.699
UG	21	75	64	160
	1.322	1.875	1.806	2.204
UI	0	86	0	86
	.0000	1.934	.0000	1.934
UK	0	14	3	17
	.0000	1.146	.477	1.230
UL	0	561	94	655
	.0000	2.749	1.973	2.816

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			
	1ST	2ND	3RD	
			SUM	
UM	Ø	89	19	108
	Ø.000	1.949	1.279	2.033
UN	122	447	16	585
	2.086	2.650	1.204	2.767
UP	495	14	20	529
	2.695	1.146	1.301	2.723
UR	21	801	1497	2319
	1.322	2.904	3.175	3.365
US	675	2147	387	3209
	2.829	3.332	2.588	3.506
UT	6	212	61	279
	.778	2.326	1.785	2.446
UX	Ø	Ø	30	30
	Ø.000	Ø.000	1.477	1.477
UY	Ø	31	Ø	31
	Ø.000	1.491	Ø.000	1.491
UZ	Ø	16	Ø	16
	Ø.000	1.204	Ø.000	1.204
VA	109	12	4	125
	2.037	1.079	.602	2.097
VD	Ø	Ø	4	4
	Ø.000	Ø.000	.602	.602
VE	857	2759	5661	9277
	2.933	3.441	3.753	3.967
VI	252	72	Ø	324
	2.401	1.857	Ø.000	2.511
VO	90	Ø	Ø	90
	1.954	Ø.000	Ø.000	1.954
VY	Ø	Ø	54	54
	Ø.000	Ø.000	1.732	1.732
WA	1117	470	4	1591
	3.048	2.672	.602	3.202
WD	Ø	Ø	3	3
	Ø.000	Ø.000	.477	.477
WE	5289	50	12	5351
	3.723	1.699	1.079	3.728
WH	4407	Ø	Ø	4407
	3.644	Ø.000	Ø.000	3.644
WI	10328	22	Ø	10350
	4.014	1.342	Ø.000	4.015
WK	Ø	Ø	14	14
	Ø.000	Ø.000	1.146	1.146

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION		
	1ST	2ND	3RD
			SUM
WL	0	0	27
	0.000	0.000	1.431
WN	0	13	1169
	0.000	1.114	3.068
WO	1214	0	0
	3.084	0.000	0.000
WR	12	0	0
	1.079	0.000	0.000
WS	0	0	297
	0.000	0.000	2.473
XE	0	17	0
	0.000	1.230	0.000
XI	0	45	16
	0.000	1.653	1.204
XT	0	0	454
	0.000	0.000	2.657
YA	69	0	0
	1.839	0.000	0.000
YD	0	0	5
	0.000	0.000	.699
YE	694	415	0
	2.841	2.618	0.000
YM	0	9	0
	0.000	.954	0.000
YN	0	5	0
	0.000	.699	0.000
YO	1227	0	0
	3.089	0.000	0.000
YP	0	200	0
	0.000	2.301	0.000
YR	0	9	0
	0.000	.954	0.000
YS	0	0	989
	0.000	0.000	2.995
YT	0	35	4
	0.000	1.544	.602
YU	6	0	0
	.778	0.000	0.000
ZB	37	0	163
	1.568	0.000	2.212
ZI	16	0	13
	1.204	0.000	1.114
			1.462

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 4

	BIGRAM POSITION			SUM
	1ST	2ND	3RD	
ZO	11	0	0	11
	1.041	0.000	0.000	1.041
ZY	0	0	14	14
	0.000	0.000	1.146	1.146
ZZ	0	0	115	115
	0.000	0.000	2.061	2.061

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

BIGRAM POSITION		1ST	2ND	3RD	4TH	SUM
AA	7	0	10	0	17	
	.845	0.000	1.000	0.000	1.230	
AB	2151	438	3	0	2592	
	3.333	2.641	.477	0.000	3.414	
AC	109	309	1625	14	2057	
	2.037	2.490	3.211	1.146	3.313	
AD	308	237	541	285	1371	
	2.489	2.375	2.733	2.455	3.137	
AF	1070	5	171	3	1249	
	3.029	.699	2.233	.477	3.097	
AG	686	241	315	0	1242	
	2.836	2.382	2.498	0.000	3.094	
AH	109	4	0	26	139	
	2.037	.602	0.000	1.415	2.143	
AI	45	581	1088	7	1721	
	1.653	2.764	3.037	.845	3.236	
AJ	0	247	0	0	247	
	0.000	2.393	0.000	0.000	2.393	
AK	0	644	120	225	989	
	0.000	2.809	2.079	2.352	2.995	
AL	875	810	1066	1616	4367	
	2.942	2.908	3.028	3.208	3.640	
AM	396	362	305	111	1174	
	2.598	2.559	2.484	2.045	3.070	
AN	176	994	1312	1228	3710	
	2.246	2.997	3.118	3.089	3.569	
AO	0	0	17	0	17	
	0.000	0.000	1.230	0.000	1.230	
AP	204	384	122	32	742	
	2.310	2.584	2.086	1.505	2.870	
AR	410	1934	2670	394	5408	
	2.613	3.286	3.427	2.595	3.733	
AS	489	884	1041	468	2882	
	2.689	2.946	3.017	2.670	3.460	
AT	69	1449	1260	736	3514	
	1.839	3.161	3.100	2.867	3.546	
AU	8	274	8	0	290	
	.903	2.438	.903	0.000	2.462	
AV	62	332	416	0	810	
	1.792	2.521	2.619	0.000	2.908	
AW	185	12	98	15	310	
	2.267	1.079	1.991	1.176	2.491	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
AX	0	65	0	19	84
	0.000	1.813	0.000	1.279	1.924
AY	0	211	74	384	669
	0.000	2.324	1.869	2.584	2.825
AZ	0	38	44	0	82
	0.000	1.580	1.643	0.000	1.914
BA	745	0	61	0	806
	2.872	0.000	1.785	0.000	2.906
BB	0	7	60	0	67
	0.000	.845	1.778	0.000	1.826
BE	1556	0	144	161	1861
	3.192	0.000	2.158	2.207	3.270
BI	269	7	77	21	374
	2.430	.845	1.886	1.322	2.573
BL	654	0	287	0	941
	2.816	0.000	2.458	0.000	2.974
BO	668	2115	162	4	2949
	2.825	3.325	2.210	.602	3.470
BR	984	0	3	0	987
	2.993	0.000	.477	0.000	2.994
BS	0	0	0	80	80
	0.000	0.000	0.000	1.903	1.903
BT	0	0	12	114	126
	0.000	0.000	1.079	2.057	2.100
BU	372	18	20	0	410
	2.571	1.255	1.301	0.000	2.613
BY	16	4	0	58	78
	1.204	.602	0.000	1.763	1.892
CA	731	86	337	0	1154
	2.864	1.934	2.528	0.000	3.062
CC	0	43	0	0	43
	0.000	1.633	0.000	0.000	1.633
CE	121	146	187	2712	3166
	2.083	2.164	2.272	3.433	3.501
CH	925	3	3	4479	5410
	2.966	.477	.477	3.651	3.733
CI	161	7	12	0	180
	2.207	.845	1.079	0.000	2.255
CK	0	0	112	1000	1112
	0.000	0.000	2.049	3.000	3.046
CL	1097	0	81	0	1178
	3.040	0.000	1.908	0.000	3.071

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
CO	2783 3.445	134 2.127	18 1.255	0 0.000	2935 3.468
CR	457 2.660	82 1.914	0 0.000	0 0.000	539 2.732
CS	0 0.000	0 0.000	0 0.000	7 .845	7 .845
CT	0 0.000	42 1.623	100 2.000	87 1.940	229 2.360
CU	113 2.053	13 1.114	88 1.944	0 0.000	214 2.330
CY	28 1.447	0 0.000	0 0.000	47 1.672	75 1.875
DA	412 2.615	50 1.699	321 2.507	5 .699	788 2.897
DD	0 0.000	211 2.324	34 1.531	0 0.000	245 2.389
DE	522 2.718	208 2.318	1814 3.259	558 2.747	3102 3.492
DG	0 0.000	44 1.643	143 2.155	0 0.000	187 2.272
DI	112 2.049	11 1.041	225 2.352	0 0.000	348 2.542
DL	0 0.000	7 .845	59 1.771	0 0.000	66 1.820
DM	0 0.000	37 1.568	0 0.000	0 0.000	37 1.568
DO	450 2.653	21 1.322	30 1.477	11 1.041	512 2.709
DR	775 2.889	0 0.000	9 .954	0 0.000	784 2.894
DS	0 0.000	0 0.000	19 1.279	1736 3.240	1755 3.244
DT	0 0.000	0 0.000	14 1.146	0 0.000	14 1.146
DU	52 1.716	25 1.398	13 1.114	3 .477	93 1.968
DW	11 1.041	11 1.041	0 0.000	0 0.000	22 1.342
DY	45 1.653	0 0.000	0 0.000	547 2.738	592 2.772
EA	573 2.758	3865 3.587	2082 3.318	12 1.079	6532 3.815

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				SUM
	1ST	2ND	3RD	4TH	
EB	0	44	4	0	48
	0.000	1.643	.602	0.000	1.681
EC	0	18	261	0	279
	0.000	1.255	2.417	0.000	2.446
ED	89	41	237	2825	3192
	1.949	1.613	2.375	3.451	3.504
EE	0	883	930	665	2478
	0.000	2.946	2.968	2.823	3.394
EF	0	27	10	210	247
	0.000	1.431	1.000	2.322	2.393
EG	14	628	18	0	660
	1.146	2.798	1.255	0.000	2.820
EH	0	5	0	0	5
	0.000	.699	0.000	0.000	.699
EI	104	763	2688	0	3555
	2.017	2.883	3.429	0.000	3.551
EK	0	0	151	95	246
	0.000	0.000	2.179	1.978	2.391
EL	77	474	490	752	1793
	1.886	2.676	2.690	2.876	3.254
EM	64	31	589	19	703
	1.806	1.491	2.770	1.279	2.847
EN	303	685	486	1686	3160
	2.481	2.836	2.687	3.227	3.500
EO	0	0	0	3	3
	0.000	0.000	0.000	.477	.477
EP	6	123	216	126	471
	.778	2.090	2.334	2.100	2.673
EQ	90	0	0	0	90
	1.954	0.000	0.000	0.000	1.954
ER	79	494	4347	7346	12266
	1.898	2.694	3.638	3.866	4.089
ES	31	166	2095	3937	6229
	1.491	2.220	3.321	3.595	3.794
ET	11	128	178	297	614
	1.041	2.107	2.250	2.473	2.788
EU	0	0	9	0	9
	0.000	0.000	.954	0.000	.954
EV	597	1088	3	0	1688
	2.776	3.037	.477	0.000	3.227
EW	0	151	51	71	273
	0.000	2.179	1.708	1.851	2.436

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
EX	151 2.179	89 1.949	0 0.000	93 1.968	333 2.522
EY	0 0.000	6 .778	0 0.000	342 2.534	348 2.542
FA	628 2.798	0 0.000	0 0.000	0 0.000	628 2.798
FE	245 2.389	0 0.000	112 2.049	84 1.924	441 2.644
FF	0 0.000	80 1.903	0 0.000	189 2.276	269 2.430
FI	2400 3.380	0 0.000	0 0.000	0 0.000	2400 3.380
FL	395 2.597	0 0.000	63 1.799	0 0.000	458 2.661
FO	1171 3.069	0 0.000	0 0.000	0 0.000	1171 3.069
FR	541 2.733	0 0.000	0 0.000	0 0.000	541 2.733
FS	0 0.000	0 0.000	0 0.000	12 1.079	12 1.079
FT	0 0.000	1438 3.158	122 2.086	162 2.210	1722 3.236
FU	234 2.369	0 0.000	17 1.230	0 0.000	251 2.400
GA	160 2.204	578 2.762	456 2.659	0 0.000	1194 3.077
GE	3 .477	44 1.643	290 2.462	1040 3.017	1377 3.139
GG	0 0.000	0 0.000	18 1.255	3 .477	21 1.322
GH	15 1.176	0 0.000	2413 3.383	138 2.140	2566 3.409
GI	665 2.823	4 .602	162 2.210	0 0.000	831 2.920
GL	201 2.303	0 0.000	56 1.748	0 0.000	257 2.410
GM	0 0.000	0 0.000	4 .602	0 0.000	4 .602
GN	0 0.000	0 0.000	68 1.833	7 .845	75 1.875
GO	521 2.717	9 .954	69 1.839	72 1.857	671 2.827

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
GR	1763 3.246	51 1.708	149 2.173	0 0.000	1963 3.293
GS	0 0.000	0 0.000	0 0.000	192 2.283	192 2.283
GU	233 2.367	0 0.000	111 2.045	0 0.000	344 2.537
GY	4 .602	14 1.146	0 0.000	23 1.362	41 1.613
HA	611 2.786	924 2.966	0 0.000	0 0.000	1535 3.186
HE	823 2.915	8507 3.930	1717 3.235	11 1.041	11058 4.044
HI	105 2.021	6209 3.793	4 .602	5 .699	6323 3.801
HN	0 0.000	0 0.000	6 .778	0 0.000	6 .778
HO	1404 3.147	2421 3.384	8 .903	4 .602	3837 3.584
HR	0 0.000	713 2.853	5 .699	0 0.000	718 2.856
HS	0 0.000	0 0.000	0 0.000	28 1.447	28 1.447
HT	0 0.000	0 0.000	0 0.000	2413 3.383	2413 3.383
HU	399 2.601	27 1.431	0 0.000	0 0.000	426 2.629
HY	6 .778	7 .845	0 0.000	4 .602	17 1.230
IA	0 0.000	87 1.940	144 2.158	123 2.090	354 2.549
IB	0 0.000	93 1.968	12 1.079	0 0.000	105 2.021
IC	0 0.000	19 1.279	4245 3.628	565 2.752	4829 3.684
ID	211 2.324	257 2.410	229 2.360	333 2.522	1030 3.013
IE	0 0.000	534 2.728	594 2.774	112 2.049	1240 3.093
IF	0 0.000	180 2.255	199 2.299	27 1.431	406 2.609
IG	0 0.000	2468 3.392	26 1.415	0 0.000	2494 3.397

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
II	0	0	0	4	4
	0.000	0.000	0.000	.602	.602
IK	0	82	20	0	102
	0.000	1.914	1.301	0.000	2.009
IL	0	535	2279	722	3536
	0.000	2.728	3.358	2.859	3.549
IM	131	438	91	98	758
	2.117	2.641	1.959	1.991	2.880
IN	228	1401	3331	1090	6050
	2.358	3.146	3.523	3.037	3.782
IO	8	24	255	156	443
	.903	1.380	2.407	2.193	2.646
IP	0	26	92	34	152
	0.000	1.415	1.964	1.531	2.182
IR	47	1914	325	2752	5038
	1.672	3.282	2.512	3.440	3.702
IS	169	226	305	447	1147
	2.228	2.354	2.484	2.650	3.060
IT	106	264	1217	339	1926
	2.025	2.422	3.085	2.530	3.285
IV	17	1053	174	0	1244
	1.230	3.022	2.241	0.000	3.095
IX	0	196	0	30	226
	0.000	2.292	0.000	1.477	2.354
IZ	0	24	34	0	58
	0.000	1.380	1.531	0.000	1.763
JA	142	0	0	0	142
	2.152	0.000	0.000	0.000	2.152
JE	86	0	0	0	86
	1.934	0.000	0.000	0.000	1.934
JI	11	0	0	0	11
	1.041	0.000	0.000	0.000	1.041
JO	155	0	291	0	446
	2.190	0.000	2.464	0.000	2.649
JU	131	0	0	0	131
	2.117	0.000	0.000	0.000	2.117
KA	27	0	0	0	27
	1.431	0.000	0.000	0.000	1.431
KE	21	0	1151	358	1530
	1.322	0.000	3.061	2.554	3.185
KI	68	82	5	0	155
	1.833	1.914	.699	0.000	2.190

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
KL	0	0	8	0	8
	.0000	.0000	.903	.0000	.903
KN	490	0	0	0	490
	2.690	.0000	0.000	0.000	2.690
KO	12	0	0	0	12
	1.079	.0000	0.000	0.000	1.079
KS	0	0	0	812	812
	.0000	.0000	0.000	2.910	2.910
KU	0	3	0	0	3
	.0000	.477	0.000	0.000	.477
KY	0	0	18	72	90
	.0000	.0000	1.255	1.857	1.954
LA	1119	1975	113	28	3235
	3.049	3.296	2.053	1.447	3.510
LB	0	16	3	0	19
	.0000	1.204	.477	0.000	1.279
LC	0	5	0	0	5
	.0000	.699	0.000	0.000	.699
LD	0	108	49	5715	5872
	.0000	2.033	1.690	3.757	3.769
LE	1063	574	731	1837	4205
	3.027	2.759	2.864	3.264	3.624
LF	0	0	8	12	20
	.0000	.0000	.903	1.079	1.301
LG	0	0	5	0	5
	.0000	.0000	.699	0.000	.699
LH	0	0	5	0	5
	.0000	.0000	.699	0.000	.699
LI	958	288	232	0	1478
	2.981	2.459	2.365	0.000	3.170
LK	0	0	65	3	68
	.0000	.0000	1.813	.477	1.833
LL	10	121	789	1806	2726
	1.000	2.083	2.897	3.257	3.436
LM	0	6	39	19	64
	.0000	.778	1.591	1.279	1.806
LO	878	1461	405	10	2754
	2.943	3.165	2.607	1.000	3.440
LP	0	0	53	4	57
	.0000	.0000	1.724	.602	1.756
LS	0	4	38	1006	1048
	.0000	.602	1.580	3.003	3.020

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
LT	0	24	40	213	277
	0.000	1.380	1.602	2.328	2.442
LU	111	191	200	0	502
	2.045	2.281	2.301	0.000	2.701
LV	0	6	26	0	32
	0.000	.778	1.415	0.000	1.505
LY	48	4	0	1055	1107
	1.681	.602	0.000	3.023	3.044
MA	978	693	581	67	2319
	2.990	2.841	2.764	1.826	3.365
MB	0	0	44	27	71
	0.000	0.000	1.643	1.431	1.851
ME	650	34	1162	283	2129
	2.813	1.531	3.065	2.452	3.328
MI	1109	115	105	22	1351
	3.045	2.061	2.021	1.342	3.131
ML	0	0	12	0	12
	0.000	0.000	1.079	0.000	1.079
MM	0	0	46	0	46
	0.000	0.000	1.663	0.000	1.663
MN	0	0	6	0	6
	0.000	0.000	.778	0.000	.778
MO	1236	416	107	0	1759
	3.092	2.619	2.029	0.000	3.245
MP	0	96	36	27	159
	0.000	1.982	1.556	1.431	2.201
MS	0	0	0	1098	1098
	0.000	0.000	0.000	3.041	3.041
MU	241	3	4	0	248
	2.382	.477	.602	0.000	2.394
MY	6	0	0	117	123
	.778	0.000	0.000	2.068	2.090
NA	281	61	178	101	621
	2.449	1.785	2.250	2.004	2.793
NC	0	62	966	0	1028
	0.000	1.792	2.985	0.000	3.012
ND	0	922	751	1331	3004
	0.000	2.965	2.876	3.124	3.478
NE	1025	144	861	607	2637
	3.011	2.158	2.935	2.783	3.421
NG	0	162	370	3346	3878
	0.000	2.210	2.568	3.525	3.589

