The use of PRECIS in indexing Chinese documents: an experimental study

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/35631

Publisher: © Li-Kuei Hsueh

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
THE USE OF PRECIS IN INDEXING CHINESE DOCUMENTS: AN EXPERIMENTAL STUDY

by

Li-Kuei Hsueh, BA, MA

A doctoral thesis submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the Loughborough University of Technology

October, 1990

Supervisors: Dr J. D. White, BSc, PhD
A. O'Brien, BA, MLIS, Dip Lib

Department of Library and Information Studies

© Li-Kuei Hsueh, 1990
Dedicated to my parents
and my wife, Ai-Jen

謹以此文獻給

雙親和內人愛珍
I deeply regret that former supervisor, the late Dr J. D. White, passed away during the course of my studies. He endured great pain during his illness while still supervising my research. I would like to express my thanks for his supervision. I also would like to thank my current supervisor, Ms A. O’Brien, for her patience, her kindness, and above all her supervision.

I am grateful to Mr F. Smith for his permission to use his English PRECIS programs and his assistance in the modifications of these programs. I am also grateful to Dr R. S. K. Chan for his advice on linguistics.

Professor A. J. Meadows and Dr R. P. Sturges were respectively my Director of Research, and I am grateful for their direction, encouragement and provision of facilities for completing this research.

The scholarship for the present work was provided by the Ministry of Education of the Republic of China. The debt is gratefully acknowledged.

I would like to thank also a number of persons within the DLIS: the computer officer Mr G. Sargent for his technical help, the lecturer Mrs I. A. Smith for her suggestions in PRECIS, and my colleagues Zimin Wu for giving me many ideas, and Ailin for her help, encouragement, and faith.

I am thankful to my former boss and also my teacher Mrs Ou-Lan Chou, for her encouragement and for being my guarantor. I am
also thankful to my former colleagues of the National Central Library for their encouragement and help, and in particular to Mr Hsi-Chang Wang, Mr Cheng-Huang Lin, Miss Hsiu-Ying Chiang, Miss Hui-Lin Chen, Mrs Shun-Hua Wang and Mrs Shu-Hua Sun.

I am also grateful to Mrs J. Stevenson and my friends: Dr C. P. Garner, Dr M. R. Luo, and Dr C. S. Feng.

I am thankful for the encouragement I have received from my family, friends and relations, both here and in the Republic of China, and in particular to my parents, my wife, Ai-Jen (for her endurance), and my son, Hsueh-Chien.

Beyond all of these, I would like to thank Thee for Thy loving kindness and care that have made me able to carry out my study.
The main factors which prompted the present study were: (1) PRECIS has a linguistic universal feature for computerized subject indexing; (2) the largest Chinese bibliography and index published by the National Central Library of Taiwan still lack subject indexes; (3) both mainland China and Taiwan have created their bibliographic databases based on UNIMARC; and (4) the field 670 of the UNIMARC is reserved for PRECIS.

This study has aimed to experiment with PRECIS for indexing Chinese documents, generate Chinese subject indexes using PRECIS, and suggest the use of PRECIS in online retrieval in Chinese bibliographic databases. The last objective is an assumption which was based on the achievement of the first objective.

An experimental approach has been used in this study. It consisted of two stages: the pilot and the main experiments. Experiments with a sample of 498 Chinese strings were randomly chosen mostly from one issue of the *Index to Chinese Periodical Literature*. Using this sample, a comprehensive test of all of the PRECIS's role operators was carried out. Software employed in the experiment included the PRECIS English version programs, the ETien Chinese Computer System, the BIG-5 Chinese Internal Code of 13,053 Chinese characters, and a word processor (GALAXY). Both the pilot and main experiments were carried out on a microcomputer following several stages, i.e. (1) analysis of subject statements; (2) coding; (3) input, computer
processing, and output; (4) analysis of the output based on Chinese syntactic rules; (5) modifications of role operators and the accompanying programs; (6) comparison of the findings of the present and previous research; and (7) demonstration of Chinese index manipulation.

Eighteen problem areas were found during the experiment that were grouped into eight categories. These were: (1) connective codes, (2) coordinate concepts, (3) agents of transitive actions, (4) role definers, (5) two-way interaction, (6) author-attributed associations, (7) following differences, and (8) typography. These problems resulted mainly from the syntactic differences between English and Chinese. To overcome these problems, modifications of and additions to the existing role operators and the accompanying programs were suggested.

The comparison of the present and previous research revealed that the present study discovered more problem areas and suggested more efficient and effective solutions.

Finally, the whole procedure of manipulating Chinese documents was demonstrated, and Chinese subject index entries were produced successfully. The recommendations are also made for the future use and research of PRECIS.
Keywords

Chinese language
Chinese documents
indexing
information retrieval
PRECIS
String indexing
Subject indexing
# CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DECLARATION</td>
<td>iii</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td></td>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td></td>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
<tr>
<td></td>
<td>LIST OF TABLES</td>
<td>xiv</td>
</tr>
<tr>
<td></td>
<td>LIST OF ABBREVIATIONS</td>
<td>xvi</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 1: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Objectives</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Significance of the Study</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Definitions of Terms</td>
<td>5</td>
</tr>
<tr>
<td>1.5</td>
<td>Introduction to Chapters</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 2: AN INTRODUCTION TO PRECIS</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Background</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Linguistic Features</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Syntax</td>
<td>14</td>
</tr>
<tr>
<td>2.2.1.1</td>
<td>Transformational-generative (TG) grammar</td>
<td>14</td>
</tr>
<tr>
<td>2.2.1.2</td>
<td>Relevance of role operators to deep cases</td>
<td>16</td>
</tr>
<tr>
<td>2.2.1.2.1</td>
<td>Case grammar</td>
<td>17</td>
</tr>
<tr>
<td>2.2.1.2.2</td>
<td>Schema of role operators</td>
<td>18</td>
</tr>
<tr>
<td>2.2.1.2.3</td>
<td>Correspondence between role operators and deep cases</td>
<td>19</td>
</tr>
<tr>
<td>2.2.1.3</td>
<td>Entry format</td>
<td>24</td>
</tr>
<tr>
<td>2.2.1.4</td>
<td>Stages of concept analysis</td>
<td>26</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Semantics</td>
<td>28</td>
</tr>
<tr>
<td>2.3</td>
<td>Multilingual Experiments</td>
<td>30</td>
</tr>
<tr>
<td>2.4</td>
<td>Status of PRECIS Applied in Chinese</td>
<td>35</td>
</tr>
<tr>
<td>2.5</td>
<td>PRECIS used at Loughborough University of Technology</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>CHAPTER 3: CHINESE SYNTAX</td>
<td>48</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction to the Chinese Language</td>
<td>48</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Mandarin (Modern Standard Chinese)</td>
<td>48</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Writing System</td>
<td>48</td>
</tr>
<tr>
<td>3.1.3</td>
<td>Romanization</td>
<td>51</td>
</tr>
<tr>
<td>3.2</td>
<td>Chinese Syntactic Features</td>
<td>53</td>
</tr>
<tr>
<td>3.3</td>
<td>Case Grammar in Chinese</td>
<td>57</td>
</tr>
<tr>
<td>3.4</td>
<td>Syntactic Differences Between Chinese and English</td>
<td>59</td>
</tr>
</tbody>
</table>
# CHAPTER 4: COMPUTER PROCESSING OF CHINESE DOCUMENTS

