This is a free page sample. Access the full version online.

SAA MP51—1986
UDC 800.866-057.3:651.7

SAA Miscellaneous Publication® MP51—1986

A STUDY OF LANGUAGE IN AUSTRALIAN BUSINESS DOCUMENTS

PUBLISHED BY THE STANDARDS ASSOCIATION OF AUSTRALIA
STANDARDS HOUSE, 80 ARTHUR ST, NORTH SYDNEY, N.S.W.

Incorporated by Royal Charter

E

This is a free page sample. Access the full version online.

PREFACE

This study was undertaken by the Association's Committee on Typing and Shorthand Tests. The study was designed and directed by Mr W L Williams of Williams Business Colleges with the assistance of committee members and Wang Computer Pty Ltd. It was based on a similar study of American business language by Dr B S Ober.

The study arose from the need to standardize the difficulty of shorthand speed test passages. Having determined that language difficulty could be controlled through the standardization of the syllabic intensity, the number of standard words, and the number of common words in the passage, the committee still did not know what levels should be set for these factors if the test passage was to resemble a passage likely to be encountered in an Australian office. Several overseas studies had been performed but there did not appear to be information available on the difficulty levels and vocabulary characteristics of Australian business language.

To overcome the problem, a sample of letters, memos, and reports from a wide range of Australian business and government was collected and submitted to computer analysis. The sample totalled 246 letters, 124 memos and 224 reports, yielding 435 738 running words and 17 545 different words.

CONTENTS

	Page
1 SCOPE	3
2 REASON FOR THE STUDY	3
3 METHOD	4
4 DATABASE	4
5 ANALYSIS OF THE DATA	5
6 DISCUSSION	6
7 CONCLUSIONS	6
BIBLIOGRAPHY	7
APPENDICES	
A THIRD AND FINAL FREQUENCY SORT OF WORDLIST (435 738 RUNNING WORDS; 17 545 DIFFERENT WORDS)	8
B 500 COMMON WORDS LIST	18
ANNEX 1. PROJECT DOCUMENTS	21

Review of Australian Standards. *To keep abreast of progress in industry, Australian standards are subject to periodic review and are kept up-to-date by the issue of amendments or new editions as necessary. It is important therefore that standards users ensure that they are in possession of the latest edition, and any amendments thereto.*

Full details of all SAA publications will be found in the Catalogue of SAA Publications; this information is supplemented each month by SAA's journal 'The Australian Standard', which subscribing members receive, and which gives details of new publications, new editions and amendments, and of withdrawn standards.

Suggestions for improvements to Australian standards, addressed to the head office of the Association, are welcomed. Notification of any inaccuracy or ambiguity found in an Australian standard should be made without delay in order that the matter may be investigated and appropriate action taken.

First published 1986

© Copyright — STANDARDS ASSOCIATION OF AUSTRALIA 1986

Users of standards are reminded that copyright subsists in all SAA publications. No part of this publication may be reproduced, stored in a retrieval system in any form or transmitted by any means without prior permission in writing of the Standards Association of Australia.

ISBN 0 7262 4414 7

This is a free page sample. Access the full version online.

STANDARDS ASSOCIATION OF AUSTRALIA

Miscellaneous Publication

A STUDY OF LANGUAGE IN AUSTRALIAN BUSINESS DOCUMENTS

1 SCOPE. This document sets out the methodology and results of a study designed to produce a list of common Australian business words and information on the difficulty of current business language.

2 REASON FOR THE STUDY. It has long been recognized that the main problem in standardizing the difficulty levels of shorthand speed test passages is one of standardizing the language used in the passages. If passages are to provide a valid measure of shorthand-writing skill, variations in writers' performances must reflect changes in skill rather than changes in passage difficulty. The same is true in many skill tests in which relative language difficulty plays a part.