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
NI	467	441	53	0	961
	2.669	2.644	1.724	0.000	2.983
NJ	0	44	0	0	44
	0.000	1.643	0.000	0.000	1.643
NK	0	8	113	711	832
	0.000	.903	2.053	2.852	2.920
NL	0	4	0	0	4
	0.000	.602	0.000	0.000	.602
NN	0	58	113	6	177
	0.000	1.763	2.053	.778	2.248
NO	546	370	152	38	1106
	2.737	2.568	2.182	1.580	3.044
NP	0	20	0	0	20
	0.000	1.301	0.000	0.000	1.301
NR	0	0	99	0	99
	0.000	0.000	1.996	0.000	1.996
NS	0	38	340	819	1197
	0.000	1.580	2.531	2.913	3.078
NT	0	565	313	1403	2281
	0.000	2.752	2.496	3.147	3.358
NU	17	0	22	0	39
	1.230	0.000	1.342	0.000	1.591
NX	0	0	0	9	9
	0.000	0.000	0.000	.954	.954
NY	0	0	0	142	142
	0.000	0.000	0.000	2.152	2.152
OA	3	573	105	0	681
	.477	2.758	2.021	0.000	2.833
OB	0	108	23	0	131
	0.000	2.033	1.362	0.000	2.117
OC	77	400	295	3	775
	1.886	2.602	2.470	.477	2.889
OD	17	419	259	361	1056
	1.230	2.622	2.413	2.558	3.024
OE	0	114	44	11	169
	0.000	2.057	1.643	1.041	2.228
OF	448	5	15	45	513
	2.651	.699	1.176	1.653	2.710
OG	0	61	0	0	61
	0.000	1.785	0.000	0.000	1.785
OH	0	9	0	0	9
	0.000	.954	0.000	0.000	.954

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

		BIGRAM POSITION				
		1ST	2ND	3RD	4TH	SUM
OI	3	1306	69	0	0	1378
	.477	3.116	1.839	0.000	0	3.139
OK	0	50	411	63	524	
	0.000	1.699	2.614	1.799	2.719	
OL	98	489	373	8	968	
	1.991	2.689	2.572	.903	2.986	
OM	0	775	99	46	920	
	0.000	2.889	1.996	1.663	2.964	
ON	53	820	1475	396	2744	
	1.724	2.914	3.169	2.598	3.438	
OO	0	663	724	3	1390	
	0.000	2.822	2.860	.477	3.143	
OP	63	109	129	25	326	
	1.799	2.037	2.111	1.398	2.513	
OR	404	2726	759	1243	5132	
	2.606	3.436	2.880	3.094	3.710	
OS	9	279	1619	28	1935	
	.954	2.446	3.209	1.447	3.287	
OT	1702	663	429	88	2882	
	3.231	2.822	2.632	1.944	3.460	
OU	105	7649	2342	0	10096	
	2.021	3.884	3.370	0.000	4.004	
OV	11	500	449	9	969	
	1.041	2.699	2.652	.954	2.986	
OW	71	581	972	309	1933	
	1.851	2.764	2.988	2.490	3.286	
OX	0	14	7	0	21	
	0.000	1.146	.845	0.000	1.322	
OY	0	94	10	44	148	
	0.000	1.973	1.000	1.643	2.170	
OZ	0	57	4	0	61	
	0.000	1.756	.602	0.000	1.785	
PA	838	301	52	5	1196	
	2.923	2.479	1.716	.699	3.078	
PE	318	446	501	134	1399	
	2.502	2.649	2.700	2.127	3.146	
PH	143	0	0	39	182	
	2.155	0.000	0.000	1.591	2.260	
PI	297	71	72	0	440	
	2.473	1.851	1.857	0.000	2.643	
PL	1117	30	146	0	1293	
	3.048	1.477	2.164	0.000	3.112	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
PO	1028	151	25	4	1208
	3.012	2.179	1.398	.602	3.082
PP	0	137	103	0	240
	0.000	2.137	2.013	0.000	2.380
PR	615	98	0	0	713
	2.789	1.991	0.000	0.000	2.853
PS	0	14	10	491	515
	0.000	1.146	1.000	2.691	2.712
PT	0	4	117	108	229
	0.000	.602	2.068	2.033	2.360
PU	62	3	20	0	85
	1.792	.477	1.301	0.000	1.929
PY	0	0	0	98	98
	0.000	0.000	0.000	1.991	1.991
QU	515	115	0	0	630
	2.712	2.061	0.000	0.000	2.799
RA	725	1278	292	123	2418
	2.860	3.107	2.465	2.090	3.383
RB	0	58	17	0	75
	0.000	1.763	1.230	0.000	1.875
RC	0	0	419	0	419
	0.000	0.000	2.622	0.000	2.622
RD	0	376	482	824	1682
	0.000	2.575	2.683	2.916	3.226
RE	486	1906	1058	4157	7607
	2.687	3.280	3.024	3.619	3.881
RF	0	0	0	7	7
	0.000	0.000	0.000	.845	.845
RG	0	84	404	0	488
	0.000	1.924	2.606	0.000	2.688
RH	108	0	0	0	108
	2.033	0.000	0.000	0.000	2.033
RI	969	1508	255	16	2748
	2.986	3.178	2.407	1.204	3.439
RK	0	0	199	91	290
	0.000	0.000	2.299	1.959	2.462
RL	0	0	1319	12	1331
	0.000	0.000	3.120	1.079	3.124
RM	0	64	408	71	543
	0.000	1.806	2.611	1.851	2.735
RN	0	23	90	116	229
	0.000	1.362	1.954	2.064	2.360

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
RO	548 2.739	1926 3.285	131 2.117	115 2.061	2720 3.435
RP	0 0.000	0 0.000	109 2.037	72 1.857	181 2.258
RR	0 0.000	64 1.806	362 2.559	0 0.000	426 2.629
RS	0 0.000	0 0.000	1718 3.235	1416 3.151	3134 3.496
RT	0 0.000	0 0.000	1028 3.012	997 2.999	2025 3.306
RU	200 2.301	512 2.709	61 1.785	5 .699	778 2.891
RV	0 0.000	0 0.000	171 2.233	0 0.000	171 2.233
RY	0 0.000	19 1.279	0 0.000	1254 3.098	1273 3.105
SA	375 2.574	24 1.380	55 1.740	6 .778	460 2.663
SC	370 2.568	9 .954	7 .845	0 0.000	386 2.587
SE	943 2.975	6 .778	519 2.715	4720 3.674	6188 3.792
SH	1655 3.219	6 .778	0 0.000	289 2.461	1950 3.290
SI	1025 3.011	222 2.346	700 2.845	0 0.000	1947 3.289
SK	85 1.929	398 2.600	51 1.708	7 .845	541 2.733
SL	254 2.405	7 .845	6 .778	0 0.000	267 2.427
SM	766 2.884	0 0.000	0 0.000	3 .477	769 2.886
SN	61 1.785	0 0.000	0 0.000	0 0.000	61 1.785
SO	750 2.875	0 0.000	31 1.491	7 .845	788 2.897
SP	896 2.952	0 0.000	5 .699	25 1.398	926 2.967
SQ	25 1.398	0 0.000	0 0.000	0 0.000	25 1.398
SS	0 0.000	191 2.281	17 1.230	787 2.896	995 2.998

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
ST	3855	0	499	2236	6590
	3.586	0.000	2.698	3.349	3.819
SU	156	96	216	0	468
	2.193	1.982	2.334	0.000	2.670
SW	306	0	0	0	306
	2.486	0.000	0.000	0.000	2.486
SY	4	0	0	28	32
	.602	0.000	0.000	1.447	1.505
TA	809	1551	362	50	2772
	2.908	3.191	2.559	1.699	3.443
TC	0	0	268	0	268
	0.000	0.000	2.428	0.000	2.428
TD	0	0	3	0	3
	0.000	0.000	.477	0.000	.477
TE	624	372	3129	2104	6229
	2.795	2.571	3.495	3.323	3.794
TH	9669	1713	44	2068	13494
	3.985	3.234	1.643	3.316	4.130
TI	517	856	591	0	1964
	2.713	2.932	2.772	0.000	3.293
TL	0	12	81	0	93
	0.000	1.079	1.908	0.000	1.968
TO	870	764	98	36	1768
	2.940	2.883	1.991	1.556	3.247
TR	1212	60	85	0	1357
	3.084	1.778	1.929	0.000	3.133
TS	0	0	0	1173	1173
	0.000	0.000	0.000	3.069	3.069
TT	0	29	58	29	116
	0.000	1.462	1.763	1.462	2.064
TU	109	301	8	0	418
	2.037	2.479	.903	0.000	2.621
TW	109	0	0	0	109
	2.037	0.000	0.000	0.000	2.037
TY	124	98	0	591	813
	2.093	1.991	0.000	2.772	2.910
TZ	0	0	0	3	3
	0.000	0.000	0.000	.477	.477
UA	0	51	211	0	262
	0.000	1.708	2.324	0.000	2.418
UB	0	62	138	9	209
	0.000	1.792	2.140	.954	2.320

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
UC	0	49	225	0	274
	0.000	1.690	2.352	0.000	2.438
UD	0	110	268	108	486
	0.000	2.041	2.428	2.033	2.687
UE	0	161	66	425	652
	0.000	2.207	1.820	2.628	2.814
UF	0	0	44	0	44
	0.000	0.000	1.643	0.000	1.643
UG	0	108	176	0	284
	0.000	2.033	2.246	0.000	2.453
UI	0	773	129	0	902
	0.000	2.888	2.111	0.000	2.955
UL	5	301	4442	17	4765
	.699	2.479	3.648	1.230	3.678
UM	0	384	57	37	478
	0.000	2.584	1.756	1.568	2.679
UN	1588	283	1541	51	3463
	3.201	2.452	3.188	1.708	3.539
UO	0	21	0	0	21
	0.000	1.322	0.000	0.000	1.322
UP	86	28	0	402	516
	1.934	1.447	0.000	2.604	2.713
UR	85	351	463	56	955
	1.929	2.545	2.666	1.748	2.980
US	261	304	902	150	1617
	2.417	2.483	2.955	2.176	3.209
UT	13	65	616	1876	2570
	1.114	1.813	2.790	3.273	3.410
UX	0	0	0	8	8
	0.000	0.000	0.000	.903	.903
UZ	0	7	0	0	7
	0.000	.845	0.000	0.000	.845
VA	262	14	51	0	327
	2.418	1.146	1.708	0.000	2.515
VE	70	583	2570	1125	4348
	1.845	2.766	3.410	3.051	3.638
VI	294	9	277	0	580
	2.468	.954	2.442	0.000	2.763
VO	309	81	81	0	471
	2.490	1.908	1.908	0.000	2.673
VY	3	0	0	114	117
	.477	0.000	0.000	2.057	2.068

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
WA	961	165	0	0	1126
	2.983	2.217	0.000	0.000	3.052
WD	0	0	7	53	60
	0.000	0.000	.845	1.724	1.778
WE	201	172	569	0	942
	2.303	2.236	2.755	0.000	2.974
WF	0	17	0	0	17
	0.000	1.230	0.000	0.000	1.230
WH	6184	0	0	0	6184
	3.791	0.000	0.000	0.000	3.791
WI	207	178	68	0	453
	2.316	2.250	1.833	0.000	2.656
WL	0	0	38	15	53
	0.000	0.000	1.580	1.176	1.724
WN	0	67	73	738	878
	0.000	1.826	1.863	2.868	2.943
WO	4632	35	0	0	4667
	3.666	1.544	0.000	0.000	3.669
WR	446	0	0	0	446
	2.649	0.000	0.000	0.000	2.649
WS	0	0	0	311	311
	0.000	0.000	0.000	2.493	2.493
WU	0	48	0	0	48
	0.000	1.681	0.000	0.000	1.681
WY	3	0	0	4	7
	.477	0.000	0.000	.602	.845
XA	0	27	74	0	101
	0.000	1.431	1.869	0.000	2.004
XE	0	11	211	0	222
	0.000	1.041	2.324	0.000	2.346
XI	0	63	3	0	66
	0.000	1.799	.477	0.000	1.820
XO	0	0	25	0	25
	0.000	0.000	1.398	0.000	1.398
XT	0	50	51	0	101
	0.000	1.699	1.708	0.000	2.004
XY	0	0	0	7	7
	0.000	0.000	0.000	.845	.845
YA	68	3	78	0	149
	1.833	.477	1.892	0.000	2.173
YB	0	0	134	0	134
	0.000	0.000	2.127	0.000	2.127

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 5

	BIGRAM POSITION				
	1ST	2ND	3RD	4TH	SUM
YC	0	24	21	0	45
	0.000	1.380	1.322	0.000	1.653
YD	0	0	0	10	10
	0.000	0.000	0.000	1.000	1.000
YE	952	0	35	0	987
	2.979	0.000	1.544	0.000	2.994
YI	35	78	5	0	118
	1.544	1.892	.699	0.000	2.072
YL	0	11	113	0	124
	0.000	1.041	2.053	0.000	2.093
YM	0	6	3	0	9
	0.000	.778	.477	0.000	.954
YN	0	0	12	0	12
	0.000	0.000	1.079	0.000	1.079
YO	492	0	38	18	548
	2.692	0.000	1.580	1.255	2.739
YP	0	123	14	0	137
	0.000	2.090	1.146	0.000	2.137
YR	0	36	0	0	36
	0.000	1.556	0.000	0.000	1.556
YS	0	0	4	74	78
	0.000	0.000	.602	1.869	1.892
YT	0	6	0	0	6
	0.000	.778	0.000	0.000	.778
ZA	0	0	0	3	3
	0.000	0.000	0.000	.477	.477
ZE	0	0	84	48	132
	0.000	0.000	1.924	1.681	2.121
ZI	0	0	12	0	12
	0.000	0.000	1.079	0.000	1.079
ZO	3	0	15	0	18
	.477	0.000	1.176	0.000	1.255
ZY	0	0	0	46	46
	0.000	0.000	0.000	1.663	1.663
ZZ	0	0	15	0	15
	0.000	0.000	1.176	0.000	1.176

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
AA	0	0	0	7	0	7
	0.000	0.000	0.000	.845	0.000	.845
AB	170	122	140	0	0	432
	2.230	2.086	2.146	0.000	0.000	2.635
AC	967	406	332	277	4	1986
	2.985	2.609	2.521	2.442	.602	3.298
AD	141	131	363	167	185	987
	2.149	2.117	2.560	2.223	2.267	2.994
AE	8	6	0	15	0	29
	.903	.778	0.000	1.176	0.000	1.462
AF	232	64	11	0	0	307
	2.365	1.806	1.041	0.000	0.000	2.487
AG	192	102	221	156	0	671
	2.283	2.009	2.344	2.193	0.000	2.827
AH	0	0	22	0	0	22
	0.000	0.000	1.342	0.000	0.000	1.342
AI	12	494	324	462	0	1292
	1.079	2.694	2.511	2.665	0.000	3.111
AJ	0	3	0	0	0	3
	0.000	.477	0.000	0.000	0.000	.477
AK	0	470	103	46	10	629
	0.000	2.672	2.013	1.663	1.000	2.799
AL	1094	1430	472	270	1558	4824
	3.039	3.155	2.674	2.431	3.193	3.683
AM	202	743	85	316	83	1429
	2.305	2.871	1.929	2.500	1.919	3.155
AN	604	1377	895	491	310	3677
	2.781	3.139	2.952	2.691	2.491	3.565
AP	185	160	148	71	4	568
	2.267	2.204	2.170	1.851	.602	2.754
AQ	0	0	6	0	0	6
	0.000	0.000	.778	0.000	0.000	.778
AR	906	1830	865	933	278	4812
	2.957	3.262	2.937	2.970	2.444	3.682
AS	344	715	437	151	220	1867
	2.537	2.854	2.640	2.179	2.342	3.271
AT	253	1950	967	370	193	3733
	2.403	3.290	2.985	2.568	2.286	3.572
AU	139	368	96	0	57	660
	2.143	2.566	1.982	0.000	1.756	2.820
AV	49	388	257	13	0	707
	1.690	2.589	2.410	1.114	0.000	2.849

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
AW	31 1.491	62 1.792	11 1.041	3 .477	10 1.000	117 2.068
AX	0 0.000	9 .954	3 .477	3 .477	20 1.301	35 1.544
AY	0 0.000	200 2.301	282 2.450	472 2.674	325 2.512	1279 3.107
AZ	0 0.000	27 1.431	41 1.613	3 .477	0 0.000	71 1.851
BA	376 2.575	0 0.000	53 1.724	58 1.763	0 0.000	487 2.688
BB	0 0.000	0 0.000	85 1.929	8 .903	0 0.000	93 1.968
BE	3022 3.480	7 .845	155 2.190	765 2.884	3 .477	3952 3.597
BI	121 2.083	0 0.000	106 2.025	21 1.322	0 0.000	248 2.394
BJ	0 0.000	68 1.833	0 0.000	0 0.000	0 0.000	68 1.833
BL	101 2.004	3 .477	490 2.690	266 2.425	0 0.000	860 2.934
BM	0 0.000	0 0.000	18 1.255	0 0.000	0 0.000	18 1.255
BO	535 2.728	25 1.398	92 1.964	152 2.182	0 0.000	804 2.905
BR	533 2.727	69 1.839	33 1.519	0 0.000	0 0.000	635 2.803
BS	0 0.000	58 1.763	8 .903	0 0.000	16 1.204	82 1.914
BT	0 0.000	42 1.623	31 1.491	16 1.204	0 0.000	89 1.949
BU	425 2.628	12 1.079	24 1.380	0 0.000	0 0.000	461 2.664
BW	0 0.000	0 0.000	7 .845	0 0.000	0 0.000	7 .845
BY	3 .477	0 0.000	0 0.000	0 0.000	60 1.778	63 1.799
CA	1468 3.167	36 1.556	436 2.639	116 2.064	45 1.653	2101 3.322
CC	0 0.000	167 2.223	0 0.000	4 .602	0 0.000	171 2.233
CE	320 2.505	44 1.643	567 2.754	929 2.968	1267 3.103	3127 3.495