- **4.1 Input Methods** ........................................... 75
- **4.2 Internal Codes** ......................................... 79
- **4.3 Output Devices** ......................................... 83
- **4.4 Hardware and Software** ................................ 84
- **4.5 Computer Processing of Chinese Documents in Libraries and Information Services** ............ 87

# CHAPTER 5: EXPERIMENTAL METHODOLOGY

- **5.1 Methodology** ........................................... 97
- **5.2 Sampling** ................................................ 101
- **5.3 Hardware and Software** ................................ 102
- **5.4 Limitations** ............................................. 103

# CHAPTER 6: DISCUSSION OF PROBLEMS

- **6.1 Generalities** ........................................... 106
  - **6.1.1 Role operators used** .......................... 107
  - **6.1.2 Problems encountered** ....................... 107
- **6.2 Philosophy** ............................................ 117
  - **6.2.1 Role operators used** .......................... 117
  - **6.2.2 Problems encountered** ....................... 118
- **6.3 Religion** ................................................ 119
  - **6.3.1 Role operators used** .......................... 119
  - **6.3.2 Problems encountered** ....................... 120
- **6.4 Natural Sciences** ..................................... 122
  - **6.4.1 Role operators used** .......................... 122
  - **6.4.2 Problems encountered** ....................... 123
- **6.5 Applied Sciences** ..................................... 123
  - **6.5.1 Role operators used** .......................... 123
  - **6.5.2 Problems encountered** ....................... 124
- **6.6 Social Sciences** ...................................... 137
  - **6.6.1 Role operators used** .......................... 137
  - **6.6.2 Problems encountered** ....................... 139
- **6.7 Chinese History & Geography** ...................... 142
  - **6.7.1 Role operators used** .......................... 142
  - **6.7.2 Problems encountered** ....................... 143
- **6.8 World History, Geography & Biography** .......... 144
  - **6.8.1 Role operators used** .......................... 144
  - **6.8.2 Problems encountered** ....................... 145
- **6.9 Language & Literature** ................................ 146
  - **6.9.1 Role operators used** .......................... 146
  - **6.9.2 Problems encountered** ....................... 147
- **6.10 The Arts** ............................................. 148
  - **6.10.1 Role operators used** .......................... 148
  - **6.10.2 Problems encountered** ....................... 148
6.11 Research on China & the Soviet Union ..... 149
  6.11.1 Role operators used ...................... 149
  6.11.2 Problems encountered ..................... 149

6.12 Chapter Summary .............................. 150
  6.12.1 Role operators used in eleven subject categories 150
  6.12.2 Problems encountered in eleven subject categories 153

CHAPTER 7: SOLUTIONS TO THE PROBLEMS ............................ 160

  7.1 Connective Codes .............................. 160
  7.2 Coordinate Concepts ........................... 161
  7.3 Agents of Transitive Actions ................. 162
  7.4 Role Definers - Operator 's' ................. 165
    7.4.1 'Applications' and 'use' ................. 166
    7.4.2 'Influence' ................................ 168
    7.4.3 'Effects, attitudes, role, policies, and participation' 169
  7.5 Two-way Interaction - Operator 'u' .......... 172
  7.6 Author-attributed Associations
      - Operator 't' ................................ 175
    7.6.1 'A compared with 'B' .................... 175
    7.6.2 'A' related to 'B' ....................... 177
    7.6.3 'A' expounded by 'B' ..................... 178
  7.7 Following Differences (Postpositions) ....... 180
  7.8 Typography .................................. 186
    7.8.1 Boldface .................................. 186
    7.8.2 Capitalization ............................ 186
    7.8.3 Italics .................................... 187
    7.8.4 Punctuation ............................... 187
  7.9 Chapter Summary .............................. 189

CHAPTER 8: COMPARISON OF THE PRESENT AND PREVIOUS RESEARCH .. 198

  8.1 Austin's Research ............................. 198
  8.2 Chor's Research ............................... 200
    8.2.1 Role definers - operator 's' .......... 201
    8.2.2 Agents of transitive actions .......... 204
    8.2.3 Two-way interaction ..................... 207
    8.2.4 Author-attributed associations - 'A'
        related to 'B' ............................ 211
    8.2.5 The set of Greek letters (α, β, γ) 215
  8.3 Chapter Summary .............................. 220

CHAPTER 9: GENERATION OF CHINESE ENTRIES .......................... 225

  9.1 Input of Chinese Strings ........................ 225
  9.2 Production of Chinese Entries .................. 228
  9.3 Production of Chinese Sort Keys ................ 228
  9.4 Sorting and Merging ........................... 229
    9.4.1 Sorting .................................. 229
    9.4.2 Merging ................................... 233
  9.5 Printout of Chinese Entries .................... 234
CHAPTER 10: CONCLUSIONS AND RECOMMENDATIONS

10.1 Conclusions ............................................. 236
10.2 Recommendations ........................................ 241