Earlier work by James[1] recommended the counting of speed test words by sounds. He suggested counting four sounds as one standard word so as to reduce passage variability. His study found the most stable measuring unit of spoken language was the sound, followed by the letter and the syllable, with the word being by far the least stable. This can better be understood by comparing the words 'the', 'thorough', and 'responsibility'. Each is one word, but the syllable counts are 1, 2, and 6 respectively, the letter counts are 3, 8, and 14, and the sound counts 2, 4, and 14. James argued that the sound length of a word most closely correlated with the actual workload on the shorthand writer, while actual words varied enormously in the problems each posed.

It was not practical to use sounds as units of measurement in the shorthand standard, but a separate study[2] concluded that counting by syllables produced slightly less variable results than counting by letters. The syllable was therefore adopted as the standard unit for determining the length of a passage and for the placement of timing marks.

While counting in standard words of a certain number of syllables will reduce variability between quarter-minutes, the specification of average syllabic intensity (SI)[3] provides a control on the length of words within the passage. The average syllabic intensity of 287 current Australian shorthand speed tests from a wide variety of sources and representing speeds between 60 w.p.m. and 120 w.p.m. was 1.46[4]. The range was from a very easy 1.24 to a very difficult 1.92.

While the wide range of difficulty underlined the need for standardization, the real problem was whether the average SI of 1.46 correctly reflected the difficulty levels of Australian business documents. In a recent American study of 2504 business documents[5], the average SI was found to be 1.66. Only 13 of the 287 Australian speed tests equalled or exceeded this level of difficulty.

A further problem in standardizing shorthand speed test passages lies in the relative commonness of the words used. 'Given' and 'gamut' are each one word, two syllables, five sounds, and five letters, but there is a great difference in their familiarity and the ease with which they would be processed at speed. The relative familiarity of the vocabulary used in the test passage may be controlled by requiring that a certain percentage of words lie within the most common 500 or 1000 words. Pullis[6] requires that 55% of the words in the passage lie within the 100 most common words, 72% lie within the top 500 words, 85% lie within the top 1500 words, and only 10% lie outside the top 2000 words. This level of standardization would be unpractical in a standard for general use, but some form of common word control was desired. The second problem, therefore, was to produce a list of the most common words in Australian business documents.

While a number of American studies and common word lists are available, none is necessarily indicative of Australian business writing. Common word lists are particularly products of their era and culture. Pitman's 'One Thousand Commonest Words', widely used by Australian educators to grade shorthand material, is based on a pre-World War II American study by Dr Leonard Ayres and includes words such as 'automobile', 'railroad', 'elaborate', 'debate', 'folks', 'vessel', and 'cordially'. Ober's 500 most frequently occurring words include 'audit', 'faculty', 'inventory', and 'status'[7]. In addition, the average word length appears to be rising; Silverthorn[8] found an average SI of 1.54, Perry[9] 1.63, and Ober[10] 1.66, all using large and generally comparable databases.

As no previous study of Australian business language appeared to exist, this study of current business language as reflected in common business documents was undertaken.

1 JAMES, T.L. Standardizing Shorthand Tests, New Zealand Council for Educational Research, 1944.

2 WILLIAMS, W.L. Shorthand Speed Measurement Research Project, Standards Association of Australia, 1984.

3 Syllabic intensity is determined by dividing the total syllables in a passage by the total words. For example 'the cat sat on the mat' is 6 syllables and 6 words, SI = 6 ÷ 6 = 1.00; 'Ethyl the student went quantity surveying' is 12 syllables and 6 words, SI = 12 ÷ 6 = 2.00.

4 WILLIAMS, 1984, op. cit.

5 OBER, B.S. The Basic Vocabulary of Written Business Communications, Delta Pi Epsilon Journal, 24(1), 1982, pp. 13–27.

6 PULLIS, J.M. Personal communication, 1983.

7 OBER op. cit.

8 SILVERTHORN, J.A. The Basic Vocabulary of Written Business Communications. Unpublished doctoral dissertation, Indiana University, 1955.