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
CH	1371 3.137	546 2.737	174 2.241	13 1.114	779 2.892	2883 3.460
CI	177 2.248	0 0.000	538 2.731	46 1.663	0 0.000	761 2.881
CK	0 0.000	0 0.000	468 2.670	186 2.270	167 2.223	821 2.914
CL	438 2.641	0 0.000	23 1.362	102 2.009	0 0.000	563 2.751
CN	0 0.000	0 0.000	15 1.176	0 0.000	0 0.000	15 1.176
CO	2059 3.314	35 1.544	998 2.999	60 1.778	27 1.431	3179 3.502
CR	311 2.493	370 2.568	119 2.076	0 0.000	0 0.000	800 2.903
CS	0 0.000	0 0.000	0 0.000	0 0.000	51 1.708	51 1.708
CT	0 0.000	555 2.744	297 2.473	5 .699	761 2.881	1618 3.209
CU	83 1.919	0 0.000	130 2.114	33 1.519	0 0.000	246 2.391
CY	10 1.000	0 0.000	0 0.000	0 0.000	287 2.458	297 2.473
DA	329 2.517	0 0.000	19 1.279	254 2.405	45 1.653	647 2.811
DD	0 0.000	22 1.342	310 2.491	0 0.000	0 0.000	332 2.521
DE	1235 3.092	16 1.204	604 2.781	1285 3.109	460 2.663	3600 3.556
DF	0 0.000	0 0.000	3 .477	0 0.000	0 0.000	3 .477
DG	0 0.000	5 .699	123 2.090	105 2.021	0 0.000	233 2.367
DH	0 0.000	4 .602	0 0.000	8 .903	0 0.000	12 1.079
DI	392 2.593	87 1.940	379 2.579	101 2.004	0 0.000	959 2.982
DJ	0 0.000	16 1.204	0 0.000	0 0.000	0 0.000	16 1.204
DL	0 0.000	0 0.000	0 0.000	469 2.671	0 0.000	469 2.671
DM	0 0.000	10 1.000	3 .477	0 0.000	0 0.000	13 1.114

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
DN	0 0.000	0 0.000	20 1.301	0 0.000	0 0.000	20 1.301
DO	293 2.467	0 0.000	8 .903	398 2.600	0 0.000	699 2.844
DR	236 2.373	0 0.000	15 1.176	11 1.041	0 0.000	262 2.418
DS	0 0.000	0 0.000	56 1.748	3 .477	535 2.728	594 2.774
DT	0 0.000	0 0.000	5 .699	0 0.000	10 1.000	15 1.176
DU	649 2.812	23 1.362	88 1.944	4 .602	0 0.000	764 2.883
DV	0 0.000	66 1.820	0 0.000	0 0.000	0 0.000	66 1.820
DW	12 1.079	45 1.653	15 1.176	0 0.000	0 0.000	72 1.857
DY	0 0.000	0 0.000	0 0.000	0 0.000	234 2.369	234 2.369
EA	231 2.364	1875 3.273	410 2.613	593 2.773	6 .778	3115 3.493
EB	0 0.000	74 1.869	11 1.041	8 .903	0 0.000	93 1.968
EC	15 1.176	1638 3.214	109 2.037	743 2.871	0 0.000	2505 3.399
ED	134 2.127	131 2.117	282 2.450	85 1.929	8585 3.934	9217 3.965
EE	0 0.000	777 2.890	274 2.438	703 2.847	224 2.350	1978 3.296
EF	358 2.554	1180 3.072	85 1.929	6 .778	130 2.114	1759 3.245
EG	0 0.000	428 2.631	11 1.041	3 .477	0 0.000	442 2.645
EH	0 0.000	304 2.483	0 0.000	0 0.000	0 0.000	304 2.483
EI	344 2.537	196 2.292	111 2.045	3 .477	13 1.114	667 2.824
EJ	0 0.000	10 1.000	0 0.000	0 0.000	0 0.000	10 1.000
EK	0 0.000	0 0.000	24 1.380	18 1.255	5 .699	47 1.672
EL	89 1.949	586 2.768	163 2.212	1509 3.179	325 2.512	2672 3.427

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
EM	62 1.792	685 2.836	360 2.556	48 1.681	435 2.638	1590 3.201
EN	886 2.947	936 2.971	803 2.905	1274 3.105	1099 3.041	4998 3.699
EO	0 0.000	976 2.989	129 2.111	3 .477	12 1.079	1120 3.049
EP	0 0.000	343 2.535	76 1.881	314 2.497	35 1.544	768 2.885
EQ	23 1.362	0 0.000	0 0.000	0 0.000	0 0.000	23 1.362
ER	73 1.863	1551 3.191	252 2.401	1461 3.165	6964 3.843	10301 4.013
ES	144 2.158	769 2.886	82 1.914	601 2.779	3735 3.572	5331 3.727
ET	32 1.505	1071 3.030	253 2.403	130 2.114	1164 3.066	2650 3.423
EU	143 2.155	9 .954	0 0.000	32 1.505	4 .602	188 2.274
EV	126 2.100	312 2.494	63 1.799	28 1.447	0 0.000	529 2.723
EW	0 0.000	177 2.248	33 1.519	24 1.380	87 1.940	321 2.507
EX	649 2.812	97 1.987	10 1.000	0 0.000	4 .602	760 2.881
EY	4 .602	183 2.262	7 .845	0 0.000	284 2.453	478 2.679
EZ	0 0.000	0 0.000	0 0.000	33 1.519	0 0.000	33 1.519
FA	997 2.999	0 0.000	57 1.756	0 0.000	0 0.000	1054 3.023
FE	165 2.217	0 0.000	434 2.637	185 2.267	6 .778	790 2.898
FF	0 0.000	791 2.898	162 2.210	12 1.079	9 .954	974 2.989
FI	803 2.905	0 0.000	301 2.479	51 1.708	0 0.000	1155 3.063
FL	218 2.338	10 1.000	31 1.491	12 1.079	0 0.000	271 2.433
FO	926 2.967	0 0.000	1256 3.099	0 0.000	0 0.000	2182 3.339
FR	573 2.758	102 2.009	52 1.716	0 0.000	0 0.000	727 2.862

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
FS	0 0.000	0 0.000	9 .954	0 0.000	23 1.362	32 1.505
FT	0 0.000	0 0.000	99 1.996	26 1.415	0 0.000	125 2.097
FU	271 2.433	0 0.000	23 1.362	61 1.785	0 0.000	355 2.550
FY	0 0.000	0 0.000	0 0.000	0 0.000	19 1.279	19 1.279
GA	196 2.292	0 0.000	129 2.111	86 1.934	0 0.000	411 2.614
GE	322 2.508	100 2.000	124 2.093	1051 3.022	844 2.926	2441 3.388
GG	0 0.000	0 0.000	117 2.068	0 0.000	0 0.000	117 2.068
GH	16 1.204	0 0.000	544 2.736	722 2.859	872 2.941	2154 3.333
GI	138 2.140	0 0.000	226 2.354	150 2.176	0 0.000	514 2.711
GL	88 1.944	0 0.000	11 1.041	205 2.312	0 0.000	304 2.483
GM	0 0.000	0 0.000	0 0.000	4 .602	0 0.000	4 .602
GN	0 0.000	19 1.279	124 2.093	0 0.000	132 2.121	275 2.439
GO	119 2.076	0 0.000	36 1.556	33 1.519	3 .477	191 2.281
GR	661 2.820	92 1.964	134 2.127	23 1.362	0 0.000	910 2.959
GS	0 0.000	0 0.000	0 0.000	0 0.000	462 2.665	462 2.665
GT	0 0.000	0 0.000	0 0.000	116 2.064	0 0.000	116 2.064
GU	172 2.236	0 0.000	301 2.479	110 2.041	0 0.000	583 2.766
GY	0 0.000	0 0.000	0 0.000	0 0.000	112 2.049	112 2.049
HA	834 2.921	848 2.928	34 1.531	29 1.462	18 1.255	1763 3.246
HE	493 2.693	328 2.516	399 2.601	1739 3.240	17 1.230	2976 3.474
HI	316 2.500	586 2.768	287 2.458	370 2.568	0 0.000	1559 3.193

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
HL	0 0.000	0 0.000	0 0.000	99 1.996	0 0.000	99 1.996
HM	0 0.000	0 0.000	0 0.000	7 .845	22 1.342	29 1.462
HN	0 0.000	0 0.000	43 1.633	0 0.000	0 0.000	43 1.633
HO	463 2.666	1991 3.299	524 2.719	241 2.382	0 0.000	3219 3.508
HR	0 0.000	361 2.558	5 .699	0 0.000	0 0.000	366 2.563
HS	0 0.000	0 0.000	0 0.000	0 0.000	273 2.436	273 2.436
HT	0 0.000	0 0.000	0 0.000	244 2.387	703 2.847	947 2.976
HU	204 2.310	359 2.555	0 0.000	55 1.740	0 0.000	618 2.791
HW	0 0.000	3 .477	0 0.000	0 0.000	0 0.000	3 .477
HY	0 0.000	22 1.342	0 0.000	0 0.000	63 1.799	85 1.929
IA	0 0.000	56 1.748	43 1.633	535 2.728	99 1.996	733 2.865
IB	0 0.000	43 1.633	25 1.398	0 0.000	0 0.000	68 1.833
IC	3 .477	284 2.453	169 2.228	1060 3.025	739 2.869	2255 3.353
ID	16 1.204	322 2.508	195 2.290	348 2.542	168 2.225	1049 3.021
IE	0 0.000	253 2.403	203 2.307	1047 3.020	46 1.663	1549 3.190
IF	0 0.000	100 2.000	32 1.505	35 1.544	0 0.000	167 2.223
IG	19 1.279	870 2.940	462 2.665	132 2.121	8 .903	1491 3.173
II	0 0.000	0 0.000	7 .845	0 0.000	16 1.204	23 1.362
IK	0 0.000	169 2.228	0 0.000	92 1.964	0 0.000	261 2.417
IL	0 0.000	548 2.739	346 2.539	570 2.756	142 2.152	1606 3.206
IM	138 2.140	421 2.624	107 2.029	97 1.987	27 1.431	790 2.898

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
IN	873 2.941	828 2.918	916 2.962	4190 3.622	915 2.961	7722 3.888
IO	0 0.000	18 1.255	36 1.556	1100 3.041	31 1.491	1185 3.074
IP	0 0.000	29 1.462	5 .699	33 1.519	34 1.531	101 2.004
IQ	0 0.000	91 1.959	58 1.763	0 0.000	0 0.000	149 2.173
IR	17 1.230	389 2.590	325 2.512	406 2.609	73 1.863	1210 3.083
IS	298 2.474	800 2.903	392 2.593	480 2.681	167 2.223	2137 3.330
IT	304 2.483	1757 3.245	901 2.955	235 2.371	444 2.647	3641 3.561
IU	0 0.000	0 0.000	0 0.000	108 2.033	0 0.000	108 2.033
IV	0 0.000	394 2.595	122 2.086	208 2.318	0 0.000	724 2.860
IX	0 0.000	21 1.322	0 0.000	0 0.000	0 0.000	21 1.322
IZ	0 0.000	3 .477	30 1.477	0 0.000	0 0.000	33 1.519
JA	50 1.699	0 0.000	0 0.000	0 0.000	0 0.000	50 1.699
JE	119 2.076	0 0.000	84 1.924	0 0.000	0 0.000	203 2.307
JO	176 2.246	0 0.000	13 1.114	0 0.000	0 0.000	189 2.276
JU	196 2.292	0 0.000	43 1.633	0 0.000	0 0.000	239 2.378
KA	31 1.491	0 0.000	0 0.000	0 0.000	22 1.342	53 1.724
KE	6 .778	16 1.204	176 2.246	1927 3.285	127 2.104	2252 3.353
KI	152 2.182	57 1.756	543 2.735	12 1.079	0 0.000	764 2.883
KL	0 0.000	0 0.000	7 .845	42 1.623	0 0.000	49 1.690
KN	25 1.398	0 0.000	0 0.000	0 0.000	0 0.000	25 1.398
KO	11 1.041	0 0.000	6 .778	7 .845	0 0.000	24 1.380

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
KS	0 0.000	0 0.000	0 0.000	0 0.000	360 2.556	360 2.556
KU	0 0.000	0 0.000	0 0.000	14 1.146	0 0.000	14 1.146
KY	0 0.000	0 0.000	0 0.000	0 0.000	32 1.505	32 1.505
LA	518 2.714	914 2.961	321 2.507	155 2.190	19 1.279	1927 3.285
LB	0 0.000	47 1.672	0 0.000	0 0.000	0 0.000	47 1.672
LD	0 0.000	28 1.447	189 2.276	93 1.968	1004 3.002	1314 3.119
LE	718 2.856	210 2.322	316 2.500	1635 3.214	3499 3.544	6378 3.805
LF	0 0.000	52 1.716	0 0.000	0 0.000	454 2.657	506 2.704
LG	0 0.000	0 0.000	10 1.000	0 0.000	0 0.000	10 1.000
LI	1605 3.205	173 2.238	773 2.888	577 2.761	0 0.000	3128 3.495
LK	0 0.000	0 0.000	234 2.369	0 0.000	0 0.000	234 2.369
LL	0 0.000	78 1.892	1471 3.168	391 2.592	81 1.908	2021 3.306
LM	0 0.000	432 2.635	117 2.068	0 0.000	0 0.000	549 2.740
LN	0 0.000	0 0.000	11 1.041	0 0.000	0 0.000	11 1.041
LO	978 2.990	536 2.729	193 2.286	342 2.534	13 1.114	2062 3.314
LP	0 0.000	0 0.000	79 1.898	0 0.000	0 0.000	79 1.898
LS	0 0.000	0 0.000	78 1.892	0 0.000	500 2.699	578 2.762
LT	0 0.000	0 0.000	144 2.158	194 2.288	259 2.413	597 2.776
LU	81 1.908	57 1.756	413 2.616	4 .602	0 0.000	555 2.744
LV	0 0.000	0 0.000	68 1.833	53 1.724	0 0.000	121 2.083
LW	0 0.000	458 2.661	0 0.000	0 0.000	0 0.000	458 2.661

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
LY	18	43	0	4	2892	2957
	1.255	1.633	0.000	.602	3.461	3.471
MA	1493	48	346	540	28	2455
	3.174	1.681	2.539	2.732	1.447	3.390
MB	0	12	788	12	0	812
	0.000	1.079	2.897	1.079	0.000	2.910
ME	801	27	420	988	1022	3258
	2.904	1.431	2.623	2.995	3.009	3.513
MI	436	85	613	194	7	1335
	2.639	1.929	2.787	2.288	.845	3.125
ML	0	0	7	79	0	86
	0.000	0.000	.845	1.898	0.000	1.934
MM	0	5	433	0	0	438
	0.000	.699	2.636	0.000	0.000	2.641
MN	0	0	19	9	105	133
	0.000	0.000	1.279	.954	2.021	2.124
MO	1637	223	707	248	0	2815
	3.214	2.348	2.849	2.394	0.000	3.449
MP	0	135	532	32	0	699
	0.000	2.130	2.726	1.505	0.000	2.844
MS	0	0	10	6	120	136
	0.000	0.000	1.000	.778	2.079	2.134
MT	0	0	0	28	0	28
	0.000	0.000	0.000	1.447	0.000	1.447
MU	215	9	37	7	0	268
	2.332	.954	1.568	.845	0.000	2.428
MY	136	4	0	0	15	155
	2.134	.602	0.000	0.000	1.176	2.190
NA	512	113	193	66	51	935
	2.709	2.053	2.286	1.820	1.708	2.971
NB	0	4	0	0	0	4
	0.000	.602	0.000	0.000	0.000	.602
NC	0	216	157	544	0	917
	0.000	2.334	2.196	2.736	0.000	2.962
ND	0	297	717	284	2237	3535
	0.000	2.473	2.856	2.453	3.350	3.548
NE	519	128	236	1362	447	2692
	2.715	2.107	2.373	3.134	2.650	3.430
NF	0	35	6	0	0	41
	0.000	1.544	.778	0.000	0.000	1.613
NG	0	107	773	765	3776	5421
	0.000	2.029	2.888	2.884	3.577	3.734

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
NI	68 1.833	673 2.828	380 2.580	71 1.851	9 .954	1201 3.080
NJ	0 0.000	43 1.633	0 0.000	0 0.000	0 0.000	43 1.633
NK	0 0.000	10 1.000	72 1.857	100 2.000	5 .699	187 2.272
NL	0 0.000	165 2.217	6 .778	75 1.875	0 0.000	246 2.391
NM	0 0.000	0 0.000	3 .477	0 0.000	0 0.000	3 .477
NN	0 0.000	113 2.053	582 2.765	67 1.826	0 0.000	762 2.882
NO	450 2.653	430 2.633	72 1.857	286 2.456	0 0.000	1238 3.093
NP	0 0.000	9 .954	0 0.000	0 0.000	0 0.000	9 .954
NR	0 0.000	25 1.398	22 1.342	0 0.000	0 0.000	47 1.672
NS	0 0.000	445 2.648	118 2.072	0 0.000	307 2.487	870 2.940
NT	0 0.000	231 2.364	955 2.980	751 2.876	1086 3.036	3023 3.480
NU	494 2.694	3 .477	176 2.246	63 1.799	0 0.000	736 2.867
NV	0 0.000	35 1.544	51 1.708	0 0.000	0 0.000	86 1.934
NW	0 0.000	12 1.079	4 .602	0 0.000	0 0.000	16 1.204
NY	0 0.000	206 2.314	12 1.079	0 0.000	87 1.940	305 2.484
NZ	0 0.000	6 .778	0 0.000	17 1.230	0 0.000	23 1.362
OA	0 0.000	131 2.117	28 1.447	116 2.064	0 0.000	275 2.439
OB	114 2.057	214 2.330	7 .845	0 0.000	0 0.000	335 2.525
OC	46 1.663	624 2.795	67 1.826	3 .477	0 0.000	740 2.869
OD	3 .477	436 2.639	71 1.851	123 2.090	407 2.610	1040 3.017
OE	0 0.000	119 2.076	0 0.000	32 1.505	12 1.079	163 2.212

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
OF	313 2.496	125 2.097	28 1.447	0 0.000	0 0.000	466 2.668
OG	0 0.000	9 .954	7 .845	0 0.000	0 0.000	16 1.204
OH	0 0.000	30 1.477	0 0.000	0 0.000	0 0.000	30 1.477
OI	0 0.000	314 2.497	116 2.064	30 1.477	0 0.000	460 2.663
OK	0 0.000	10 1.000	531 2.725	39 1.591	0 0.000	580 2.763
OL	21 1.322	1189 3.075	114 2.057	85 1.929	598 2.777	2007 3.303
OM	0 0.000	781 2.893	179 2.253	481 2.682	223 2.348	1664 3.221
ON	8 .903	849 2.929	105 2.021	1030 3.013	2267 3.355	4259 3.629
OO	0 0.000	592 2.772	193 2.286	526 2.721	0 0.000	1311 3.118
OP	197 2.294	104 2.017	1007 3.003	171 2.233	22 1.342	1501 3.176
OR	178 2.250	1523 3.183	392 2.593	2101 3.322	762 2.882	4956 3.695
OS	0 0.000	428 2.631	268 2.428	803 2.905	35 1.544	1534 3.186
OT	323 2.509	853 2.931	72 1.857	107 2.029	288 2.459	1643 3.216
OU	82 1.914	1606 3.206	2869 3.458	123 2.090	0 0.000	4680 3.670
OV	8 .903	407 2.610	130 2.114	62 1.792	0 0.000	607 2.783
OW	38 1.580	598 2.777	504 2.702	83 1.919	573 2.758	1796 3.254
OX	61 1.785	0 0.000	0 0.000	0 0.000	3 .477	64 1.806
OY	6 .778	26 1.415	0 0.000	10 1.000	31 1.491	73 1.863
OZ	0 0.000	18 1.255	27 1.431	0 0.000	0 0.000	45 1.653
PA	849 2.929	44 1.643	127 2.104	0 0.000	0 0.000	1020 3.009
PE	1450 3.161	278 2.444	511 2.708	639 2.806	197 2.294	3075 3.488