BIBLIOGRAPHY ....................................................... 247

LIST OF APPENDICES

Appendix

2.1 PRECIS operators and codes ......................... 267
4.1 Two-byte internal code (The BIG-5 code as example) .............. 269
4.2 The three-dimensional structure of the three-byte seven-bit code .................. 270
4.3 Structure of the Chinese operating system .................. 271
5.1 Sample PRECIS output from the pilot experiment with romanization ............... 272
5.2 Sample PRECIS output from the main experiment with Chinese .................. 273
7.1 Modification to connective codes ..................... 274
7.2 New algorithm for the Chinese downward-reading connective .................. 279
7.3 New algorithm for the Chinese upward-reading connective ..................... 281
7.4 Modification to the Chinese following differences ..................... 283
7.5 Modification to capitalization ......................... 285
7.6 Modification to Chinese punctuation ................... 286
9.1 Input and output Chinese records ................... 287
9.2 File records .................................................. 302
9.3 Production of Chinese entries ....................... 304
9.4 Production of Chinese sort keys .................... 309
9.5 Sorted Chinese records .................................. 314
9.6 Printout of Chinese entries ......................... 317
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Chapter 2:</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1 Relationship between syntax and semantics</td>
<td>13</td>
</tr>
<tr>
<td>2.2 The active-passive correspondence in English</td>
<td>15</td>
</tr>
<tr>
<td>2.3 PRECIS entry format</td>
<td>25</td>
</tr>
<tr>
<td>2.4 Relationship between 'Thing', 'Qualifier' and 'Lead', 'Qualifier'</td>
<td>26</td>
</tr>
<tr>
<td>2.5 Stages of concept analysis</td>
<td>27</td>
</tr>
<tr>
<td>2.6 Network of terms linked by thesaural relationships</td>
<td>29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 3:</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 3.1 The units of written English</td>
<td>62</td>
</tr>
<tr>
<td>3.2 The units of written Chinese</td>
<td>62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5:</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 5.1 Seven stages of the experiment</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 7:</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 7.1 The structure of instrumental verbs and the coverb 在</td>
<td>167</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 9:</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 9.1 Five stages of generating Chinese PRECIS index entries</td>
<td>226</td>
</tr>
<tr>
<td>9.2 Online input and output of a record</td>
<td>227</td>
</tr>
<tr>
<td>9.3 The structure of the record</td>
<td>227</td>
</tr>
<tr>
<td>9.4 The structure of the entry</td>
<td>228</td>
</tr>
<tr>
<td>9.5 Procedure of merging</td>
<td>233</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Chapter 2:</th>
<th>Table 2.1</th>
<th>Correspondences between deep cases and role operators</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 3:</td>
<td>Table 3.1</td>
<td>Category of coverbs</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Table 3.2</td>
<td>Guidelines to singular versus plural froms in English usage</td>
<td>69</td>
</tr>
<tr>
<td>Chapter 4:</td>
<td>Table 4.1</td>
<td>Category of Chinese characters' input methods</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>4.2</td>
<td>Chinese character internal codes</td>
<td>82</td>
</tr>
<tr>
<td></td>
<td>4.3</td>
<td>Four types of storage for Chinese characters</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>4.4</td>
<td>Databases of Chinese documents</td>
<td>89</td>
</tr>
<tr>
<td>Chapter 5:</td>
<td>Table 5.1</td>
<td>Subject categories of the sample</td>
<td>102</td>
</tr>
<tr>
<td>Chapter 6</td>
<td>Table 6.1</td>
<td>Distribution of 'Generalities' strings by role operators</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>6.2</td>
<td>Problems encountered in 'Generalities'</td>
<td>116</td>
</tr>
<tr>
<td></td>
<td>6.3</td>
<td>Distribution of 'Philosophy' strings by role operators</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>6.4</td>
<td>New type of problem encountered in 'Philosophy'</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>6.5</td>
<td>Distribution of 'Religion' strings by role operators</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>New type of problems encountered in 'Religion'</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>6.7</td>
<td>Distribution of 'Natural Sciences' strings by role operators</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>6.8</td>
<td>Distribution of 'Applied Sciences' strings by role operators</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td>6.9</td>
<td>New types of problems encountered in 'Applied Sciences'</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td>6.10</td>
<td>Distribution of 'Social Sciences' strings by role operators</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>6.11</td>
<td>New types of problems encountered in 'Social Sciences'</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>6.12</td>
<td>Distribution of 'Chinese History &amp; Geography' strings by role operators</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>6.13</td>
<td>Distribution of 'World History, Geography &amp; Biography' strings by role operators</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>6.14</td>
<td>New type of problem encountered in 'World History, Geography, &amp; Biography'</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>6.15</td>
<td>Distribution of 'Language &amp; Literature' strings by role operators</td>
<td>147</td>
</tr>
<tr>
<td></td>
<td>6.16</td>
<td>Distribution of 'The Arts' strings by role operators</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>6.17</td>
<td>Distribution of 'Research of China &amp; the Soviet Union' strings by role operators</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td>6.18</td>
<td>Percentage of strings according to subject categories and role operators</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>6.19</td>
<td>Problems encountered in eleven subject categories</td>
<td>154</td>
</tr>
</tbody>
</table>
Chapter 7
Table 7.1 The algorithm of the English downward-reading connective ($v$) ............................. 163
7.2 The algorithm of the Chinese downward-reading connective ($V$) ............................. 164
7.3 The algorithm of the English upward-reading connective ($w$) ............................. 174
7.4 The algorithm of the Chinese upward-reading connective ($W$) ............................. 175
7.5 Prepositional and postpositional codes .......... 181
7.6 The first number in standard differencing codes ... 185
7.7 Chinese differencing codes ..................... 189
7.8 The punctuation differences between English and Chinese ............................. 188
7.9 Solutions to the problems ....................... 190
7.10 Comparison of role operators used in English and Chinese ............................. 192

Chapter 8:
Table 8.1 Comparison of two proposed solutions to the problem related to operator 's' - 'applications' 204
8.2 Comparison of two proposed solutions to the problem related to agents of transitive actions 207
8.3 Comparison of two proposed solutions to the problem related to two-way interaction .......... 211
8.4 Comparison of two proposed solutions to the problem relate to author-attributed associations 214
8.5 Comparison of the use of the Greek letters and the operator 's' - 'applications' ................ 219
8.6 Comparison of Chor's and the author's solutions to the problems brought about by the role operators ............................. 221