9 PERRY, D.J. An Analytical Comparison of the Relative Word-combination Frequencies of Business Correspondence with Phrase Frequencies of Selected Shorthand Textbooks, Vols I and II. Unpublished doctoral dissertation, University of North Dakota, 1968.

10 OBER, op. cit.

COPYRIGHT

This is a free page sample. Access the full version online.

3 METHOD. The following method was used:

(a) A sample of letters, memos, and reports was collected from a wide range of Australian business and government. Documents were sought already on diskettes from users of Wang word processing systems throughout Australia so that cost and the enormous keying effort inherent in such a study could be reduced.

(b) The documents were analysed by computer to produce a word frequency list and statistics on language difficulty, in particular, syllabic intensity.

(c) The word frequency list was edited to produce a list of 500 common words.

4 DATABASE. In August 1984, 1200 clients of Wang Computer were each asked to provide on diskette approximately 5000 words of running text representative of their normal letters, memos, reports, etc. To preserve confidentiality, participants could replace names with the word 'NAME', figures with the word 'NUMBER', and blocks of text with the word 'BLOCK', all upper case. Participants also completed a short questionnaire seeking the name and address of the organization, contact details, and industry category according to the Australian Bureau of Statistics (ABS) subdivisions listed in Table 1.

One hundred and nine organizations responded, but this number was reduced to 101 participants, as some of the diskettes were in a form incompatible with the Wang OIS systems used to perform the study. Thirteen documents were removed from the remaining body of documents because they contained Wang Glossaries, form blanks, or mostly tabular material. There remained 246 letters, 124 memos, and 224 reports.

The 'reports' category was deliberately interpreted widely to allow the natural variety of documents representative of industry groups to be counted. They included manuals, schedules, speeches, press releases, forms, regulations, legal documents and agreements, job descriptions, notices, and specifications, as well as standard reports. Documents, or portions of documents, were only rejected if they did not meet the criterion of running text.

Table 1 shows the numbers of participants in each of the 12 ABS industry subdivisions, the letters, memos, and reports counted in the study, the percentage of the total word count provided by each industry subdivision, and the percentage of the whole work force employed in each industry subdivision as at February 1984.

It would have been desirable to stratify the samples of text according to the proportion of document producers in each industry category. Unfortunately this was not possible.

Firstly, no statistics exist to show the distribution of clerical, secretarial, and related workers by industry. While comparison of the total workforce percentages and the percentage of words analysed in Table 1 give an indication of the industry categories which are over-represented or underrepresented in the study, it is not easy to produce justifiable alternative figures on which to base stratification.

TABLE 1

SUMMARY OF PARTICIPANTS AND DOCUMENTS RECEIVED BY INDUSTRY CATEGORY

Industry Subdivision	Number of participants	Letters	Memos	Reports	Percentage of total words analysed	Percentage of total workforce*
Agriculture, forestry, fishing, and hunting	3	9	1	4	2.22	6.34
Mining	11	19	11	26	10.59	1.59
Manufacturing	12	23	13	34	10.42	17.75
Electricity, gas, and water	3	15	4	6	1.92	2.28
Construction	3	5	0	5	1.83	6.51
Wholesale and retail trade	1	4	4	1	0.50	19.65
Transport and storage	4	4	2	9	3.09	5.75
Communication	3	15	3	4	2.26	2.12
Finance, property and business services	34	80	29	80	23.95	9.68
Public administration and defence	15	53	42	32	26.00	5.05
Community services	12	19	15	23	17.22	16.79
Recreation, personal, and other services	0	0	0	0	0.00	6.49
TOTALS	101	246	124	224	100.00	100.00

* Source: Australian Bureau of Statistics, Catalogue No 6203.0 (February 1984)

COPYRIGHT

This is a free page sample. Access the full version online.