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
PH	108 2.033	35 1.544	9 .954	15 1.176	55 1.740	222 2.346
PI	313 2.496	193 2.286	156 2.193	52 1.716	0 0.000	714 2.854
PK	0 0.000	6 .778	3 .477	0 0.000	0 0.000	9 .954
PL	688 2.838	3 .477	12 1.079	1564 3.194	0 0.000	2267 3.355
PO	1075 3.031	101 2.004	233 2.367	46 1.663	0 0.000	1455 3.163
PP	0 0.000	195 2.290	298 2.474	3 .477	0 0.000	496 2.695
PR	610 2.785	239 2.378	6 .778	0 0.000	0 0.000	855 2.932
PS	7 .845	11 1.041	7 .845	7 .845	211 2.324	243 2.386
PT	0 0.000	22 1.342	19 1.279	0 0.000	298 2.474	339 2.530
PU	727 2.862	0 0.000	17 1.230	75 1.875	0 0.000	819 2.913
PW	0 0.000	27 1.431	0 0.000	0 0.000	0 0.000	27 1.431
PY	0 0.000	0 0.000	0 0.000	0 0.000	25 1.398	25 1.398
QU	70 1.845	166 2.220	91 1.959	64 1.806	0 0.000	391 2.592
RA	817 2.912	674 2.829	269 2.430	41 1.613	40 1.602	1841 3.265
RB	0 0.000	9 .954	132 2.121	44 1.643	40 1.602	225 2.352
RC	0 0.000	21 1.322	404 2.606	528 2.723	0 0.000	953 2.979
RD	0 0.000	73 1.863	360 2.556	118 2.072	875 2.942	1426 3.154
RE	2643 3.422	709 2.851	1436 3.157	909 2.959	2364 3.374	8061 3.906
RG	0 0.000	91 1.959	417 2.620	381 2.581	0 0.000	889 2.949
RH	30 1.477	0 0.000	0 0.000	0 0.000	0 0.000	30 1.477
RI	312 2.494	1231 3.090	1756 3.245	268 2.428	0 0.000	3567 3.552

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
RK	0	0	542	15	44	601
	0.000	0.000	2.734	1.176	1.643	2.779
RL	0	0	140	237	0	377
	0.000	0.000	2.146	2.375	0.000	2.576
RM	0	15	745	14	49	823
	0.000	1.176	2.872	1.146	1.690	2.915
RN	0	30	561	10	399	1000
	0.000	1.477	2.749	1.000	2.601	3.000
RO	254	1520	953	199	32	2958
	2.405	3.182	2.979	2.299	1.505	3.471
RP	0	0	50	0	0	50
	0.000	0.000	1.699	0.000	0.000	1.699
RR	0	100	311	42	0	453
	0.000	2.000	2.493	1.623	0.000	2.656
RS	0	0	340	489	1200	2029
	0.000	0.000	2.531	2.689	3.079	3.307
RT	0	154	286	363	602	1405
	0.000	2.188	2.456	2.560	2.780	3.148
RU	220	85	124	22	0	451
	2.342	1.929	2.093	1.342	0.000	2.654
RV	0	4	293	0	0	297
	0.000	.602	2.467	0.000	0.000	2.473
RY	0	207	0	0	543	750
	0.000	2.316	0.000	0.000	2.735	2.875
SA	530	11	35	61	0	637
	2.724	1.041	1.544	1.785	0.000	2.804
SC	722	74	220	4	0	1020
	2.859	1.869	2.342	.602	0.000	3.009
SD	0	0	44	0	0	44
	0.000	0.000	1.643	0.000	0.000	1.643
SE	1846	58	677	1384	819	4784
	3.266	1.763	2.831	3.141	2.913	3.680
SH	1464	6	276	3	188	1937
	3.166	.778	2.441	.477	2.274	3.287
SI	836	0	897	228	0	1961
	2.922	0.000	2.953	2.358	0.000	3.292
SK	73	67	25	45	0	210
	1.863	1.826	1.398	1.653	0.000	2.322
SL	291	196	3	0	0	490
	2.464	2.292	.477	0.000	0.000	2.690
SM	146	0	34	13	13	206
	2.164	0.000	1.531	1.114	1.114	2.314

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
SN	36	0	0	0	0	36
1.556	0.000	0.000	0.000	0.000	0.000	1.556
SO	886	0	25	817	0	1728
2.947	0.000	1.398	2.912	0.000	0.000	3.238
SP	698	56	21	0	0	775
2.844	1.748	1.322	0.000	0.000	0.000	2.889
SQ	143	0	0	0	0	143
2.155	0.000	0.000	0.000	0.000	0.000	2.155
SR	0	15	0	3	0	18
0.000	1.176	0.000	.477	0.000	0.000	1.255
SS	0	313	562	0	565	1440
0.000	2.496	2.750	0.000	2.752	0.000	3.158
ST	2484	59	1132	142	1104	4921
3.395	1.771	3.054	2.152	3.043	0.000	3.692
SU	704	0	607	84	0	1395
2.848	0.000	2.783	1.924	0.000	0.000	3.145
SW	83	0	152	0	0	235
1.919	0.000	2.182	0.000	0.000	0.000	2.371
SY	481	7	0	0	28	516
2.682	.845	0.000	0.000	1.447	0.000	2.713
TA	519	1251	416	142	38	2366
2.715	3.097	2.619	2.152	1.580	0.000	3.374
TC	0	0	0	76	0	76
0.000	0.000	0.000	1.881	0.000	0.000	1.881
TE	203	57	505	4357	380	5502
2.307	1.756	2.703	3.639	2.580	0.000	3.741
TF	0	0	16	0	0	16
0.000	0.000	1.204	0.000	0.000	0.000	1.204
TH	1419	363	1739	335	637	4493
3.152	2.560	3.240	2.525	2.804	0.000	3.653
TI	145	31	1352	197	0	1725
2.161	1.491	3.131	2.294	0.000	0.000	3.237
TL	0	0	50	1462	0	1512
0.000	0.000	1.699	3.165	0.000	0.000	3.180
TM	0	7	0	0	0	7
0.000	.845	0.000	0.000	0.000	0.000	.845
TO	508	202	74	685	54	1523
2.706	2.305	1.869	2.836	1.732	0.000	3.183
TP	0	0	35	0	0	35
0.000	0.000	1.544	0.000	0.000	0.000	1.544
TR	614	887	63	130	0	1694
2.788	2.948	1.799	2.114	0.000	0.000	3.229

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

BIGRAM POSITION		1ST	2ND	3RD	4TH	5TH	SUM
TS	0	304	58	0	1256	1618	
	0.000	2.483	1.763	0.000	3.099	3.209	
TT	0	199	2414	118	5	2736	
	0.000	2.299	3.383	2.072	.699	3.437	
TU	370	103	809	148	0	1430	
	2.568	2.013	2.908	2.170	0.000	3.155	
TW	144	0	0	0	0	144	
	2.158	0.000	0.000	0.000	0.000	2.158	
TY	7	23	0	8	779	817	
	.845	1.362	0.000	.903	2.892	2.912	
UA	0	44	175	374	4	597	
	0.000	1.643	2.243	2.573	.602	2.776	
UB	0	577	82	7	0	666	
	0.000	2.761	1.914	.845	0.000	2.823	
UC	0	52	29	145	0	226	
	0.000	1.716	1.462	2.161	0.000	2.354	
UD	0	203	73	54	4	334	
	0.000	2.307	1.863	1.732	.602	2.524	
UE	0	96	5	394	343	838	
	0.000	1.982	.699	2.595	2.535	2.923	
UF	0	59	4	4	0	67	
	0.000	1.771	.602	.602	0.000	1.826	
UG	0	134	316	883	0	1333	
	0.000	2.127	2.500	2.946	0.000	3.125	
UI	0	133	44	48	0	225	
	0.000	2.124	1.643	1.681	0.000	2.352	
UL	0	209	47	1145	64	1465	
	0.000	2.320	1.672	3.059	1.806	3.166	
UM	0	759	40	311	128	1238	
	0.000	2.880	1.602	2.493	2.107	3.093	
UN	874	333	352	934	4	2497	
	2.942	2.522	2.547	2.970	.602	3.397	
UO	0	31	0	43	0	74	
	0.000	1.491	0.000	1.633	0.000	1.869	
UP	74	188	150	162	14	588	
	1.869	2.274	2.176	2.210	1.146	2.769	
UR	31	1544	1086	1078	62	3801	
	1.491	3.189	3.036	3.033	1.792	3.580	
US	69	316	356	177	338	1256	
	1.839	2.500	2.551	2.248	2.529	3.099	
UT	31	492	41	149	62	775	
	1.491	2.692	1.613	2.173	1.792	2.889	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
UU	0 0.000	0 0.000	0 0.000	20 1.301	0 0.000	20 1.301
UX	0 0.000	27 1.431	0 0.000	0 0.000	4 .602	31 1.491
UY	0 0.000	35 1.544	0 0.000	0 0.000	0 0.000	35 1.544
UZ	0 0.000	20 1.301	0 0.000	0 0.000	0 0.000	20 1.301
VA	383 2.583	0 0.000	48 1.681	19 1.279	17 1.230	467 2.669
VE	123 2.090	163 2.212	322 2.508	929 2.968	347 2.540	1884 3.275
VI	309 2.490	0 0.000	1185 3.074	13 1.114	0 0.000	1507 3.178
VO	251 2.400	20 1.301	51 1.708	23 1.362	0 0.000	345 2.538
VU	7 .845	0 0.000	0 0.000	0 0.000	0 0.000	7 .845
WA	898 2.953	36 1.556	1039 3.017	65 1.813	0 0.000	2038 3.309
WB	0 0.000	0 0.000	16 1.204	0 0.000	0 0.000	16 1.204
WD	0 0.000	0 0.000	28 1.447	12 1.079	8 .903	48 1.681
WE	226 2.354	147 2.167	145 2.161	418 2.621	0 0.000	936 2.971
WH	90 1.954	4 .602	0 0.000	0 0.000	0 0.000	94 1.973
WI	910 2.959	79 1.898	87 1.940	7 .845	0 0.000	1083 3.035
WL	0 0.000	0 0.000	0 0.000	115 2.061	3 .477	118 2.072
WM	0 0.000	0 0.000	10 1.000	0 0.000	0 0.000	10 1.000
WN	0 0.000	38 1.580	5 .699	0 0.000	51 1.708	94 1.973
WO	341 2.533	4 .602	0 0.000	0 0.000	0 0.000	345 2.538
WR	180 2.255	0 0.000	0 0.000	0 0.000	0 0.000	180 2.255
WS	0 0.000	0 0.000	3 .477	0 0.000	48 1.681	51 1.708

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
WT	0 0.000	0 0.000	6 .778	155 2.190	0 0.000	161 2.207
WY	0 0.000	0 0.000	43 1.633	0 0.000	0 0.000	43 1.633
XA	0 0.000	0 0.000	9 .954	0 0.000	0 0.000	9 .954
XC	0 0.000	276 2.441	0 0.000	0 0.000	0 0.000	276 2.441
XE	0 0.000	5 .699	0 0.000	13 1.114	0 0.000	18 1.255
XF	0 0.000	18 1.255	0 0.000	0 0.000	0 0.000	18 1.255
XI	0 0.000	42 1.623	49 1.690	0 0.000	0 0.000	91 1.959
XL	0 0.000	0 0.000	6 .778	0 0.000	0 0.000	6 .778
XO	0 0.000	11 1.041	0 0.000	0 0.000	0 0.000	11 1.041
XP	0 0.000	169 2.228	0 0.000	0 0.000	0 0.000	169 2.228
XT	0 0.000	146 2.164	10 1.000	0 0.000	0 0.000	156 2.193
XU	0 0.000	0 0.000	80 1.903	0 0.000	0 0.000	80 1.903
XY	0 0.000	43 1.633	0 0.000	0 0.000	3 .477	46 1.663
YA	26 1.415	0 0.000	17 1.230	0 0.000	0 0.000	43 1.633
YC	0 0.000	7 .845	7 .845	0 0.000	0 0.000	14 1.146
YD	0 0.000	5 .699	6 .778	0 0.000	0 0.000	11 1.041
YE	89 1.949	4 .602	15 1.176	332 2.521	0 0.000	440 2.643
YG	0 0.000	0 0.000	43 1.633	0 0.000	0 0.000	43 1.633
YH	0 0.000	0 0.000	20 1.301	0 0.000	0 0.000	20 1.301
YI	7 .845	0 0.000	435 2.638	0 0.000	0 0.000	442 2.645
YL	0 0.000	0 0.000	49 1.690	0 0.000	0 0.000	49 1.690

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 6

	BIGRAM POSITION					
	1ST	2ND	3RD	4TH	5TH	SUM
YM	0	54	9	6	0	69
	.0000	1.732	.954	.778	.0000	1.839
YN	0	12	0	0	4	16
	.0000	1.079	0.000	0.000	.602	1.204
YO	18	0	326	12	0	356
	1.255	0.000	2.513	1.079	0.000	2.551
YP	0	10	0	0	0	10
	.0000	1.000	0.000	0.000	0.000	1.000
YR	0	22	0	0	8	30
	.0000	1.342	0.000	0.000	.903	1.477
YS	0	551	0	0	482	1033
	.0000	2.741	0.000	0.000	2.683	3.014
YT	0	0	26	0	0	26
	.0000	0.000	1.415	0.000	0.000	1.415
YW	0	0	46	0	0	46
	.0000	0.000	1.663	0.000	0.000	1.663
ZA	0	0	22	0	17	39
	.0000	0.000	1.342	0.000	1.230	1.591
ZE	0	0	11	73	47	131
	.0000	0.000	1.041	1.863	1.672	2.117
ZI	0	0	11	8	0	19
	.0000	0.000	1.041	.903	0.000	1.279
ZL	0	0	0	24	0	24
	.0000	0.000	0.000	1.380	0.000	1.380
ZO	6	0	0	0	0	6
	.778	0.000	0.000	0.000	0.000	.778
ZY	0	0	6	0	6	12
	.0000	0.000	.778	0.000	.778	1.079
ZZ	0	0	24	17	0	41
	.0000	0.000	1.380	1.230	0.000	1.613

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
AB	178 2.250	101 2.004	64 1.806	166 2.220	5 .699	0 0.000	514 2.711
AC	356 2.551	521 2.717	543 2.735	65 1.813	432 2.635	5 .699	1922 3.284
AD	285 2.455	87 1.940	593 2.773	100 2.000	328 2.516	180 2.255	1573 3.197
AE	0 0.000	4 .602	0 0.000	0 0.000	11 1.041	0 0.000	15 1.176
AF	150 2.176	5 .699	81 1.908	0 0.000	0 0.000	0 0.000	236 2.373
AG	629 2.799	66 1.820	175 2.243	200 2.301	635 2.803	3 .477	1708 3.232
AH	0 0.000	0 0.000	0 0.000	6 .778	0 0.000	0 0.000	6 .778
AI	34 1.531	433 2.636	838 2.923	270 2.431	661 2.820	0 0.000	2236 3.349
AJ	0 0.000	3 .477	0 0.000	0 0.000	0 0.000	0 0.000	3 .477
AK	0 0.000	4 .602	40 1.602	67 1.826	40 1.602	0 0.000	151 2.179
AL	439 2.642	574 2.759	652 2.814	650 2.813	185 2.267	3015 3.479	5515 3.742
AM	315 2.498	78 1.892	459 2.662	89 1.949	40 1.602	561 2.749	1542 3.188
AN	1162 3.065	740 2.869	894 2.951	582 2.765	1086 3.036	352 2.547	4816 3.683
AO	0 0.000	4 .602	5 .699	0 0.000	0 0.000	0 0.000	9 .954
AP	253 2.403	465 2.667	232 2.365	67 1.826	319 2.504	4 .602	1340 3.127
AQ	14 1.146	0 0.000	0 .602	0 0.000	0 0.000	0 0.000	18 1.255
AR	304 2.483	1936 3.287	1160 3.064	365 2.562	953 2.979	547 2.738	5265 3.721
AS	238 2.377	480 2.681	555 2.744	111 2.045	138 2.140	82 1.914	1604 3.205
AT	189 2.276	972 2.988	477 2.679	853 2.931	393 2.594	32 1.505	2916 3.465
AU	44 1.643	125 2.097	3 .477	908 2.958	5 .699	3 .477	1088 3.037
AV	157 2.196	80 1.903	258 2.412	21 1.322	4 .602	0 0.000	520 2.716

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
AW	49	26	85	10	0	0	170
	1.690	1.415	1.929	1.000	0.000	0.000	2.230
AX	0	101	0	17	0	0	118
	0.000	2.004	0.000	1.230	0.000	0.000	2.072
AY	0	154	202	35	18	252	661
	0.000	2.188	2.305	1.544	1.255	2.401	2.820
AZ	0	20	33	0	0	0	53
	0.000	1.301	1.519	0.000	0.000	0.000	1.724
BA	308	17	50	175	3	0	553
	2.489	1.230	1.699	2.243	.477	0.000	2.743
BB	0	0	63	23	0	0	86
	0.000	0.000	1.799	1.362	0.000	0.000	1.934
BD	0	6	8	0	0	0	14
	0.000	.778	.903	0.000	0.000	0.000	1.146
BE	2413	13	177	502	188	7	3300
	3.383	1.114	2.248	2.701	2.274	.845	3.519
BI	150	79	43	73	33	0	378
	2.176	1.898	1.633	1.863	1.519	0.000	2.577
BJ	0	65	161	0	0	0	226
	0.000	1.813	2.207	0.000	0.000	0.000	2.354
BL	182	21	8	397	364	0	972
	2.260	1.322	.903	2.599	2.561	0.000	2.988
BO	195	8	14	98	17	0	332
	2.290	.903	1.146	1.991	1.230	0.000	2.521
BR	702	11	127	4	0	0	844
	2.846	1.041	2.104	.602	0.000	0.000	2.926
BS	0	95	8	0	0	18	121
	0.000	1.978	.903	0.000	0.000	1.255	2.083
BT	0	0	0	9	0	0	9
	0.000	0.000	0.000	.954	0.000	0.000	.954
BU	197	0	31	76	0	0	304
	2.294	0.000	1.491	1.881	0.000	0.000	2.483
BV	0	92	0	0	0	0	92
	0.000	1.964	0.000	0.000	0.000	0.000	1.964
BY	0	0	0	0	6	60	66
	0.000	0.000	0.000	0.000	.778	1.778	1.820
CA	813	45	1059	119	516	194	2746
	2.910	1.653	3.025	2.076	2.713	2.288	3.439
CC	0	157	108	0	24	0	289
	0.000	2.196	2.033	0.000	1.380	0.000	2.461
CE	782	26	153	711	535	1291	3498
	2.893	1.415	2.185	2.852	2.728	3.111	3.544