Chapter 9:
Table 9.1 Example of sorting by radicals and numbers of total strokes ............................. 230
9.2 Example of sorting by numbers of total strokes and radicals ............................. 230
9.3 Example of sorting by Pinyin ............................. 231
9.4 Example of sorting by Mandarin Phonetic Symbol 232
LIST OF ABBREVIATIONS

Abbreviations Terms

Agt Agent
ASCII American Standard Code for Information Interchange
AUX auxiliary
Ben Benefactive
BLAISE British Library Automated Information Service
BNB British National Bibliography
CCII Chinese Character Code for Information Interchange
CISCO Chinese Character Cross Reference File
CISII Chinese Industrial Standard Code for Information Interchange
CJK Chinese, Japanese and Korean
CL classifier *
Com Comitative
Cov coverb
CRS currently relevant state (-le) *
DDC Dewey Decimal Classification
DUR durative aspect
EBCDIC Extended Binary Coded Decimal Interchange Code
Exp experiential aspect *
Exper Experiencer
GEN genitive *
G.R. Guoyeu Romatzyh
Inst Instrument
KIPS Korean Information Processing System
MARC Machine Readable Cataloguing
MPS Mandarin Phonetic Symbol
NP noun phrase
OCLC Online Computer Library Center
OCR Optical Character Recognition
Pat Patient
PFV perfective aspect (-le) *
PREMIS Preserved Context Index System
Q question
REACC RLIN East Asian Character Code
RIN Reference Indicator Number
RLIN Research Libraries Information Network
SICGCC Standard Interchange Code for Generally-used Chinese Characters
SIN Subject Indicator Number
SOV subject-object-verb
SVO subject-verb-object
TG Transformational-generative grammar
UTLAS University of Toronto Library Automated System
VP verb phrase
VSO verb-subject-object
WLN Western Library Network

[* Note: The abbreviations are according to Li and Thompson (1981).]
CHAPTER 1
INTRODUCTION

1.1 Background

PRECIS is a string indexing system based on the theories of case grammar and transformational grammar. Role operators used in PRECIS are founded upon these theories. From the point of view of grammatical principles, this system is a closer approximation to the universal features of natural languages and thus it is considered universally valid (Foskett, 1982). Consequently, there is a possibility of applying PRECIS to more than one language (Sørensen, 1977), even though it was originally devised for indexing in English.

It has been proved that PRECIS can be used with various languages other than English, such as Danish, Dutch, Finnish, French, German, Hungarian, Hindi, Italian, Polish, Spanish, and so on (Ramsden, 1981; Curwen, 1985). Therefore, a case could be made for the application of PRECIS in the Chinese language.

PRECIS is meant for subject indexing and can be separated into two aspects: (1) primarily for the printout of subject index entries, and (2) latterly on-line retrieval using subject terms. These two aspects are now described in more detail.

(1) Printout of subject index entries

In the English language, PRECIS has been used to produce printed subject indexes for:

a. the national bibliography: e.g. the Australian National Bibliography (Dykstra, 1986b); and the BNB (British
National Bibliography); 
b. the periodical index: e.g. the British Education Index (until 1986) (Vickers and Howarth, 1980); and an annual PRECIS index to the journal Argus provided by the Professional Librarians of Quebec (Bidd, Chevigny and Marshall, 1986).

In Taiwan, however, the national bibliography and the periodical index published by the National Central Library have only author and title indexes. Neither the Chinese National Bibliography nor the Index to Chinese Periodical Literature has a subject index. This causes a great inconvenience for users to retrieve information from these Bibliography and Index.

Accordingly, there is a necessity to experiment with subject index production in Chinese, and thus to help Chinese libraries produce printed Chinese subject indexes. The PRECIS system was chosen for this experiment.

(2) On-line retrieval using subject terms

It has been proved that UKMARC files on BLAISE (British Library Automated Information Service) can be searched using PRECIS words (Austin, 1986). In Canada, the UTLAS (University of Toronto Library Automation Systems) has developed a PRECIS software which can be used for online access to the database (Cain, 1986). In addition, the National Film Board of Canada has developed a national information system using PRECIS, known as FORMAT, which allows the users to do their online searchings (Dykstra, 1986a). Also, a project is being carried out by J. Congreve at the Middlesex Polytechnic of the United Kingdom,
"to analyse the problems of subject access in MARC based systems and to evaluate PRECIS as a retrieval tool" (Congreve, 1989:6). The data file evaluated in this project is based on the RIN (Reference Indicator Number) data which is supplied by the British Library. The RIN file will be used to switch an unused term input by the searcher to the preferred synonym in PRECIS.

This current development of PRECIS for online subject retrieval has prompted the possibility of using PRECIS to help the following situation in Chinese bibliographic databases.

The development of Chinese subject indexing is very diverse. Two Chinese subject indexing systems have been separately developed, i.e. in mainland China and Taiwan respectively by different institutions.

In mainland China, a Chinese subject heading list, the Chinese Thesaurus of Subject Terms, was compiled in 1980 by the joint effort of 1,300 experts from more than 500 information units and libraries. It covers all fields of science and technology and is divided into two parts: social sciences and natural sciences. This Thesaurus is regarded as the first national standard thesaurus in mainland China (Jiang and Liu, 1983; Lin, 1988). In Taiwan, a Chinese subject heading list, the First Edition of Chinese Subject Headings List, was compiled in 1984 by the Library Automation Planning Committee (Library Automation Planning Committee, 1984). Obvious problems result from not having a shared subject retrieval tool.
However, both mainland China and Taiwan have created their bibliographic databases based on the UNIMARC, the field 670 of which is reserved for PRECIS (Holt, 1987). If Chinese PRECIS data were available in these databases, then it might be possible to also search the data online.

In summary, the main factors which prompted the present study are:

1. PRECIS has a linguistic universal feature for computerized subject indexing;
2. the largest Chinese bibliography and index published by the National Central Library of Taiwan still lack subject indexes;
3. both mainland China and Taiwan have created their bibliographic databases based on UNIMARC; and
4. field 670 of UNIMARC is reserved for PRECIS.