Secondly, the obvious underrepresentation in wholesale and retail trade and recreation, personal and other services has proved difficult to correct given the resources of the study.

Finally, it can be argued that the distribution of responses from users of one of the country's largest word processor suppliers is likely to be substantially proportional to document producers, at least in larger companies and government departments. Special measures were, however, taken at the analysis stage of the study, in view of the apparent overrepresentation of the finance, property and business services and public administration and defence categories. These are described later.

Distribution of participants by location is shown in Table 2. Representation compared with population is good except for the national capital where the response was increased by Commonwealth Government departments, and Victoria where the response was inexplicably low.

TABLE 2

DISTRIBUTION OF PARTICIPANTS BY STATE

State	Percentage
Australian Capital Territory	13
New South Wales	33
Northern Territory	0
Queensland	16
South Australia	17
Tasmania	2
Victoria	9
Western Australia	10
TOTAL	100

5 ANALYSIS OF THE DATA. As diskettes of documents were received, they were catalogued and transferred to 203 mm diskettes compatible with the Wang OIS-115 II system which was used for analysis. Some documents which would not transfer had to be abandoned, and others which did not meet the criteria explained before were eliminated. Codes were then assigned to documents to identify them as letters, memos, or reports.

The documents were scanned to remove tables and other non-running text. Hyphens were searched for and dealt with as follows:

(a) Compound hyphenated words, such as 'trade-in', were written out as separate words.

(b) Hyphenated words, one part of which was not a separate word, such as 're-enter', were left standing.

(c) Hyphens forming part of a word division at the end of a line were removed so as to rejoin the two parts of the word unless (a) or (b) above applied.

The words 'NAME', 'NUMBER', and 'BLOCK', where inserted for confidentiality, were then removed from the documents. Finally, underscoring, page breaks and format lines, which would have hindered the running of programs at the next stage, were removed.

Next, a specially written program analysed each document character by character. The program stripped away numbers, symbols, computer control marks and all punctuation apart from hyphens. Upper case letters were converted to lower case. The resultant documents consisted entirely of the remaining letter groups (words) and intervening spaces. The program then created a file for each new word it found, and set a counter in the file to 1. Subsequently, if the word recurred, the counter in its file was advanced by one. At the end of processing all the documents, 17 545 word files had been created with their respective counters indicating the number of times each word had occurred.

Because of the limited resources available for the study, much of the proofing and document preparation work that has characterized previous studies could not be carried out. The effects of this on the raw word-frequency lists and the corrections which had to be made to arrive at the final common word list are described in Clause 6. The final raw word-frequency list of the 700 most commonly occurring words is included as Appendix A.

Finally, the Wang Readability Index proprietary program was run through all documents to provide a measure of average syllabic intensity. The results of this, broken also into industry categories, are shown in Table 3. The average syllabic intensity for all documents was 1.712. Unfortunately, because of the organization of the documents on the diskettes, it was not possible to ascertain separate syllabic intensities for each letter, memo, or report, but a partial study showed the average SI of 53 letters to be 1.66, the average SI of 22 memos to be 1.66, and the average SI of 85 reports to be 1.74.

TABLE 3

SYLLABIC INTENSITY OF ALL DOCUMENTS WITHIN EACH INDUSTRY GROUP

Industry group	SI
Agriculture, forestry, fishing, and hunting	1.75
Mining	1.68
Manufacturing	1.70
Electricity, gas, and water	1.69
Construction	1.70
Wholesale and retail trade	1.59
Transport and storage	1.69
Communication	1.69
Finance, property, and business services	1.67
Public administration and defence	1.72
Community services	1.78
Recreational, personal, and other services	—
AVERAGE	1.712

COPYRIGHT

SAI GLOBAL

This is a free preview. Purchase the entire publication at the link below:

Product Page

Looking for additional Standards? Visit Intertek Inform Infostore

Learn about LexConnect, All Jurisdictions, Standards referenced in Australian legislation