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
CH	1075 3.031	267 2.427	202 2.305	718 2.856	32 1.505	78 1.892	2372 3.375
CI	94 1.973	131 2.117	554 2.744	513 2.710	139 2.143	0 0.000	1431 3.156
CK	0 0.000	0 0.000	307 2.487	304 2.483	22 1.342	32 1.505	665 2.823
CL	748 2.874	0 0.000	293 2.467	72 1.857	127 2.104	0 0.000	1240 3.093
CN	0 0.000	0 0.000	3 .477	0 0.000	0 0.000	0 0.000	3 .477
CO	2983 3.475	101 2.004	372 2.571	156 2.193	0 0.000	27 1.431	3639 3.561
CQ	0 0.000	27 1.431	8 .903	0 0.000	0 0.000	0 0.000	35 1.544
CR	413 2.616	56 1.748	49 1.690	0 0.000	0 0.000	0 0.000	518 2.714
CS	0 0.000	6 .778	0 0.000	0 0.000	0 0.000	103 2.013	109 2.037
CT	0 0.000	125 2.097	719 2.857	234 2.369	340 2.531	911 2.960	2329 3.367
CU	344 2.537	5 .699	145 2.161	94 1.973	0 0.000	0 0.000	588 2.769
CY	16 1.204	0 0.000	5 .699	0 0.000	0 0.000	71 1.851	92 1.964
DA	140 2.146	18 1.255	3 .477	40 1.602	96 1.982	46 1.663	343 2.535
DB	0 0.000	0 0.000	0 0.000	9 .954	3 .477	0 0.000	12 1.079
DC	0 0.000	0 0.000	0 0.000	4 .602	0 0.000	0 0.000	4 .602
DD	0 0.000	81 1.908	98 1.991	13 1.114	0 0.000	0 0.000	192 2.283
DE	1546 3.189	10 1.000	385 2.585	502 2.701	727 2.862	629 2.799	3799 3.580
DF	0 0.000	0 0.000	0 0.000	18 1.255	0 0.000	0 0.000	18 1.255
DG	0 0.000	0 0.000	37 1.568	41 1.613	0 0.000	0 0.000	78 1.892
DH	0 0.000	5 .699	7 .845	0 0.000	0 0.000	0 0.000	12 1.079
DI	517 2.713	63 1.799	285 2.455	1004 3.002	5 .699	0 0.000	1874 3.273

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
DJ	0	6	0	0	0	0	6
	0.000	.778	0.000	0.000	0.000	0.000	.778
DL	0	0	20	78	136	0	234
	0.000	0.000	1.301	1.892	2.134	0.000	2.369
DM	0	20	0	0	13	0	33
	0.000	1.301	0.000	0.000	1.114	0.000	1.519
DN	0	0	11	4	0	0	15
	0.000	0.000	1.041	.602	0.000	0.000	1.176
DO	218	45	35	116	199	7	620
	2.338	1.653	1.544	2.064	2.299	.845	2.792
DR	372	0	129	178	5	0	684
	2.571	0.000	2.111	2.250	.699	0.000	2.835
DS	0	0	5	0	0	867	872
	0.000	0.000	.699	0.000	0.000	2.938	2.941
DT	0	0	4	0	7	0	11
	0.000	0.000	.602	0.000	.845	0.000	1.041
DU	32	7	138	251	11	0	439
	1.505	.845	2.140	2.400	1.041	0.000	2.642
DV	0	110	0	0	0	0	110
	0.000	2.041	0.000	0.000	0.000	0.000	2.041
DW	0	3	11	5	0	0	19
	0.000	.477	1.041	.699	0.000	0.000	1.279
DY	21	4	0	0	0	510	535
	1.322	.602	0.000	0.000	0.000	2.708	2.728
EA	231	2239	934	683	364	0	4451
	2.364	3.350	2.970	2.834	2.561	0.000	3.648
EB	0	28	4	0	52	0	84
	0.000	1.447	.602	0.000	1.716	0.000	1.924
EC	85	1753	454	336	640	0	3268
	1.929	3.244	2.657	2.526	2.806	0.000	3.514
ED	70	614	60	159	210	7760	8873
	1.845	2.788	1.778	2.201	2.322	3.890	3.948
EE	0	560	226	105	919	28	1838
	0.000	2.748	2.354	2.021	2.963	1.447	3.264
EF	236	386	8	203	23	0	856
	2.373	2.587	.903	2.307	1.362	0.000	2.932
EG	0	261	18	32	267	0	578
	0.000	2.417	1.255	1.505	2.427	0.000	2.762
EH	0	48	0	72	0	0	120
	0.000	1.681	0.000	1.857	0.000	0.000	2.079
EI	0	243	28	251	35	0	557
	0.000	2.386	1.447	2.400	1.544	0.000	2.746

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
EJ	0	11	0	0	0	0	11
	0.000	1.041	0.000	0.000	0.000	0.000	1.041
EK	0	0	74	0	0	0	74
	0.000	0.000	1.869	0.000	0.000	0.000	1.869
EL	139	869	372	200	1007	127	2714
	2.143	2.939	2.571	2.301	3.003	2.104	3.434
EM	168	696	167	146	270	313	1760
	2.225	2.843	2.223	2.164	2.431	2.496	3.246
EN	747	1751	291	1159	1505	1205	6658
	2.873	3.243	2.464	3.064	3.178	3.081	3.823
EO	0	97	0	123	27	0	247
	0.000	1.987	0.000	2.090	1.431	0.000	2.393
EP	12	265	175	20	85	0	557
	1.079	2.423	2.243	1.301	1.929	0.000	2.746
EQ	70	135	0	0	0	0	205
	1.845	2.130	0.000	0.000	0.000	0.000	2.312
ER	37	1985	761	2156	2492	4146	11577
	1.568	3.298	2.881	3.334	3.397	3.618	4.064
ES	45	1214	574	59	1100	2944	5936
	1.653	3.084	2.759	1.771	3.041	3.469	3.773
ET	58	1557	476	147	693	96	3027
	1.763	3.192	2.678	2.167	2.841	1.982	3.481
EU	0	50	0	10	36	0	96
	0.000	1.699	0.000	1.000	1.556	0.000	1.982
EV	205	766	111	621	264	0	1967
	2.312	2.884	2.045	2.793	2.422	0.000	3.294
EW	0	48	32	47	9	4	140
	0.000	1.681	1.505	1.672	.954	.602	2.146
EX	954	67	0	0	38	91	1150
	2.980	1.826	0.000	0.000	1.580	1.959	3.061
EY	11	12	3	0	26	121	173
	1.041	1.079	.477	0.000	1.415	2.083	2.238
EZ	0	0	0	3	11	0	14
	0.000	0.000	0.000	.477	1.041	0.000	1.146
FA	674	0	95	318	7	0	1094
	2.829	0.000	1.978	2.502	.845	0.000	3.039
FE	554	0	403	94	17	0	1068
	2.744	0.000	2.605	1.973	1.230	0.000	3.029
FF	0	595	59	101	0	20	775
	0.000	2.775	1.771	2.004	0.000	1.301	2.889
FI	707	0	231	98	166	0	1202
	2.849	0.000	2.364	1.991	2.220	0.000	3.080

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
FL	206 2.314	0 0.000	29 1.462	19 1.279	72 1.857	0 0.000	326 2.513
FO	668 2.825	0 0.000	161 2.207	137 2.137	0 0.000	0 0.000	966 2.985
FR	421 2.624	28 1.447	10 1.000	0 0.000	0 0.000	0 0.000	459 2.662
FS	0 0.000	0 0.000	0 0.000	0 0.000	0 0.000	23 1.362	23 1.362
FT	0 0.000	0 0.000	76 1.881	55 1.740	8 .903	0 0.000	139 2.143
FU	318 2.502	0 0.000	98 1.991	19 1.279	188 2.274	0 0.000	623 2.794
FW	0 0.000	0 0.000	0 0.000	18 1.255	0 0.000	0 0.000	18 1.255
FY	0 0.000	0 0.000	0 0.000	0 0.000	0 0.000	89 1.949	89 1.949
GA	157 2.196	626 2.797	107 2.029	47 1.672	18 1.255	21 1.322	976 2.989
GD	0 0.000	0 0.000	0 0.000	26 1.415	0 0.000	0 0.000	26 1.415
GE	910 2.959	0 0.000	89 1.949	429 2.632	768 2.885	927 2.967	3123 3.495
GG	0 0.000	0 0.000	144 2.158	43 1.633	0 0.000	0 0.000	187 2.272
GH	6 .778	0 0.000	221 2.344	161 2.207	907 2.958	974 2.989	2269 3.356
GI	15 1.176	3 .477	120 2.079	327 2.515	46 1.663	0 0.000	511 2.708
GL	117 2.068	0 0.000	361 2.558	56 1.748	0 0.000	0 0.000	534 2.728
GM	0 0.000	0 0.000	19 1.279	0 0.000	0 0.000	0 0.000	19 1.279
GN	4 .602	34 1.531	71 1.851	0 0.000	41 1.613	158 2.199	308 2.489
GO	25 1.398	0 0.000	0 0.000	4 .602	0 0.000	106 2.025	135 2.130
GR	708 2.850	0 0.000	94 1.973	416 2.619	0 0.000	0 0.000	1218 3.086
GS	0 0.000	0 0.000	0 0.000	0 0.000	4 .602	102 2.009	106 2.025
GT	0 0.000	0 0.000	0 0.000	34 1.531	0 0.000	0 0.000	34 1.531

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
GU	40 1.602	0 0.000	267 2.427	21 1.322	19 1.279	0 0.000	347 2.540
GY	0 0.000	0 0.000	0 0.000	0 0.000	0 0.000	35 1.544	35 1.544
HA	427 2.630	884 2.946	67 1.826	397 2.599	32 1.505	0 0.000	1807 3.257
HE	505 2.703	781 2.893	19 1.279	155 2.190	2568 3.410	32 1.505	4060 3.609
HF	0 0.000	0 0.000	0 0.000	9 .954	0 0.000	0 0.000	9 .954
HI	1033 3.014	385 2.585	148 2.170	731 2.864	56 1.748	0 0.000	2353 3.372
HL	0 0.000	0 0.000	9 .954	0 0.000	58 1.763	0 0.000	67 1.826
HM	0 0.000	0 0.000	0 0.000	0 0.000	12 1.079	0 0.000	12 1.079
HN	0 0.000	0 0.000	34 1.531	0 0.000	0 0.000	0 0.000	34 1.531
HO	863 2.936	761 2.881	217 2.336	794 2.900	85 1.929	0 0.000	2720 3.435
HR	0 0.000	1053 3.022	0 0.000	0 0.000	0 0.000	0 0.000	1053 3.022
HS	0 0.000	0 0.000	0 0.000	0 0.000	0 0.000	26 1.415	26 1.415
HT	0 0.000	0 0.000	0 0.000	117 2.068	60 1.778	907 2.958	1084 3.035
HU	401 2.603	27 1.431	7 .845	0 0.000	3 .477	0 0.000	438 2.641
HW	0 0.000	0 0.000	0 0.000	40 1.602	0 0.000	0 0.000	40 1.602
HY	3 .477	72 1.857	0 0.000	0 0.000	0 0.000	70 1.845	145 2.161
IA	0 0.000	55 1.740	0 0.000	79 1.898	719 2.857	76 1.881	929 2.968
IB	0 0.000	203 2.307	40 1.602	85 1.929	0 0.000	0 0.000	328 2.516
IC	0 0.000	526 2.721	288 2.459	888 2.948	794 2.900	350 2.544	2846 3.454
ID	17 1.230	57 1.756	108 2.033	432 2.635	498 2.697	49 1.690	1161 3.065
IE	0 0.000	41 1.613	446 2.649	853 2.931	1323 3.122	69 1.839	2732 3.436

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
IF	0	90	126	38	93	0	347
	0.000	1.954	2.100	1.580	1.968	0.000	2.540
IG	34	489	98	195	161	0	977
	1.531	2.689	1.991	2.290	2.207	0.000	2.990
IK	0	3	0	20	19	7	49
	0.000	.477	0.000	1.301	1.279	.845	1.690
IL	32	791	429	201	393	126	1972
	1.505	2.898	2.632	2.303	2.594	2.100	3.295
IM	249	899	280	252	52	19	1751
	2.396	2.954	2.447	2.401	1.716	1.279	3.243
IN	825	1068	538	941	7787	682	11841
	2.916	3.029	2.731	2.974	3.891	2.834	4.073
IO	0	67	24	933	1351	6	2381
	0.000	1.826	1.380	2.970	3.131	.778	3.377
IP	0	16	100	10	4	39	169
	0.000	1.204	2.000	1.000	.602	1.591	2.228
IQ	0	6	0	12	0	0	18
	0.000	.778	0.000	1.079	0.000	0.000	1.255
IR	18	162	80	166	264	21	711
	1.255	2.210	1.903	2.220	2.422	1.322	2.852
IS	42	927	192	220	704	47	2132
	1.623	2.967	2.283	2.342	2.848	1.672	3.329
IT	51	957	976	419	636	235	3274
	1.708	2.981	2.989	2.622	2.803	2.371	3.515
IU	0	0	22	0	48	0	70
	0.000	0.000	1.342	0.000	1.681	0.000	1.845
IV	0	117	291	216	190	0	814
	0.000	2.068	2.464	2.334	2.279	0.000	2.911
IX	0	75	0	14	0	9	98
	0.000	1.875	0.000	1.146	0.000	.954	1.991
IZ	0	22	15	57	79	0	173
	0.000	1.342	1.176	1.756	1.898	0.000	2.238
JA	111	0	3	0	0	0	114
	2.045	0.000	.477	0.000	0.000	0.000	2.057
JE	19	0	76	254	0	0	349
	1.279	0.000	1.881	2.405	0.000	0.000	2.543
JO	130	0	62	0	0	0	192
	2.114	0.000	1.792	0.000	0.000	0.000	2.283
JU	211	0	26	0	0	0	237
	2.324	0.000	1.415	0.000	0.000	0.000	2.375
KA	21	0	4	20	0	0	45
	1.322	0.000	.602	1.301	0.000	0.000	1.653

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
KE	213	0	3	276	394	56	942
	2.328	0.000	.477	2.441	2.595	1.748	2.974
KI	183	43	4	898	0	0	1128
	2.262	1.633	.602	2.953	0.000	0.000	3.052
KL	0	0	0	3	127	0	130
	0.000	0.000	0.000	.477	2.104	0.000	2.114
KM	0	0	0	7	0	0	7
	0.000	0.000	0.000	.845	0.000	0.000	.845
KN	104	0	47	0	0	0	151
	2.017	0.000	1.672	0.000	0.000	0.000	2.179
KO	0	0	0	15	0	0	15
	0.000	0.000	0.000	1.176	0.000	0.000	1.176
KP	0	0	0	16	0	0	16
	0.000	0.000	0.000	1.204	0.000	0.000	1.204
KR	11	0	0	0	0	0	11
	1.041	0.000	0.000	0.000	0.000	0.000	1.041
KS	0	0	0	39	0	81	120
	0.000	0.000	0.000	1.591	0.000	1.908	2.079
KU	0	0	0	0	3	0	3
	0.000	0.000	0.000	0.000	.477	0.000	.477
KW	0	0	11	6	0	0	17
	0.000	0.000	1.041	.778	0.000	0.000	1.230
KY	0	5	0	0	0	0	5
	0.000	.699	0.000	0.000	0.000	0.000	.699
LA	312	890	341	525	465	59	2592
	2.494	2.949	2.533	2.720	2.667	1.771	3.414
LB	0	0	3	0	0	0	3
	0.000	0.000	.477	0.000	0.000	0.000	.477
LC	0	13	73	0	0	0	86
	0.000	1.114	1.863	0.000	0.000	0.000	1.934
LD	0	13	135	71	0	18	237
	0.000	1.114	2.130	1.851	0.000	1.255	2.375
LE	706	429	226	661	1478	993	4493
	2.849	2.632	2.354	2.820	3.170	2.997	3.653
LF	0	0	88	5	0	733	826
	0.000	0.000	1.944	.699	0.000	2.865	2.917
LG	0	0	21	0	9	0	30
	0.000	0.000	1.322	0.000	.954	0.000	1.477
LH	0	0	14	0	0	0	14
	0.000	0.000	1.146	0.000	0.000	0.000	1.146
LI	497	219	378	2380	73	9	3556
	2.696	2.340	2.577	3.377	1.863	.954	3.551

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
LK	0 0.000	0 0.000	154 2.188	7 .845	0 0.000	0 0.000	161 2.207
LL	0 0.000	133 2.124	1528 3.184	209 2.320	569 2.755	110 2.041	2549 3.406
LM	0 0.000	0 0.000	4 .602	0 0.000	0 0.000	7 .845	11 1.041
LN	0 0.000	0 0.000	25 1.398	0 0.000	0 0.000	47 1.672	72 1.857
LO	429 2.632	425 2.628	225 2.352	288 2.459	113 2.053	16 1.204	1496 3.175
LP	0 0.000	0 0.000	76 1.881	0 0.000	0 0.000	0 0.000	76 1.881
LR	0 0.000	273 2.436	0 0.000	0 0.000	35 1.544	0 0.000	308 2.489
LS	0 0.000	0 0.000	13 1.114	0 0.000	20 1.301	484 2.685	517 2.713
LT	0 0.000	22 1.342	107 2.029	133 2.124	292 2.465	33 1.519	587 2.769
LU	39 1.591	87 1.940	92 1.964	214 2.330	27 1.431	0 0.000	459 2.662
LV	0 0.000	4 .602	13 1.114	16 1.204	44 1.643	0 0.000	77 1.886
LW	0 0.000	0 0.000	5 .699	19 1.279	0 0.000	0 0.000	24 1.380
LY	7 .845	16 1.204	5 .699	17 1.230	0 0.000	1924 3.284	1969 3.294
MA	901 2.955	213 2.328	235 2.371	440 2.643	162 2.210	58 1.763	2009 3.303
MB	0 0.000	30 1.477	561 2.749	114 2.057	7 .845	0 0.000	712 2.852
ME	1222 3.087	273 2.436	251 2.400	241 2.382	442 2.645	246 2.391	2675 3.427
MF	0 0.000	0 0.000	43 1.633	4 .602	0 0.000	0 0.000	47 1.672
MI	833 2.921	78 1.892	306 2.486	297 2.473	35 1.544	0 0.000	1549 3.190
ML	0 0.000	0 0.000	5 .699	11 1.041	0 0.000	0 0.000	16 1.204
MM	0 0.000	19 1.279	177 2.248	38 1.580	25 1.398	0 0.000	259 2.413
MN	0 0.000	0 0.000	0 0.000	7 .845	36 1.556	13 1.114	56 1.748