1.2 Objectives

The above necessities and possibilities regarding the application of PRECIS for Chinese bibliographic databases have motivated the present study to formulate its objectives as follows:

1. to experiment with PRECIS for indexing Chinese documents;
2. to generate Chinese subject indexes using PRECIS; and
3. to suggest the use of PRECIS in on-line retrieval on Chinese bibliographic databases across different Chinese-speaking countries.
1.3 Significance of the Study

To the author's knowledge, there has not been any research into PRECIS using experimental methods and microcomputers in Chinese documents. Also, the PRECIS system has never been thoroughly examined or tested for its application to the Chinese language. The work of this thesis takes a substantial sample of Chinese documents and systematically applies PRECIS rules of indexing to them. This should indicate most of the possible advantages and disadvantages of PRECIS to the Chinese language.

Consequently, the success of the present study may have theoretical as well as practical contributions to the research into Chinese subject indexing using experimental method.

1.4 Definitions of Terms

In this thesis, certain terms are used with the meanings shown below:

Algorithm: A finite set of rules giving a sequence of operations for solving a specific type of problem (British Computer Society, 1987).

Display: The second line of a PRECIS entry, containing the concepts which depend upon the lead term (Buchanan, 1976).

Entry: A single record in an index, consisting of a heading followed by a location reference or by a 'see' cross-reference (British Standards Institution, 1988).
**Heading:** A term chosen to represent in the index an item or concept in a document. The heading of a PRECIS entry consists of the lead and the qualifier (British Standards Institution, 1988; Austin, 1984).

**Lead:** A term offered as the reader's access point in the index (Austin, 1984).

**Qualifier:** The terms which follow the lead on the first line of a PRECIS entry; the qualifier sets the lead into its wider context (Buchanan, 1976; Austin, 1984).

**Role operator:** Symbols which represent a relationship between concepts, and regulate the order of terms in a PRECIS string. Thus, it includes codes which initiate certain algorithms (Buchanan, 1976; Ramsden, 1981).

**String:** An ordered sequence of terms preceded by role operators (Ramsden, 1981).

### 1.5 Introduction to Chapters

The next three chapters give an overview of the PRECIS system, describe the Chinese syntax, and review the computer processing of Chinese documents, respectively.

Chapter 5 describes the methodology of the experiment which includes the stages followed during the experiment, sampling, hardware and software, and limitations. Chapter 6 presents the problems discovered through the analysis of the results of the experiment, based on the Chinese syntactic rules. In Chapter 7, solutions to the problems are suggested in the forms of modifications to the PRECIS' role operators and the
accompanying programs. The modifications are then compared with previous research in Chapter 8.

The whole procedure of manipulating the Chinese PRECIS system is demonstrated in Chapter 9, including the input, production of Chinese entries, production of Chinese sort keys, the sorting and merging, and printout of Chinese subject entries.

Finally, recommendations are given in the final chapter along with conclusions.
REFERENCES


CHAPTER 2

AN INTRODUCTION TO PRECIS

2.1 Background

The origins of PRECIS can be traced back to the nineteen-sixties. Two separate but related streams of research were developing at that time. The first of these was classification research. This research was supported by NATO (the North Atlantic Treaty Organization) and was carried out by the Classification Research Group (CRG) during the period of 1964-1968 in London (Austin, 1974b). The original purpose of this research was to develop a new classification scheme for the arrangement of the BNB to replace Dewey Decimal Classification (DDC) (Foskett, 1982).

The second stream was the development of MARC (Machine Readable Cataloguing). The BNB was looking towards the feasibility of using a computer to automate the production of its catalogue and to join the Library of Congress in a US/UK Project MARC in 1966. This project intended amongst other things to provide a co-extensive subject index for documents in the MARC database (Chan, 1981). A compatible US/UK communications format for the MARC records had been established in 1968 (Kent, Lancour and Nasri, 1970).

Before the PRECIS was used in the BNB, a chain index was the main index used in the BNB. Foskett (1982) suggested that a chain index has two defects: (a) it is difficult for the computer system to process, and (b) only the first entry was a specific entry. Moreover, the BNB's chain index to the DDC was
based on a human intellectual decision which was beyond the capacity of the computer (Austin, 1976b).

For the above reasons, therefore, a special research project was set up which was under the direction of D. Austin. This research was established with five principal goals (Austin, 1976b; Austin, 1982; Austin, 1984b):

1. An indexer should write a single string of terms which together summarize the subject of a document, and complete this string by adding instruction codes which tell the computer how and where each term should be printed.

2. Each of the entries produced in this mechanical fashion should be co-extensive with the subject as seen by the indexer; i.e. each entry should contain an equally complete summary of the subject.

3. Each entry should be printed in a meaningful order which is close to natural language to allow the user to understand the subject without the need for complex instructions in how to use the index.

4. The system should be based on a single set of logical criteria which could be applied consistently throughout the entire subject spectrum.

5. The terms selected as entry points in the index should be supported by a system of See and See also reference. These references would be extracted from a machine-held thesaurus.

After initial experiments, the first operational version of PRECIS was adopted by the BNB in January 1971 (Austin, 1974a).
The earliest version was called PRECIS I. Further research was conducted until the end of 1973. A new revised version of PRECIS, i.e. PRECIS II, has been regarded as the standard system since January 1974 (Bakewell, 1978; Robinson, 1979; Curwen, 1985). This new version of PRECIS is now being used by the BNB, and all earlier versions were abandoned.

The first edition of the manual - PRECIS: a manual of concept and subject indexing was published in 1974 (Austin, 1974c). This manual had 551 pages. As Curwen (1985:149-50) pointed out:

"it was a massive volume, far too long for teaching purposes, full of over-long explanations and prone to go off into interesting digressions and theorizing. What was lacking was a reasonably complete yet concise and convenient statement of the essentials of the system."

After ten years of experiments, the second edition of the Manual was published in 1984 (Austin, 1984a). This new edition corrected previous deficiencies and with a more concise exposition, reduced the number of pages to 397 and size (from 21 x 30 cm to 15 x 23 cm). A grid for differencing codes was used to replace the old codes; i.e. $01=\text{h}$, $21=\text{i}$, $02=\text{k}$, and $22=\text{m}$. In addition, two new operators, 'f' and 'u', were introduced. Curwen (1985) suggested that a reference handbook was needed which includes all the rules, codes and procedures. The handbook should be tabulated rather than context-oriented so it can be referred more easily. Further descriptions of PRECIS are available in Ramsden (1981), Richmond (1981) and Dykstra (1985).
2.2 Linguistic Features

PRECIS is based on the fundamentals of natural language which include two linguistic features; i.e. syntax and semantics. Syntactic theories, such as transformational-generative (TG) grammar and deep cases are applied in this system. The semantics is related to vocabulary control; namely the creation and maintenance of a thesaurus.

Although the syntax and semantics are treated as separate components and are manipulated by different procedures in PRECIS, a very strong relationship exists between them. Broxis (1976:54) noted that PRECIS "is an attempt to bridge the syntactic and semantic aspects of indexing through natural language as used in everyday speech." The relationship between syntax and semantics (Austin, 1977; 1984a) is illustrated in the diagram in Figure 2.1.

The diagram shows a distinction between syntactical and semantic relationships. The syntactic relation links the terms that make up a string, i.e.

*(0) England
*(1) foxes
*(2) hunting
The semantic relations set the terms in their network of meanings. In this case, only the hierarchical relations are shown, e.g. 'foxes' is a narrower term to 'canidae', which is part of a longer chain of generic relationships of 'animals'.

2.2.1 Syntax

As mentioned above, PRECIS is based on natural language grammars, particularly the notion of transformational-generative (TG) and deep cases. The schema of role operators is language-independent as is true in deep cases. Also, the stages of concept analysis is based on deep cases. These syntactic features of PRECIS will be described in the following sections.

2.2.1.1 Transformational-generative (TG) grammar

The origins of the transformational-generative school is associated with Chomsky. The aim of the TG school is to develop a transformational model for linguistic structure which "is more powerful than the immediate constituent model in certain important respects and which does account for such relations in a natural way" (Chomsky, 1965:6).

Transformational grammar consists of two major kinds of rules, i.e. phrase-structure rules and transformational rules (Caroll, 1986). The essential nature of phrase-structure rules is NP (noun phrase) and VP (verb phrase) as the basic constituents of a sentence in opposition to the more traditional grammatical concepts such as subject and object (Chomsky, 1965:68-69; Fillmore, 1969: 361). The grammatical names are divided into grammatical categories by the school of TG, such as N (noun),
NP, V (verb), VP, AUX (auxiliary), etc., these meanings are the same as in earlier grammars. Chomsky (1965:29) explained the functions of the transformation rule as follows:

"the syntactic component of a grammar must contain transformational rules mapping semantically interpreted deep structures into phonetically interpreted surface structures."

Amongst Chomsky's transformation rules, passive transformation is particularly emphasized in this section. A passive transformation rule was suggested by Chomsky (1957:43) as follows:

(Active) NP1 - Aux - V - NP2

(Passive) NP2 - Aux + be + en - V - by + NP1

If a passive transformation is applied, a deep structure (active sentence) would convert into a surface structure (passive sentence) (Jacobs and Rosenbaum, 1968). For example, an active sentence, "The child broke the glass" changed to passive sentence as, "The glass was broken by the child" through the passive transformational rule, as demonstrated by Figure 2.2 below.

<table>
<thead>
<tr>
<th>NP1</th>
<th>V</th>
<th>NP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active (Deep structure)</td>
<td>The child</td>
<td>broke</td>
</tr>
<tr>
<td>Passsive Transformation</td>
<td></td>
<td>Active verb</td>
</tr>
<tr>
<td>Passive (Surface structure)</td>
<td>The glass</td>
<td>was broken</td>
</tr>
<tr>
<td></td>
<td>Passive subject</td>
<td>Passive verb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.2: The active-passive correspondence in English

(Quirk et al., 1985:160)
The subject (the child) in the active sentence is changed to the agent of the passive sentence, and the object (the glass) in the active sentence became the subject of the passive sentence.

There are reasons to believe the passive voice reduces the syntactic complexity in indexing. Yngve (Bar-Hillel, 1964) stressed that by changing its voice from the active to passive, the syntactic complexity of a given sentence may be reduced. Batty (1984) further analysed both active and passive voices in indexing and found that the use of phrases was in passive, rather than in the active voice in natural language.

PRECIS input strings and entries are based upon the passive voice. Keen's (1977a) laboratory investigation had found that PRECIS input strings were in passive order. Furthermore, Keen's (1977b) experiment had shown that passive term order was used in 39% PRECIS index entries, compared with 0% in three other types of index entries (i.e. articulated prepositional, rotated string, and rotated term).

In addition to the passive transformation, PRECIS also applies the predicate transformation to the following role operators: agents of transitive actions, operators 's', 't', and 'u', and intake of internal processes (Austin, 1984a).

2.2.1.2 Relevance of role operators to deep cases

A PRECIS string is written based on the schema of role operators. These role operators are mostly in the area of deep cases. Both case grammar and the schema of role operators will
be described first, and the correspondence between them is explained in the later sections.

2.2.1.2.1 Case grammar

Chomsky proposed the notions of noun phrase and verb phrase stated in 2.2.1.1, these were questioned by Fillmore (1969) with respect to the validity of deep-structure. He suggested that the distinction between noun phrase and preposition phrase was unnecessary; for example, preposition phrases such as, "in the room, toward the moon, on the next day, in a careless way, with a sharp knife, and by my brother", are adverbials of "location, direction, time, manner, instrument, and agent" respectively (Fillmore, 1969: 362). Therefore, Fillmore (1968) suggested the "case" plays a role in the groundwork of grammars. He points out that:

"each simple sentence in a language is made up of a verb and a collection of nouns in various 'cases' in the deep structure sense."

(Fillmore, 1969:375)

The 'case' consists of a set of cases, i.e. Agentive, Instrumental, Dative, Objective, Factitive, Locative, Benefactive, and Comitative (Fillmore, 1968). Case grammar is organized around verbs and the above cases can occur with given verbs. In a later paper, Fillmore (1971) changed two names of his original case; i.e. 'Factitive' to 'Goal', and 'Dative' to 'Experiencer'. These cases are based upon deep-structure.