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
MO	471 2.673	90 1.954	136 2.134	145 2.161	0 0.000	0 0.000	0 2.925
MP	0 0.000	197 2.294	602 2.780	348 2.542	117 2.068	0 0.000	1264 3.102
MR	0 0.000	0 0.000	4 .602	0 0.000	0 0.000	0 0.000	0 .602
MS	0 0.000	0 0.000	603 2.780	8 .903	0 0.000	227 2.356	838 2.923
MU	180 2.255	17 1.230	0 0.000	77 1.886	159 2.201	0 0.000	433 2.636
MY	42 1.623	0 0.000	0 0.000	0 0.000	0 0.000	112 2.049	154 2.188
NA	358 2.554	131 2.117	476 2.678	81 1.908	92 1.964	38 1.580	1176 3.070
NB	0 0.000	3 .477	0 0.000	13 1.114	0 0.000	0 0.000	0 1.204
NC	0 0.000	195 2.290	448 2.651	228 2.358	509 2.707	0 0.000	1380 3.140
ND	0 0.000	127 2.104	596 2.775	244 2.387	450 2.653	530 2.724	1947 3.289
NE	394 2.595	38 1.580	718 2.856	424 2.627	638 2.805	502 2.701	2714 3.434
NF	0 0.000	36 1.556	55 1.740	25 1.398	0 0.000	0 0.000	0 2.064
NG	0 0.000	498 2.697	329 2.517	309 2.490	230 2.362	7212 3.858	8578 3.933
NH	0 0.000	50 1.699	6 .778	0 0.000	0 0.000	0 0.000	0 1.748
NI	3 .477	220 2.342	193 2.286	1231 3.090	87 1.940	9 .954	1743 3.241
NJ	0 0.000	77 1.886	0 0.000	0 0.000	0 0.000	0 0.000	0 1.886
NK	0 0.000	47 1.672	115 2.061	73 1.863	0 0.000	0 0.000	0 2.371
NL	0 0.000	7 .845	3 .477	35 1.544	21 1.322	0 0.000	0 1.820
NM	0 0.000	9 .954	4 .602	0 0.000	0 0.000	0 0.000	0 1.114
NN	0 0.000	15 1.176	373 2.572	149 2.173	32 1.505	0 0.000	569 2.755
NO	585 2.767	782 2.893	150 2.176	142 2.152	10 1.000	10 1.000	1679 3.225

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
NQ	0	28	10	0	0	0	38
	0.000	1.447	1.000	0.000	0.000	0.000	1.580
NR	0	3	10	0	6	0	19
	0.000	.477	1.000	0.000	.778	0.000	1.279
NS	0	357	238	89	946	1197	2827
	0.000	2.553	2.377	1.949	2.976	3.078	3.451
NT	0	242	1110	732	332	1653	4069
	0.000	2.384	3.045	2.865	2.521	3.218	3.609
NU	288	69	293	10	42	0	702
	2.459	1.839	2.467	1.000	1.623	0.000	2.846
NV	0	88	29	0	0	0	117
	0.000	1.944	1.462	0.000	0.000	0.000	2.068
NX	0	71	0	0	0	5	76
	0.000	1.851	0.000	0.000	0.000	.699	1.881
NY	0	46	10	0	3	467	526
	0.000	1.663	1.000	0.000	.477	2.669	2.721
NZ	0	11	0	0	3	0	14
	0.000	1.041	0.000	0.000	.477	0.000	1.146
OA	0	130	47	12	0	0	189
	0.000	2.114	1.672	1.079	0.000	0.000	2.276
OB	229	87	318	51	0	0	685
	2.360	1.940	2.502	1.708	0.000	0.000	2.836
OC	51	460	295	5	8	0	819
	1.708	2.663	2.470	.699	.903	0.000	2.913
OD	0	43	185	14	249	25	516
	0.000	1.633	2.267	1.146	2.396	1.398	2.713
OE	0	0	22	0	58	0	80
	0.000	0.000	1.342	0.000	1.763	0.000	1.903
OF	237	0	48	7	0	16	308
	2.375	0.000	1.681	.845	0.000	1.204	2.489
OG	0	39	399	9	29	0	476
	0.000	1.591	2.601	.954	1.462	0.000	2.678
OH	0	34	0	13	0	0	47
	0.000	1.531	0.000	1.114	0.000	0.000	1.672
OI	0	115	37	27	80	0	259
	0.000	2.061	1.568	1.431	1.903	0.000	2.413
OJ	0	5	93	0	0	0	98
	0.000	.699	1.968	0.000	0.000	0.000	1.991
OK	0	0	230	24	3	36	293
	0.000	0.000	2.362	1.380	.477	1.556	2.467
OL	7	888	162	70	329	264	1720
	.845	2.948	2.210	1.845	2.517	2.422	3.236

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
OM	25 1.398	942 2.974	133 2.124	125 2.097	194 2.288	229 2.360	1648 3.217
ON	5 .699	1044 3.019	143 2.155	150 2.176	925 2.966	1664 3.221	3931 3.595
OO	0 0.000	356 2.551	55 1.740	208 2.318	153 2.185	0 0.000	772 2.888
OP	331 2.520	140 2.146	379 2.579	21 1.322	6 .778	89 1.949	966 2.985
OQ	0 0.000	0 0.000	0 0.000	11 1.041	0 0.000	0 0.000	11 1.041
OR	200 2.301	1337 3.126	274 2.438	526 2.721	1105 3.043	195 2.290	3637 3.561
OS	0 0.000	99 1.996	187 2.272	115 2.061	278 2.444	0 0.000	679 2.832
OT	0 0.000	704 2.848	1023 3.010	82 1.914	57 1.756	13 1.114	1879 3.274
OU	343 2.535	1348 3.130	1108 3.045	1101 3.042	1261 3.101	0 0.000	5161 3.713
OV	24 1.380	164 2.215	238 2.377	229 2.360	53 1.724	0 0.000	708 2.850
OW	0 0.000	681 2.833	413 2.616	98 1.991	245 2.389	112 2.049	1549 3.190
OX	0 0.000	0 0.000	0 0.000	0 0.000	0 0.000	.954 .954	9 9
OY	8 .903	51 1.708	3 .477	64 1.806	13 1.114	48 1.681	187 2.272
PA	949 2.977	59 1.771	79 1.898	403 2.605	0 0.000	0 0.000	1490 3.173
PD	0 0.000	3 .477	0 0.000	0 0.000	0 0.000	0 0.000	3 .477
PE	810 2.908	508 2.706	379 2.579	263 2.420	571 2.757	0 0.000	2531 3.403
PF	0 0.000	0 0.000	0 0.000	29 1.462	0 0.000	0 0.000	29 1.462
PG	0 0.000	3 .477	0 0.000	0 0.000	0 0.000	0 0.000	3 .477
PH	57 1.756	4 .602	3 .477	40 1.602	0 0.000	22 1.342	126 2.100
PI	317 2.501	160 2.204	265 2.423	382 2.582	23 1.362	0 0.000	1147 3.060
PK	0 0.000	0 0.000	7 .845	0 0.000	0 0.000	0 0.000	7 .845

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
PL	468	0	361	289	349	0	1467
	2.670	0.000	2.558	2.461	2.543	0.000	3.166
PM	0	0	0	6	0	0	6
	0.000	0.000	0.000	.778	0.000	0.000	.778
PO	412	51	223	585	0	0	1271
	2.615	1.708	2.348	2.767	0.000	0.000	3.104
PP	0	294	397	437	26	0	1154
	0.000	2.468	2.599	2.640	1.415	0.000	3.062
PR	2730	51	166	0	0	0	2947
	3.436	1.708	2.220	0.000	0.000	0.000	3.469
PS	3	3	5	15	19	313	358
	.477	.477	.699	1.176	1.279	2.496	2.554
PT	0	63	141	160	6	203	573
	0.000	1.799	2.149	2.204	.778	2.307	2.758
PU	321	6	128	46	0	0	501
	2.507	.778	2.107	1.663	0.000	0.000	2.700
PW	0	6	0	0	0	0	6
	0.000	.778	0.000	0.000	0.000	0.000	.778
PY	0	0	0	0	0	38	38
	0.000	0.000	0.000	0.000	0.000	1.580	1.580
QU	358	113	199	26	23	0	719
	2.554	2.053	2.299	1.415	1.362	0.000	2.857
RA	311	774	280	456	2059	32	3912
	2.493	2.889	2.447	2.659	3.314	1.505	3.592
RB	0	5	50	8	18	15	96
	0.000	.699	1.699	.903	1.255	1.176	1.982
RC	0	8	173	99	41	0	321
	0.000	.903	2.238	1.996	1.613	0.000	2.507
RD	0	93	102	37	190	317	739
	0.000	1.968	2.009	1.568	2.279	2.501	2.869
RE	2781	1463	870	661	1219	1087	8081
	3.444	3.165	2.940	2.820	3.086	3.036	3.907
RF	0	0	340	13	0	0	353
	0.000	0.000	2.531	1.114	0.000	0.000	2.548
RG	0	63	212	183	18	0	476
	0.000	1.799	2.326	2.262	1.255	0.000	2.678
RH	12	0	307	0	0	0	319
	1.079	0.000	2.487	0.000	0.000	0.000	2.504
RI	118	1457	858	1205	74	0	3712
	2.072	3.163	2.933	3.081	1.869	0.000	3.570
RK	0	0	350	7	53	39	449
	0.000	0.000	2.544	.845	1.724	1.591	2.652

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
RL	0	40	199	183	219	0	641
	0.000	1.602	2.299	2.262	2.340	0.000	2.807
RM	0	3	372	90	20	109	594
	0.000	.477	2.571	1.954	1.301	2.037	2.774
RN	0	0	494	222	45	450	1211
	0.000	0.000	2.694	2.346	1.653	2.653	3.083
RO	247	2640	1047	289	302	4	4529
	2.393	3.422	3.020	2.461	2.480	.602	3.656
RP	0	0	207	41	6	0	254
	0.000	0.000	2.316	1.613	.778	0.000	2.405
RQ	0	0	4	0	0	0	4
	0.000	0.000	.602	0.000	0.000	0.000	.602
RR	0	111	630	52	7	0	800
	0.000	2.045	2.799	1.716	.845	0.000	2.903
RS	0	3	459	86	53	2299	2900
	0.000	.477	2.662	1.934	1.724	3.362	3.462
RT	0	123	842	373	388	356	2082
	0.000	2.090	2.925	2.572	2.589	2.551	3.318
RU	257	182	5	23	5	0	472
	2.410	2.260	.699	1.362	.699	0.000	2.674
RV	0	0	502	3	74	0	579
	0.000	0.000	2.701	.477	1.869	0.000	2.763
RW	0	0	142	15	0	0	157
	0.000	0.000	2.152	1.176	0.000	0.000	2.196
RY	0	26	42	0	0	1485	1553
	0.000	1.415	1.623	0.000	0.000	3.172	3.191
SA	163	0	15	128	18	5	329
	2.212	0.000	1.176	2.107	1.255	.699	2.517
SB	0	0	131	0	0	0	131
	0.000	0.000	2.117	0.000	0.000	0.000	2.117
SC	454	22	104	0	0	0	580
	2.657	1.342	2.017	0.000	0.000	0.000	2.763
SD	0	0	3	59	0	0	62
	0.000	0.000	.477	1.771	0.000	0.000	1.792
SE	1655	17	241	1283	777	1832	5805
	3.219	1.230	2.382	3.108	2.890	3.263	3.764
SF	0	0	0	0	16	0	16
	0.000	0.000	0.000	0.000	1.204	0.000	1.204
SH	554	23	204	189	0	455	1425
	2.744	1.362	2.310	2.276	0.000	2.658	3.154
SI	569	0	416	1154	68	0	2207
	2.755	0.000	2.619	3.062	1.833	0.000	3.344

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
SK	48	0	3	22	0	6	79
	1.681	0.000	.477	1.342	0.000	.778	1.898
SL	186	30	14	0	4	0	234
	2.270	1.477	1.146	0.000	.602	0.000	2.369
SM	160	0	5	4	0	30	199
	2.204	0.000	.699	.602	0.000	1.477	2.299
SN	56	0	0	0	0	0	56
	1.748	0.000	0.000	0.000	0.000	0.000	1.748
SO	665	8	29	290	170	17	1179
	2.823	.903	1.462	2.462	2.230	1.230	3.072
SP	586	70	415	29	0	0	1100
	2.768	1.845	2.618	1.462	0.000	0.000	3.041
SQ	29	3	0	0	0	0	32
	1.462	.477	0.000	0.000	0.000	0.000	1.505
SR	0	0	5	0	0	0	5
	0.000	0.000	.699	0.000	0.000	0.000	.699
SS	0	168	666	326	31	644	1835
	0.000	2.225	2.823	2.513	1.491	2.809	3.264
ST	1586	8	1550	338	140	1204	4826
	3.200	.903	3.190	2.529	2.146	3.081	3.684
SU	1124	206	317	233	0	0	1880
	3.051	2.314	2.501	2.367	0.000	0.000	3.274
SW	84	0	44	0	0	0	128
	1.924	0.000	1.643	0.000	0.000	0.000	2.107
SY	173	3	0	0	0	40	216
	2.238	.477	0.000	0.000	0.000	1.602	2.334
TA	203	584	214	826	167	35	2029
	2.307	2.766	2.330	2.917	2.223	1.544	3.307
TB	0	0	3	0	0	0	3
	0.000	0.000	.477	0.000	0.000	0.000	.477
TC	0	4	312	0	35	0	351
	0.000	.602	2.494	0.000	1.544	0.000	2.545
TD	0	0	27	0	0	0	27
	0.000	0.000	1.431	0.000	0.000	0.000	1.431
TE	374	161	426	1412	2777	595	5745
	2.573	2.207	2.629	3.150	3.444	2.775	3.759
TF	0	0	3	3	0	0	6
	0.000	0.000	.477	.477	0.000	0.000	.778
TG	0	0	4	8	0	0	12
	0.000	0.000	.602	.903	0.000	0.000	1.079
TH	1813	38	1242	1675	96	115	4979
	3.258	1.580	3.094	3.224	1.982	2.061	3.697

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
TI	62 1.792	68 1.833	773 2.888	2558 3.408	97 1.987	0 0.000	3558 3.551
TL	0 0.000	35 1.544	57 1.756	108 2.033	378 2.577	0 0.000	578 2.762
TN	0 0.000	0 0.000	36 1.556	46 1.663	0 0.000	0 0.000	82 1.914
TO	283 2.452	297 2.473	64 1.806	633 2.801	246 2.391	17 1.230	1540 3.188
TP	0 0.000	0 0.000	8 .903	0 0.000	0 0.000	0 0.000	8 .903
TR	693 2.841	278 2.444	176 2.246	484 2.685	390 2.591	0 0.000	2021 3.306
TS	7 .845	0 0.000	210 2.322	0 0.000	0 0.000	1678 3.225	1895 3.278
TT	0 0.000	177 2.248	1007 3.003	244 2.387	19 1.279	16 1.204	1463 3.165
TU	192 2.283	331 2.520	203 2.307	747 2.873	10 1.000	0 0.000	1483 3.171
TW	28 1.447	0 0.000	773 2.888	11 1.041	0 0.000	0 0.000	812 2.910
TY	76 1.881	3 .477	0 0.000	0 0.000	0 0.000	1010 3.004	1089 3.037
UA	0 0.000	204 2.310	294 2.468	57 1.756	102 2.009	0 0.000	657 2.818
UB	0 0.000	233 2.367	29 1.462	142 2.152	0 0.000	4 .602	408 2.611
UC	0 0.000	269 2.430	112 2.049	105 2.021	224 2.350	0 0.000	710 2.851
UD	0 0.000	44 1.643	337 2.528	15 1.176	134 2.127	5 .699	535 2.728
UE	0 0.000	83 1.919	42 1.623	71 1.851	107 2.029	77 1.886	380 2.580
UF	0 0.000	40 1.602	8 .903	0 0.000	0 0.000	0 0.000	48 1.681
UG	0 0.000	71 1.851	127 2.104	783 2.894	974 2.989	0 0.000	1955 3.291
UI	0 0.000	217 2.336	19 1.279	214 2.330	57 1.756	0 0.000	507 2.705
UK	0 0.000	4 .602	0 0.000	0 0.000	0 0.000	0 0.000	4 .602
UL	0 0.000	138 2.140	28 1.447	490 2.690	95 1.978	188 2.274	939 2.973

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION	1ST	2ND	3RD	4TH	5TH	6TH	SUM
UM	Ø	216	29	212	30	207	694	
	Ø.000	2.334	1.462	2.326	1.477	2.316	2.841	
UN	247	481	812	120	162	23	1845	
	2.393	2.682	2.910	2.079	2.210	1.362	3.266	
UO	Ø	3	Ø	10	Ø	Ø	13	
	Ø.000	.477	Ø.000	1.000	Ø.000	Ø.000	1.114	
UP	29	337	34	5	13	6	424	
	1.462	2.528	1.531	.699	1.114	.778	2.627	
UR	12	1185	313	444	936	39	2929	
	1.079	3.074	2.496	2.647	2.971	1.591	3.467	
US	223	631	291	156	933	664	2898	
	2.348	2.800	2.464	2.193	2.970	2.822	3.462	
UT	92	508	98	258	95	627	1678	
	1.964	2.706	1.991	2.412	1.978	2.797	3.225	
UZ	Ø	29	Ø	Ø	Ø	Ø	29	
	Ø.000	1.462	Ø.000	Ø.000	Ø.000	Ø.000	1.462	
VA	374	5	130	278	80	Ø	867	
	2.573	.699	2.114	2.444	1.903	Ø.000	2.938	
VE	196	298	790	245	1026	629	3184	
	2.292	2.474	2.898	2.389	3.011	2.799	3.503	
VI	409	56	358	892	Ø	Ø	1715	
	2.612	1.748	2.554	2.950	Ø.000	Ø.000	3.234	
VO	60	27	143	27	Ø	Ø	257	
	1.778	1.431	2.155	1.431	Ø.000	Ø.000	2.410	
VU	4	Ø	Ø	Ø	Ø	Ø	4	
	.602	Ø.000	Ø.000	Ø.000	Ø.000	Ø.000	.602	
WA	498	44	77	169	136	Ø	924	
	2.697	1.643	1.886	2.228	2.134	Ø.000	2.966	
WB	Ø	Ø	10	Ø	Ø	Ø	10	
	Ø.000	Ø.000	1.000	Ø.000	Ø.000	Ø.000	1.000	
WD	Ø	Ø	11	32	Ø	Ø	43	
	Ø.000	Ø.000	1.041	1.505	Ø.000	Ø.000	1.633	
WE	559	32	576	867	129	Ø	2163	
	2.747	1.505	2.760	2.938	2.111	Ø.000	3.335	
WF	Ø	10	Ø	Ø	Ø	Ø	10	
	Ø.000	1.000	Ø.000	Ø.000	Ø.000	Ø.000	1.000	
WH	446	Ø	29	Ø	Ø	Ø	475	
	2.649	Ø.000	1.462	Ø.000	Ø.000	Ø.000	2.677	
WI	1018	43	Ø	377	Ø	Ø	1438	
	3.008	1.633	Ø.000	2.576	Ø.000	Ø.000	3.158	
WK	Ø	11	Ø	Ø	Ø	Ø	11	
	Ø.000	1.041	Ø.000	Ø.000	Ø.000	Ø.000	1.041	