Psychologists also pay attention to case grammar because notions like agent, instrument and location appear to be much closer to speaker's and listener's intuitions than notions like noun phrase and verb phrase (Slobin, 1979; Carroll, 1986).
Apart from Fillmore, different sets of deep cases were suggested by several linguists, such as Chafe (1970), Grimes (1972), and Longacre (1976).

2.2.1.2.2 Schema of role operators

Vickery (1953) emphasized that the ideal index system should use one of the three mechanisms in natural languages for revealing the linkage between system and terms, i.e.

1. to add relational term 'operators' to identify and join terms;
2. to add an affix which reveals what 'case' it is in the phrase;
3. to analyse the elementary terms into fundamental categories.

A schema of role operators (see Appendix 2.1) includes two mechanisms that Vickery emphasized, i.e. 'operators' and 'cases'. The role operators consist of primary operators ('0' to '6') and secondary operators ('f', 'g', 'p', 'q', 'r', 's', 't', 'u'). Additionally, special codes, such as primary codes ($s, $y, $z, $a, $c, $d), secondary codes ($0 to $3), and typographic codes ($e, $f, $g, $h, $i) are devised also. Three principal functions for these role operators are (Austin, 1976d: 13):

"(a) to ensure, within reasonable limits, that compound subjects should be analysed into conceptual units according to common frames of reference shared by the members of an indexing team.
(b) to ensure that these concepts are set down consistently, in the same order in input strings. For this purpose, a filing order has been built into the schema - in particular, into the numbered or 'main line' operators."
(c) to serve as computer instructions which regulate the format of entries, the typography of terms, and their associated punctuation."

According to these three functions, the manipulation of this system is based on a single set of logical criteria which can be applied consistently and complete a string of terms to produce entries by the computer.

2.2.1.2.3 Correspondence between role operators and deep cases

Sørensen and Austin (1976a:114) analysed the role operators from the point of view of deep cases and these fall into three main groups:

"(a) those which identify deep cases: operators '0' to '6', 's', 't' and $d$ (or, occasionally, differencing codes);
(b) those indicating semantic relationships between concepts occupying the same deep case: operators 'p' to 'r', and $h$ to $o$. For convenience, the operator 'g' could be regarded as a member of this group...;
(c) those which function essentially as case markers: operators 's', 't', $v$ and $w$. It should be noted that the operators '4' to '6' identify concepts which lie outside the kernel subject statement...".

Austin (1982:158-166) has further investigated deep cases by classifying verbs into four types: states, process, action, and action/process. His method is borrowed for the purpose of this research, but verbs are represented by 'Action'. Furthermore, the cases that are used are different from Austin's. Eight cases were used in Austin's thesis; i.e. Patient, Complement, Factitive, Beneficiary, Agent, Instrument, Experiencer, and Location, whereas nine cases are chosen in this research; namely Agent, Instrument, Experiencer, Patient, Goal, Location,
Benefactive, Comitative, and Time. These are mostly based on Fillmore's (1968, 1971) cases. The differences of cases between Austin and those used in this study are: Factitive changed to Goal, Complement changed to Comitative, and the addition of Time as a case.

(1) Agent

Definition: the case of the typically animate perceived instigator of the action identified by the verb (Fillmore, 1968:24).

Example 2.1

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digestion of cattle</td>
<td>*1)cattle</td>
<td>Agent (Action)</td>
</tr>
<tr>
<td></td>
<td>*2)digestion</td>
<td></td>
</tr>
<tr>
<td>Damage to houses by typhoons</td>
<td>*1)houses</td>
<td>Patient (Action)</td>
</tr>
<tr>
<td></td>
<td>*2)damage$by$wto</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*3)typhoons</td>
<td>Agent</td>
</tr>
</tbody>
</table>

The above example shows that the agent could be the operator '1' for intransitive action (digestion) or '3' for transitive action (damage).

+ Note: The asterisks ('*') are used to indicate lead terms (terms that are offered as the user's access points in the index) throughout this chapter and the following chapters.

(2) Instrument

Definition: the case of the inanimate force or object causally involved in the action or state identified by the verb. (Fillmore, 1968:24).
Example 2.2

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of concrete in reconstruction of buildings</td>
<td>*1)buildings</td>
<td>Patient (Action)</td>
</tr>
<tr>
<td></td>
<td>*2)reconstruction$wof</td>
<td>(Action)</td>
</tr>
<tr>
<td></td>
<td>s)use$vofo$win</td>
<td>Instrument</td>
</tr>
<tr>
<td></td>
<td>*3)concrete</td>
<td></td>
</tr>
</tbody>
</table>

(3) Experiencer

Definition: the case of the animate being affected by the state or action identified by the verb (Fillmore, 1968:24; Fillmore, 1971:42).

Example 2.3

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading by Children in Japan</td>
<td>*0)Japan</td>
<td>Location</td>
</tr>
<tr>
<td></td>
<td>*2)reading$vby</td>
<td>(Action)</td>
</tr>
<tr>
<td></td>
<td>*3)children</td>
<td>Experiencer</td>
</tr>
</tbody>
</table>

(4) Patient

Definition: who or what is affected by an action. The Patient may be changed or moved, depending upon the meaning of the predicate (Grimes, 1972:148).

Example 2.4

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death of dolphins</td>
<td>*1)dolphins</td>
<td>Patient (Action)</td>
</tr>
<tr>
<td></td>
<td>*2)death</td>
<td></td>
</tr>
<tr>
<td>Doors broken by children</td>
<td>*1)doors</td>
<td>Patient (Action)</td>
</tr>
<tr>
<td></td>
<td>*2)b$breaking$by$swo$of</td>
<td>Agent</td>
</tr>
<tr>
<td></td>
<td>*3)children</td>
<td></td>
</tr>
</tbody>
</table>
(5) Goal

Definition: the case of the object or being resulting from the action or state identified by the verb, or understood as a part of the meaning of the verb (Fillmore, 1968:25; 1971:42).

Example 2.5

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of bridges</td>
<td>*1)bridges</td>
<td>Goal</td>
</tr>
<tr>
<td></td>
<td>*2)construction</td>
<td>(Action)</td>
</tr>
</tbody>
</table>

(6) Location

Definition: the case which identifies the location or spatial orientation of the state or action identified by the verb (Fillmore, 1968:25).