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
WL	0	0	3	31	0	0	34
	0.000	0.000	.477	1.491	0.000	0.000	1.531
WM	0	0	0	3	0	0	3
	0.000	0.000	0.000	.477	0.000	0.000	.477
WN	0	0	0	21	4	56	81
	0.000	0.000	0.000	1.322	.602	1.748	1.908
WO	389	21	0	44	0	0	454
	2.590	1.322	0.000	1.643	0.000	0.000	2.657
WP	0	0	27	0	0	0	27
	0.000	0.000	1.431	0.000	0.000	0.000	1.431
WR	370	0	4	0	0	0	374
	2.568	0.000	.602	0.000	0.000	0.000	2.573
WS	0	0	4	0	0	198	202
	0.000	0.000	.602	0.000	0.000	2.297	2.305
WY	9	0	23	0	0	0	32
	.954	0.000	1.362	0.000	0.000	0.000	1.505
XA	0	435	8	0	0	0	443
	0.000	2.638	.903	0.000	0.000	0.000	2.646
XC	0	49	0	0	0	0	49
	0.000	1.690	0.000	0.000	0.000	0.000	1.690
XE	0	20	0	0	31	0	51
	0.000	1.301	0.000	0.000	1.491	0.000	1.708
XH	0	32	0	0	0	0	32
	0.000	1.505	0.000	0.000	0.000	0.000	1.505
XI	0	40	177	0	0	0	217
	0.000	1.602	2.248	0.000	0.000	0.000	2.336
XP	0	298	0	0	0	0	298
	0.000	2.474	0.000	0.000	0.000	0.000	2.474
XT	0	80	118	0	0	38	236
	0.000	1.903	2.072	0.000	0.000	1.580	2.373
XW	0	0	11	0	0	0	11
	0.000	0.000	1.041	0.000	0.000	0.000	1.041
YA	28	0	32	0	0	0	60
	1.447	0.000	1.505	0.000	0.000	0.000	1.778
YB	0	0	81	0	0	0	81
	0.000	0.000	1.908	0.000	0.000	0.000	1.908
YC	0	0	11	5	0	0	16
	0.000	0.000	1.041	.699	0.000	0.000	1.204
YE	9	11	4	64	99	6	193
	.954	1.041	.602	1.806	1.996	.778	2.286
YF	0	0	0	3	0	0	3
	0.000	0.000	0.000	.477	0.000	0.000	.477

BIGRAM FREQUENCIES FOR WORDS OF LETTER LENGTH 7

	BIGRAM POSITION						
	1ST	2ND	3RD	4TH	5TH	6TH	SUM
YG	0	3	0	0	0	0	3
	0.000	.477	0.000	0.000	0.000	0.000	.477
YH	0	0	8	0	0	0	8
	0.000	0.000	.903	0.000	0.000	0.000	.903
YI	12	0	0	194	0	0	206
	1.079	0.000	0.000	2.288	0.000	0.000	2.314
YL	0	0	12	0	0	0	12
	0.000	0.000	1.079	0.000	0.000	0.000	1.079
YM	0	41	81	11	0	3	136
	0.000	1.613	1.908	1.041	0.000	.477	2.134
YN	0	33	13	0	0	0	46
	0.000	1.519	1.114	0.000	0.000	0.000	1.663
YO	44	9	0	0	0	0	53
	1.643	.954	0.000	0.000	0.000	0.000	1.724
YP	0	72	3	0	0	0	75
	0.000	1.857	.477	0.000	0.000	0.000	1.875
YR	0	18	54	0	0	0	72
	0.000	1.255	1.732	0.000	0.000	0.000	1.857
YS	0	179	45	0	7	57	288
	0.000	2.253	1.653	0.000	.845	1.756	2.459
YT	0	0	31	27	0	0	58
	0.000	0.000	1.491	1.431	0.000	0.000	1.763
YW	0	0	17	0	0	0	17
	0.000	0.000	1.230	0.000	0.000	0.000	1.230
YZ	0	0	0	0	10	0	10
	0.000	0.000	0.000	0.000	1.000	0.000	1.000
ZA	0	0	32	0	0	3	35
	0.000	0.000	1.505	0.000	0.000	.477	1.544
ZE	7	0	10	0	33	100	150
	.845	0.000	1.000	0.000	1.519	2.000	2.176
ZI	0	0	0	39	0	0	39
	0.000	0.000	0.000	1.591	0.000	0.000	1.591
ZL	0	0	0	23	0	0	23
	0.000	0.000	0.000	1.362	0.000	0.000	1.362
ZO	0	0	0	9	27	0	36
	0.000	0.000	0.000	.954	1.431	0.000	1.556
ZU	0	0	0	6	0	0	6
	0.000	0.000	0.000	.778	0.000	0.000	.778
ZY	0	0	11	0	0	0	11
	0.000	0.000	1.041	0.000	0.000	0.000	1.041
ZZ	0	0	29	0	0	0	29
	0.000	0.000	1.462	0.000	0.000	0.000	1.462

This page intentionally left blank

6 Summary and Conclusion

The goal of this research project was to evaluate how the reader's higher-order knowledge about orthographic structure interacts with featural information during the processing of letters and words. The research was carried out utilizing a general information processing model, which provided a formal framework from which hypotheses could be derived and tested. The model along with supporting evidence was described in Chapter 1. Since the study of visual processing in reading requires an examination of the visual features that are functional during reading, the discussion in Chapter 2 described those features utilized in the recognition of lowercase letters. Understanding feature detection proved useful in accounting for some of the findings in the perceptual recognition tasks described in Chapter 4. The significant finding was that the search for a target letter was critically dependent on the featural similarity between the target letter and the letters in the test string.

The fundamental properties of the orthographic structure of English were described in Chapter 3. We distinguished two broad categories of descriptions. The first was derived from statistical-redundancy measures based on the frequency with which letters, letter sequences, and words occur in natural texts. The second category, rule-governed regularity, was based on phonological constraints of English and scribal conventions governing the spelling of English. Also included in Chapter 3 was a brief review of previous studies of orthography.

In Chapter 4, we examined the psychological reality of various descriptions of orthographic structure in perceptual recognition tasks and overt judgment tasks. The recognition

tasks assessed the degree to which knowledge of orthographic structure is utilized in word perception, and the overt judgment tasks assessed the degree to which this knowledge is consciously available.

Given our previous results (Massaro et al., 1979), we had reason to suspect differences between reaction time tasks and accuracy tasks in terms of their sensitivities to the utilization of orthographic structure. We began with a direct comparison between these two tasks, while simultaneously contrasting a rule-governed measure and a statistical-redundancy measure of orthographic structure. Replicating our previous work, we found highly significant effects of orthographic structure in the accuracy task and relatively small, and sometimes nonsignificant, effects in the reaction time task. The next six experiments assessed what aspects of the two tasks accounted for the differences that were observed. We evaluated the influences of presenting the target letter before or after the test display, the contribution of a feature detection strategy, the role of limited stimulus information, and the sensitivity of speeded reaction-time experiments. The reason that the magnitude of the effect appears smaller for a high accuracy reaction-time measure is that the differences in visual processing time are small relative to the overall reaction time to complete the task. In contrast, the stimulus parameters are modified in the accuracy task to keep average performance at 75% correct; any differences in accuracy are then influenced almost completely by visual processing. Supporting this interpretation, the post hoc evaluation of various measures of orthographic structure revealed reasonably good correlations with reaction time but much higher correlations with accuracy.

To provide a quantitative description of task accuracy, we formalized the information processing model developed in Chapter 1. This model allows a formal description of the facilitative effect of orthographic structure on both reaction time and task accuracy. The basic assumption of the model is that knowledge of orthographic structure contributes an independent source of information about the letter string. Consequently, fewer visual features are necessary to resolve well-structured strings than to resolve poorly-structured strings. The model was applied to the target search task by formalizing a decision algorithm assumed to be used by the subject when faced with partial information. The model provided a good quantitative description of the results from the accuracy task. The parameters of the model were

psychologically meaningful and the parameter values corresponding to the number of letters seen in the test strings provided a quantitative measure of the contribution of orthographic structure. According to the model, our readers were able to recognize two additional letters in well-structured strings compared to poorly-structured strings. This is a substantial effect considering that two letters represent one-third of the six-letter test string.

We also carried out a series of overt judgment experiments to assess which descriptions of orthographic structure are consciously available. We asked whether subjects could reliably assign ratings of how much a letter string resembles written English on the basis of the typicality of the letter sequencing in the test strings. In addition, other subjects were asked to discriminate the items on the basis of rule-governed regularity or on the basis of statistical redundancy. Subjects' judgments were more accurately described by rule-governed regularity than by statistical redundancy. Finally, the ratings of the individual items correlated quite strongly with their accuracies and reaction times in the target search task.

The factorial design of our experiments contrasted just one measure of rule-governed regularity with one measure of statistical redundancy. Therefore, we conducted a large number of post hoc correlational analyses to evaluate a wide range of measures of orthographic structure. The dependent measures were the reaction times, accuracy, and rating responses to each of 200 test items. Position-sensitive log bigram frequency provided the best description of performance on the individual test items. However, our rule-based regularity measure also provided a very good description. In fact, our regularity measure correlated very highly with the best frequency-based measure. Therefore, we were not successful in choosing between rule-governed regularity and statistical redundancy in these experiments.

We were also able to evaluate the most appropriate measures for a statistical-redundancy description. We obtained information about the utility of type versus token counts, the appropriate scale for describing frequency of occurrence, position-sensitive versus position-insensitive counts of letter occurrences, the role of word frequency, and the relative contributions of various serial positions in determining the orthographic structure of an item.

What conclusions can be reached and what are the implications for future research and classroom

practice? Although we were not successful in choosing between rule-governed regularity and statistical redundancy in these experiments, additional experiments might be successful. Given the best statistical-redundancy measure, it should be possible to develop a new set of test items to contrast this measure with a rule-governed regularity measure. We plan to follow the logic of the present studies by contrasting log bigram frequency and regularity measures in target search and overt judgment tasks. Although these two measures are highly correlated, it should be possible to distinguish between them. Given the results, it will still be necessary to carry out post hoc correlations to determine if some other measure might provide even a better description. This research has the promise of leading to the description that best reflects the knowledge of orthographic structure used in the course of normal reading.

Arriving at a good description of orthographic structure should be beneficial in the development of reading instruction. Given that the utilization of orthographic structure is an important component in letter and word recognition, instructional materials should be designed to facilitate children's understanding of this structure. Current phonics programs address only the correspondences between spelling and sound, not the constraints in English orthography. Massaro and Taylor (1979) sketch a series of classroom games that might be employed to teach children common letter patterns and the locations in words these letter patterns are most likely to be found. Delineating the best description of orthographic structure might facilitate the teaching and learning of this structure.

References

- Allport, A. Word recognition in reading (Tutorial paper). In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible language I. New York: Plenum, 1979.
- Anderson, I. H., & Dearborn, W. F. The psychology of teaching reading. New York: Ronald Press, 1952.
- Anisfeld, M. A. A comment on "The role of grapheme-phoneme correspondence in the perception of words." American Journal of Psychology, 1964, 77, 320-326.
- Atkinson, R. C., & Juola, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology. San Francisco: Freeman, 1974.
- Baddeley, A. D. Immediate memory and the "perception" of letter sequences. Quarterly Journal of Experimental Psychology, 1964, 16, 364-367.
- Baron, J. The word superiority effect. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 6). Hillsdale, N.J.: Erlbaum, 1978.
- Baron, J., & Thurston, I. An analysis of the word-superiority effect. Cognitive Psychology, 1973, 4, 207-228.

- Bjork, E. L., & Estes, W. K. Letter identification in relation to linguistic context and masking conditions. Memory and Cognition, 1973, 1, 217-223.
- Bouma, H. Visual recognition of isolated lower-case letters. Vision Research, 1971, 11, 459-474.
- Bouwhuis, D. G. Visual recognition of words. Unpublished doctoral dissertation, Katholieke Universiteit, Nijmegen, Holland, 1979.
- Breitmeyer, B. G., & Ganz, L. Implications of sustained and transient channels for theories of visual pattern masking, saccadic suppression and information processing. Psychological Review, 1976, 83, 1-36.
- Broadbent, D. E. Word-frequency effect and response bias. Psychological Review, 1967, 74, 1-15.
- Broadbent, D. E. The hidden preattentive processes. American Psychologist, 1977, 32, 109-118.
- Cattell, J. M. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. Philosophische Studien, 1885, 2, 635-650.
- Chandler, J. P. Subroutine STEPIT--Finds local minima of a smooth function of several parameters. Behavioral Science, 1969, 14, 81-82.
- Chomsky, N., & Halle, M. The sound pattern of English. New York: Harper & Row, 1968.
- Clark, H. H. The language-as-fixed-effect-fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.
- Estes, W. K. The locus of inferential and perceptual processes in letter identification. Journal of Experimental Psychology: General, 1975, 104, 122-145. (a)

- Estes, W. K. Memory, perception, and decision in letter identification. In R. L. Solso (Ed.), Information procesing and cognition: The Loyola Symposium. Hillsdale, N.J.: Erlbaum, 1975. (b)
- Estes, W. K. On the interaction of perception and memory in reading. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and Comprehension. Hillsdale, N.J.: Erlbaum, 1977.
- Estes, W. K. Perceptual processing in letter recognition and reading. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception: Perceptual processing (Vol. 9). New York: Academic press, 1978.
- Estes, W. K., Bjork, E. L., & Skaar, E. Detection of single letters and letters in words with changing versus unchanging mask characters. Bulletin of the Psychonomic Society, 1974, 3, 201-203.
- Geoffrion, L. D. Positional uncertainty in lateral masking and the perceptual superiority of words. Perception and Psychophysics, 1976, 19(3), 273-278.
- Gibson, E. J. On the perception of words. American Journal of Psychology, 1964, 77, 667-669.
- Gibson, E. J., & Levin, H. The psychology of reading. Cambridge: MIT Press, 1975.
- Gibson, E. J., Osser, H., Schiff, W., & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In Final report on a basic research program on reading (Cooperative Research Project No. 639). Ithaca, N. Y.: Cornell University and U.S. Office of Education, 1963.
- Gibson, E. J., Pick, A., Osser, H., & Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. American Journal of Psychology, 1962, 75, 554-570.

- Gibson, E. J., Shurcliff, A., & Yonas, A. A utilization of spelling patterns used by deaf and hearing subjects. In H. Levin and J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.
- Gilmore, G. C., & Egeth, H. E. When are nonwords easy to see? Memory and Cognition, 1976, 4, 519-524.
- Green, D. M., & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.
- Henderson, L. Word recognition. In N. S. Sutherland (Ed.), Tutorial essays in psychology (Vol. 1). Hillsdale, N.J.: Erlbaum, 1977.
- Huey, E. B. The psychology and pedagogy of reading. Cambridge, Mass.: MIT Press, 1968. (Originally published, New York: Macmillan, 1908.)
- Jakobson, R., Fant, C. G. M., & Halle, M. Preliminaries to speech analysis: The distinctive features and their correlates. Cambridge, Mass.: MIT Press, 1961.
- James, W. Principles of psychology. New York: Dover Paperback, 1950. (Originally published, Holt, 1890.)
- Johnson, N. F. A pattern-unit model of word identification. In D. LaBerge & S. J. Samuels (Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Erlbaum, 1977.
- Johnston, J. C., & McClelland, J. L. Perception of letters in words: Seek not and ye shall find. Science, 1974, 184, 1192-1194.
- Kallman, H. J., & Massaro, D. W. Similarity effects in backward recognition masking. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 110-128.

- Kleiman, G. M. Speech recoding in reading. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 323-340.
- Kolers, P. A. Three stages of reading. In H. Levin & J. P. Williams (Eds.), Basic studies on reading. New York: Basic Books, 1970.
- Krueger, L. E. Familiarity effects in visual information processing. Psychological Bulletin, 1975, 82, 949-974.
- Krueger, L. E. Features versus redundancy: Comments on Massaro, Venezky, and Taylor's "Orthographic regularity, positional frequency, and visual processing of letter strings." Journal of Experimental Psychology: General, 1979, 108, 125-130.
- Krueger, L. E., & Shapiro, R. G. Letter detection with rapid serial visual presentation: Evidence against word superiority at feature extraction. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 657-673.
- Kucera, H., & Francis, W. N. Computational analysis of present-day American English. Providence: Brown University Press, 1967.
- Kurath, H. A phonology and prosody of Modern English. Ann Arbor, Michigan: University of Michigan Press, 1964.
- Landauer, T. K., & Streeter, L. A. Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 119-131.
- Lefton, L. A., Spragins, A. B., & Byrnes, J. English orthography: Relation to reading experience. Bulletin of the Psychonomic Society, 1973, 2, 281-282.
- Levy, B. A. Speech analysis during sentence processing: Reading and listening. Visible Language, 1978, 12, 81-101.

- Mason, M. Reading ability and letter search time: Effects of orthographic structure defined by single-letter positional frequency. Journal of Experimental Psychology: General, 1975, 104, 146-166.
- Massaro, D. W. Perception of letters, words, and nonwords. Journal of Experimental Psychology, 1973, 100, 349-353.
- Massaro, D. W. Primary and secondary recognition in reading. In D. W. Massaro (Ed.), Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975.
- Massaro, D. W. Reading and listening (Technical Report No. 423). Madison: Wisconsin Research and Development Center for Individualized Schooling, December 1977.
- Massaro, D. W. A stage model of reading and listening. Visible Language, 1978, 12, 3-26.
- Massaro, D. W. Letter information and orthographic context in word perception. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 595-609. (a)
- Massaro, D. W. Reading and listening (Tutorial paper). In P. A. Kolers, M. Wrolstad, & H. Bouma. (Eds.), Processing of visible language I. New York: Plenum, 1979 (b).
- Massaro, D. W. How does orthographic structure facilitate reading? In R. L. Venezky & J. F. Kavanagh (Eds.), Orthography, reading, and dyslexia. Baltimore: University Park Press, in press.
- Massaro, D. W., Jones, R. D., Lipscomb, C., & Scholz, R. Role of prior knowledge on naming and lexical decisions with good and poor stimulus information. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 498-512.

- Massaro, D. W., & Schmuller, J. Visual features, perceptual storage, and processing time in reading. In D. W. Massaro (Ed.), Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975.
- Massaro, D. W., & Taylor, G. A. Reading ability and the utilization of orthographic structure in reading. (Technical Report No. 515). Madison: Wisconsin Research and Development Center for Individualized Schooling, July 1979.
- 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. Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. Psychonomic Monograph Supplements, 1965, 1, 13-32.
- 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, 79-143. (a)
- Mayzner, M. S., Tresselt, M. E., & Wolin, B. R. Tables of trigram frequency counts for various word-length and letter-position combinations. Psychonomic Monographic Supplements, 1965, 1, 33-78. (b)
- McClelland, J. L., & Johnston, J. C. The role of familiar units in perception of words and nonwords. Perception & Psychophysics, 1977, 22, 249-261.
- Mewhort, D. J. K., & Beal, A. L. Mechanisms of word identification. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 629-640.

- Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. Loci of contextual effects in visual word recognition. In P. Rabbit (Ed.), Attention and Performance V. New York: Academic Press, 1975.
- Miller, G. A., Bruner, J. S., & Postman, L. Familiarity of letter sequences and tachistoscopic identification. Journal of General Psychology, 1954, 50, 129-139.
- Neisser, U. Decision time without reaction time: Experiments in visual scanning. American Journal of Psychology, 1963, 76, 376-385.
- Neisser, U. Cognitive psychology. New York: Appleton Century Crofts, 1967.
- Oden, G. C. A fuzzy logical model of letter identification. Journal of Experimental Psychology: Human Perception and Performance, 1979, 5, 336-352.
- Oden, G. C., & Massaro, D. W. Integration of featural information in speech perception. Psychological Review, 1978, 85(3), 172-191.
- Olson, G. A., & Kausler, D. H. Orthographic distinctiveness of homophones. Behavior Research Methods and Instrumentation, 1971, 3, 298-299.
- Pollatsek, A., & Carr, T. H. Rule-governed and wholistic encoding processes in word perception. In P. A. Kolers, M. Wrolstad, & H. Bouma (Eds.), Processing of visible language I. New York: Plenum, 1979.
- Postman, L., & Conger, B. Verbal habits and the visual recognition of words. Science, 1954, 119, 671-673.
- Pyles, T. The origins and development of the English language. New York: Harcourt, Brace, & Jovanovich, 1964.
- Reed, A. V. Speed-accuracy trade-off in recognition memory, Science, 1973, 181, 574-576.

- Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. Journal of Experimental Psychology, 1969, 81, 275-280.
- Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.
- Savin, H. B. Word-frequency effects and errors in the perception of speech. Journal of the Acoustical Society of America, 1963, 35, 200-206.
- Selfridge, O. G., & Neisser, U. Pattern recognition by machine. Scientific American, August 1960, 203, 60-68.
- Shannon, C. E. A mathematical theory of communication. Bell System Technical Journal, 1948, 27, 379-423; 623-656.
- Shannon, C. E. Prediction and entropy of printed English. Bell System Technical Journal, 1951, 30, 50-64.
- Shebilske, W. Reading eye movements from an information processing point of view. In D. W. Massaro (Ed.), Understanding language: An information processing analysis of speech perception, reading, and psycholinguistics. New York: Academic Press, 1975.
- Shepard, R. N., & Podgorny, P. Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), Handbook of learning and cognitive processes: Human information processing (Vol. 5). Hillsdale, N.J.: Erlbaum, 1978.
- Sidman, M. Tactics of scientific research: Evaluating experimental data in psychology. New York: Basic books, 1960.
- Singer, M. H. The primacy of visual information in the analysis of letter strings. Perception and Psychophysics, in press.

- Smith, E. E., Haviland, S. E., Reder, L. M., Brownell, H., and Adams, N. When preparation fails: Disruptive effects of prior information on perceptual recognition. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 151-161.
- Smith, E. E., & Spoehr, K. T. The perception of printed English: A theoretical perspective. In B. H. Kantowitz (Ed.), Human information processing: Tutorials in performance and cognition. Potomac, Md.: Erlbaum, 1974.
- Smith, F. Understanding reading. New York: Holt, 1971.
- Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. Journal of Experimental Psychology, 1952, 43, 195-201.
- 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.
- 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(5), 475-484.
- Solso, R. L., & King, J. F. Frequency and versatility of letters in the English language. Behavior Research Methods & Instrumentation, 1976, 8, 283-286.
- Solso, R. L., Topper, G. E., & Macey, W. H. Anagram solution as a function of bigram versatility. Journal of Experimental Psychology, 1973, 100, 259-262.
- Spencer, H. The visible word. New York: Hastings, 1969.
- Spoehr, K. T., & Smith, E. E. The role of orthographic and phonotactic rules in perceiving letter patterns. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 21-34.

- Spreen, O., & Schulz, R. W. Parameters of abstraction, meaningfulness, and pronounceability for 329 nouns. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 459-468.
- Taylor, G. A. The contribution of orthographic structure to the perception of letter strings. Unpublished doctoral dissertation, University of Kansas, 1977.
- Taylor, G. A., Klitzke, D., & Massaro, D. W. A visual display system for reading and visual perception research. Behavior Research Methods & Instrumentation, 1978, 10, 148-153.
- Taylor, G. A., Miller, T. J., & Juola, J. F. Isolating visual units in the perception of words and nonwords. Perception and Psychophysics, 1977, 21, 377-386.
- Taylor, M. M. & Creelman, C. D. PEST: Efficient estimates of probability functions. Journal of the Acoustical Society of America, 1967, 41, 782-787.
- Thompson, M. C., & Massaro, D. W. Visual information and redundancy in reading. Journal of Experimental Psychology, 1973, 98, 49-54.
- Thorndike, E. L., & Lorge, I. The teacher's workbook of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.
- Tinker, M. A. Bases for effective reading. Minneapolis: University of Minnesota Press, 1965.
- Townsend, J. T. Theoretical analysis of an alphabetic confusion matrix. Perception and Psychophysics, 1971, 9, 40-50.
- Topper, G. E., Macey, W. H., & Solso, R. L. Bigram versatility and bigram frequency. Behavior Research Methods & Instrumentation, 1973, 5, 51-53.

- Travers, J. R., & Olivier, D. G. Pronounceability and statistical "Englishness" as determinants of letter identification. American Journal of Psychology, 1978, 91, 523-538.
- Tulving, E. Familiarity of letter-sequences and tachistoscopic identification. American Journal of Psychology, 1963, 76, 143-144.
- Turvey, M. T. On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. Psychological Review, 1973, 80, 1-52.
- Tzeng, O. J. L., Hung, D. L., & Wang, W. S-Y. Speech recoding in reading Chinese characters. Journal of Experimental Psychology: Human Learning and Memory, 1977, 3, 621-630.
- Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. New York: Academic Press, Lippincott, 1960.
- Vachek, J. Two chapters on written English. Brno Studies in English, 1959, 1, 7-38.
- Vachek, J. Written language. The Hague: Mouton, 1973.
- Venezky, R. L. The structure of English orthography. The Hague: Mouton, 1970.
- Venezky, R. L., & Massaro, D. W. The role of orthographic regularity on word recognition. In L. Resnick & P. Weaver (Eds.), Theory practice of early reading. Hillsdale, N. J.: Erlbaum, 1979.
- Wallach, M. A. Perceptual recognition of approximations to English in relation to spelling achievement. Journal of Educational Psychology, 1963, 54, 57-62.
- Wheeler, D. D. Processes in word recognition. Cognitive Psychology, 1970, 1, 59-85.

Whorf, B. L. Linguistics as an exact science. In J. B. Carroll (Ed.), Language, thought, and reality. Cambridge, Mass.: The MIT Press, 1956.

Woodworth, R. S. Experimental psychology. New York: Holt, 1938.

This page intentionally left blank

Glossary

accuracy task: a task in which accuracy is the primary dependent measure; in the target search task, average performance is maintained at 75% correct by adjusting the duration of the test string and/or the processing time available.

anagram: a permutation of the letters comprising a word.

bigram: a two-letter cluster.

catch trial: a trial during which the target letter does not appear in the test string.

feature detection: a process which answers the question of whether or not a particular feature is present and/or provides information about the degree to which the feature is present.

fourier analysis: the description of a visual display in terms of intensity and frequency of component sine waves.

full report: given a display of items, subjects are asked to report all of the items in the display.

generated abstract memory: modality-free (abstract) storage, analogous to short-term memory.

graphemic constraints: see scribal constraints.

high spatial frequencies: the high frequencies of a fourier analyses of a visual display.

homophones: different words that are pronounced identically, such as right/write.

information: source of knowledge available to the reader; for example, the featural information in preperceptual storage or the reader's memory of letter patterns in the language.

lexicon: storage of perceptual and conceptual codes of words in long-term memory; each word has perceptual codes such as the sound and visual pattern of the word and conceptual codes such as its meaning.

long-term memory: relatively permanent store of knowledge in term of codes, rules, and procedures (schemata).

low spatial frequencies: the low frequencies of a fourier analysis of a visual display.

morpheme identity: retention of spelling similarity of semantically-related words, such as cone/conic.

nonword: a letter string that does not spell a word; sometimes used to mean that the letter string also does not conform to standard English spelling, i.e., it violates English orthography.

orthographic regularity: description of orthographic structure based upon rules and generalizations about the observed patterns of English orthography; also referred to as rule-based description.

orthographic structure: the spelling constraints in a written language; in English, for example, given th in an initial position, r is more likely to follow than s.

overt judgments: judgments about the spelling of letter strings; studied to assess what knowledge of orthography is consciously available and capable of report.

parallel processing: when information about two or more items is processed simultaneously.

perceptual recognition: see primary recognition.

phonological constraints: the constraints on English orthography that derive from the allowable sequences of sounds in English words.

positional frequency: see position-sensitive measure.

position-dependent measure: see position-sensitive measure.

position-independent measure: see position-insensitive measure.

position-insensitive measure: a statistical-redundancy description of orthographic structure which counts letter sequences on the bases of occurrence in words of specific lengths without regard to position of occurrence; also referred to as position-independent measure.

position-sensitive measure: a statistical-redundancy description of orthographic structure which counts letter sequences on the basis of occurrence in specific positions in words of specific lengths; also referred to as position-dependent measure or positional-frequency.

postcue: when a target letter follows a test string.

precue: when a target letter precedes a test string.

preperceptual visual storage: holds features passed on by feature detection for roughly one quarter of a second.

primary recognition: joint evaluation and integration of the features in preperceptual storage and knowledge of orthographic structure in long-term memory; the outcome is a synthesized percept; also referred to as perceptual recognition.

probabilistic description: see statistical redundancy.

pseudoword: a nonword that conforms to English spelling.

reaction time (RT) task: task in which reaction time is the primary dependent measure; in the high-accuracy target search task, for example, accuracy is kept near perfect by

keeping the test string present during the response interval.

regularity: see orthographic regularity.

rule-based description: see orthographic regularity.

scribal constraints: the constraints in the selection of letters for either representing sounds or marking graphemic, graphotactic, or morphological functions; also referred to as graphemic constraints.

secondary recognition: transforms visual percept held in synthesized memory into meaning held in generated abstract memory.

single-letter positional frequency: the frequency with which a letter occurs in a specific position in words of a specific length in a sample of text.

speeded RT task: subjects are instructed to respond more quickly than the time needed for high accuracy; in the target search task, subjects learn to respond in synchrony with an external signal whose relative onset is adjusted to keep performance at 75% correct; both accuracy and RT are evaluated as dependent measures.

statistical redundancy: category of descriptions of orthographic structure derived from counts of letters or letter sequences occurring in the written language; also called probabilistic description.

summed counts: the sum of individual frequencies of the letters or letter clusters in a letter string.

synthesized visual memory: holds perceptual outcomes of primary recognition process.

target letter: in the target search task, the letter whose presence or absence in the test string of letters must be reported.

target search task: the subject is given a target letter and the task is to indicate whether or not the target letter is contained in a test string of letters.

test string: in the target search task, the letters that must be evaluated in terms of presence or absence of the target letter.

token count: a count based on the number of total occurrences of an item; for example, the total number of occurrences of all words containing a particular letter.

trigram: a three-letter cluster.

type count: a count based on the number of unique occurrences of an item; for example, the number of unique words containing a particular letter.

This page intentionally left blank

Author Index

A

- Adams, N., 58, 67
Allport, G. W., viii
Anderson, J. H., 11-12
Anisfeld, M. A., 25
Atkinson, R. C., 56

B

- Baddeley, A. D., 7, 64
Barbuto, P. F., 27
Baron, J., viii, 62
Beal, A. L., 24
Bjork, E. L., 62
Bouma, H., 12, 15-16, 43-45, 52-53
Bouwhuis, D., G. 26
Breitmeyer, B. G., 19
Broadbent, D. E., 19, 141,
Brownell, H., 58, 67
Bruner, J. S., 6, 24
Byrnes, J., 24

C

- Carr, T. H., ix
Cattell, J. M., 11-12
Chandler, J. P., 59
Chomsky, N., 35
Clark, H. H., 51

Conger, B., 132
Creelman, C. D., 93

D

Dearborn, W. F., 11-12

E

Egert, H. E., 25
Estes, W. R., viii, 5, 62

F

Fant, C. G. M., 14
Francis, W. N., 26-27, 126-127, 149

G

Ganz, L., 19
Geoffrion, L. D., 5
Gibson, E. J., viii, 1, 15, 25, 29-30, 32
Gilmore, G. C., 25
Green, D. M., 17

H

Halle, M., 14
Hammond, M., 25, 29-30
Haviland, S. E., 58, 67
Henderson, L., viii
Huey, E. B., 1, 13
Hung, D. L., 29

J

Jakobson, R., 14
Johnson, N. F., viii
Johnston, J. C., 26, 101, 132
Jones, R. D., 74, 77-78
Juel, C. L., 27
Juola, J. F., 56, 68

K

- Kallman, H. J., 58
Kausler, D. H., 33
King, J. F., 27
Klitzke, D., 105
Kleiman, G. M., 29
Kolers, P. A., 108
Krueger, L. E., viii, 7, 47
Kucera, H., 26-27, 126-127, 149
Kurath, H., 35

L

- Landauer, T. K., 27-28
Lefton, L. A., 24
Levin, H., viii, 1
Levy, B. A., 29
Lewis, S. S., 29-30
Lipscomb, C., 74, 77-78
Lorge, I., 25

M

- Macey, W. H., 27
Mason, M., 26, 37-38, 132
Massaro, D. W., ix, 2, 5-8, 17, 19, 21, 26, 29,
38-47, 43, 56-58, 62, 74, 77-78, 96, 105,
246, 248
Mayzner, M. S., 26, 48, 91
McClelland, J. L., 26, 101, 132
Mehwort, D. J., 24
Meyer, D. E., 74,
Miller, G. A., 6, 24
Miller, T. J., 68

N

- Neisser, U., 14, 43

O

- Oden, G. C., 21,
Olivier, D. G., 127-128, 131
Olsen, G. A., 33

Osser, H., 15, 25, 29-30

P

Pick, A., 25, 29-30
Podgorny, P., 131
Pollatsek, A., ix
Postman, L., 6, 24, 27-28, 131-132
Pyles, T., 32

R

Reder, L. M., 58, 67
Reed, A. V., 75, 105
Reicher, G. M., 6, 62
Rubenstein, H., 29-30
Rubenstein, M. A., 29-30
Ruddy, M. G., 74

S

Savin, H. B., 27-28
Schiff, W., 15
Schmuller, J., 7, 17, 19, 43
Scholz, R., 74, 77-78
Schulz, R. W., 25-26
Schvandeveldt, R. W., 74
Selfridge, O. G., 14
Shannon, C. E., 6, 24
Shapiro, R. G., 7, 47
Shebilske, W., 3
Shepard, R. N., 131
Shurcliff, A., 25-26, 29-30, 132
Sidman, M., 47
Skaar, E., 62
Smith, E. E., 4, 29-30, 58, 67, 132
Smith, J., 15
Solomon, R. L., 27-28, 131
Solso, R. L., 27
Spencer, H., 11-13
Spoehr, R. T., 132
Spragins, A. B., 24
Spreen, O., 26
Streeter, L. A., 27-28
Swets, J. A., 17

T

Taylor, G. A., 26, 38-47, 57-58, 68, 93, 105,
131, 248
Thompson, M. C., 6, 57-58, 62
Thorndike, L. L., 25
Thurston, I., 62
Tinker, M. A., 11
Townsend, J. T., 15
Topper, G. E., 27
Travers, J. R., 127-128, 131
Tresselt, M. E., 26, 48, 91
Tulving, E., 6
Turvey, M. T., 56
Tzeng, O. J. L., 29

U

Underwood, B. J., 25-26

V

Vachek, J., 33
Venezky, R. L., 26, 29, 31, 33, 35, 38-47, 57-58, 96

W

Wallach, M. A., 24
Wang, W. S-Y., 29
Wheeler, D. D., 6, 62
Whorf, B. L., 35
Wolin, B. R., 26
Woodworth, R. S., 3

Y

Yonas, A., 25-26, 29-30, 132

This page intentionally left blank

Subject Index

A

accuracy task, 40, 47-48, 56-61, 246-247
anagram, 38
awareness, 81

B

bigram, 25-29
bigram counts, 160-244

C

catch trial, 39
classroom practice, 779
confusion matrices, 14-15
consonant spelling, 31-32

D

decision, 17-18, 56, 57
 d' measure, 17-18

E

Englishness, 127-128
envelope, 15-16, 19

F

feature analysis, 2, 13-14
feature detection, 3-4, 19-21
feature detection strategy, 43-46, 68, 71-73
features, 3, 5, 14, 14-21
fourier analysis, 19
frequency scale, 130-132
full report, 6, 24-25

G

generated abstract memory, 8

H

high spatial frequencies, 19
homophones, 33

I

information, 2, 73-74
information processing model, 1-8, 55-62, 66-67

K

knowledge, 2, 74

L

legibility of text, 11-13
letter-sound correspondences, 29-30
lexicon, 8
lingiust, 34
long-term memory, 3, 8
low spatial frequencies, 19

M

masking stimulus, 6, 15, 17, 40
memory, 2-3
morpheme identity, 32
morphemic features, 32-33

N

nonword, 25

O

ordered approximations to English, 24-25
orthographic regularity, 38
orthographic structure, 2, 4-9, 23-36, 245
overt judgments, 81-90, 247

P

parallel processing, 5, 56
paired judgments, 81, 87-89, 120-123
perceptual recognition, 8
perceptual unit, 5
peripheral vision, 13, 15, 19
phonological constraints, 30-31
phonologically-based descriptions, 29-30
position-insensitive measure, 132
position-sensitive measure, 132
postcue, 47-48, 62-66
precue, 47-48, 62-66
preperceptual visual storage, 3-4, 56
primary recognition, 4-5, 7-8, 56
priming, 74
process, 2-3
processing time, 14
pseudowords, 24, 82-85, 115
psychologists, 34

R

ratings, 81, 81-87, 113-119
reaction-time task, 37-38, 47-48, 60-61, 246-247
reading, 1-2
reading instruction, 1-2, 247-248
recoding, 8
redundancy, 6
rehearsal, 8
Reicher-Wheeler paradigm, 6-7, 40
rule-based description, 23-24, 30-36, 38, 96, 129, 247
rule-governed regularity, 23-24, 30-36, 38, 96, 129, 247

S

saccades, 3
scribal constraints, 31-32
secondary recognition, 4, 8
short-term memory, 7-8
similarity, 43-46, 51-54, 64-65, 72-73
single-letter positional frequency, 37, 48
single-letter positional frequency counts, 149-159
speed-accuracy paradigm, 75
speeded RT task, 75-82
stage of processing, 2-3
statistical redundancy, 23-29, 247
stimulus items, 97-98, 143-148
summed counts, 37
synthesized visual memory, 8, 56

T

target letter, 37-39
target search task, 37, 37-81
target trial, 39
template, 14
test string, 37-39
token count, 27-28, 128-130
trigram, 27
type count, 27-28, 128-130

V

vowel spellings, 31-32

W

word frequency, 141
word recognition, 1-2, 4