Example 2.6

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development of agriculture in Taiwan</td>
<td>*0)Taiwan</td>
<td>Location</td>
</tr>
<tr>
<td></td>
<td>*1)agriculture</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>2)development</td>
<td>(Action)</td>
</tr>
</tbody>
</table>

(7) Benefactive

Definition: someone or something on whom an action has a secondary effect. The action involves transfer, such as 'give', 'offer', 'lose', 'teach', etc. (Fillmore, 1971:52; Grimes, 1972:161).

Example 2.7

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Losses of jobs</td>
<td>*1)jobs</td>
<td>Benefactive</td>
</tr>
<tr>
<td></td>
<td>2)losses</td>
<td>(Action)</td>
</tr>
<tr>
<td>Reading books for children</td>
<td>*1)reading books</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>*6)children$01for</td>
<td>Benefactive</td>
</tr>
</tbody>
</table>
Both of the operators '1' and '6' are used in this case. Operator '1' (jobs) is regarded as 'loser' and the operator '6' (children) is regarded as 'receiver'.

(8) Comitative

Definition: the phenomenon of coordinate conjunction of noun phrases. A preposition 'with' or the conjunction 'and' is used to link terms (Fillmore, 1968:81).

Example 2.8

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebra &amp; calculus</td>
<td>*2) algebra &amp;v &amp;</td>
<td>Comitative</td>
</tr>
<tr>
<td></td>
<td>*f) calculus</td>
<td></td>
</tr>
<tr>
<td>Installation of security systems and fire systems for buildings</td>
<td>*1) buildings</td>
<td>Patient</td>
</tr>
<tr>
<td></td>
<td>*p) security systems$u &amp;</td>
<td>Comitative</td>
</tr>
<tr>
<td></td>
<td>*g) fire systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td>*2) installation</td>
<td>(Action)</td>
</tr>
</tbody>
</table>

(9) Time

Definition: any time predicate, e.g. Saturday, on the action (Bruce, 1975:346).

Example 2.9

<table>
<thead>
<tr>
<th>Subject statement</th>
<th>String</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education in China, 1127-1279</td>
<td>*0) China</td>
<td>Location</td>
</tr>
<tr>
<td></td>
<td>*2) education$ d1127-1279</td>
<td>Time</td>
</tr>
</tbody>
</table>

The above examples have shown that the role operators which correspond to deep cases are: '0', '1', '3', '6', 'f', 'g', 's' and '$d', as seen in Table 2.1.
Table 2.1: Correspondences between deep cases and role operators

<table>
<thead>
<tr>
<th>Cases</th>
<th>Operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent</td>
<td>1 (intransitive)</td>
</tr>
<tr>
<td></td>
<td>3 (transitive)</td>
</tr>
<tr>
<td>Instrument</td>
<td>3</td>
</tr>
<tr>
<td>Experiencer</td>
<td>3</td>
</tr>
<tr>
<td>Patient</td>
<td>1</td>
</tr>
<tr>
<td>Goal</td>
<td>1</td>
</tr>
<tr>
<td>Location</td>
<td>0</td>
</tr>
<tr>
<td>Benefactive</td>
<td>1 (loser)</td>
</tr>
<tr>
<td></td>
<td>6 (receiver)</td>
</tr>
<tr>
<td>Comitative</td>
<td>f (bound coordinate concept)</td>
</tr>
<tr>
<td></td>
<td>g (standard coordinate concept)</td>
</tr>
<tr>
<td>Time</td>
<td>$d</td>
</tr>
</tbody>
</table>

2.2.1.3 Entry format

The basic structure of the PRECIS entry format is two-line and made up of three parts. This design is based upon two major notions; namely context-dependency order and one-to-one relationships that are closely related. The former notion means that each term in a string sets the next term into its logical context. Each selected term will dominate the entry when it moves into the lead position (Austin, 1976a). For example, a string of four terms summarizing a subject:

Philippines - rural regions - women - life styles

These four terms are linked together to show the context of the subject. It is easy for the user realize an entry's meaning.
The one-to-one relationship is a sequence mainly used to explicate the relationship between terms within an entry. In order to preserve the one-to-one relationships, an algorithm is used for generating multiple entries out of a single input. This procedure is known as "shunting", producing entries in the 'standard format' of PRECIS (Austin, 1976a). PRECIS entry format is made up of three positions: i.e. lead, qualifier, and display, a mechanism to preserve the one-to-one relationships. This is pictured as follows:

```
<table>
<thead>
<tr>
<th>Heading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Lead</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Display</td>
</tr>
</tbody>
</table>
```

Figure 2.3 : PRECIS entry format

The diagram shows a three-part structure in which each part has a different function. The lead acts as the user's access point to the index, the qualifier contains the terms which set the lead in its wider context, and the display is context-dependent terms to the lead. Both lead and qualifier make up the heading.

The qualifier is the term which follows the lead. Evidence has shown that this term has been used in functional grammar. Halliday (1985) stated that qualifier is the element which follows the thing. The elements are words or sometimes word complexes, like two hundred, very big. The qualifier has the function of characterizing the thing. The thing is the semantic core of the nominal group which could be regarded as
The relationship between *thing*, *qualifier* and *lead, qualifier* is shown below:

<table>
<thead>
<tr>
<th>Thing</th>
<th>Qualifier</th>
<th>Lead</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>the car</td>
<td>which is made in Japan</td>
<td>cars</td>
<td>Japan</td>
</tr>
</tbody>
</table>

Figure 2.4: Relationship between 'Thing', 'Qualifier' and 'Lead', 'Qualifier'

2.2.1.4 Stages of concept analysis

The stages of analysing a subject depend upon deep cases theory and can be divided into three steps as follows (Sørensen & Austin, 1976a:122; Sørensen & Austin, 1976b:205):

Stage 1: Identify basic concepts in terms of their deep cases.

Stage 2: Identify the components of any concepts analysed at stage 1 which still contain embedded deep cases. This stage is recursive.

Stage 3: Establish the relationships between any concepts which occupy the same deep case. This stage is also recursive and might be applied in practice before stage 2.

Both Stage 1 and 2 are mainly used in syntactic analysis, while Stage 3 involves the analysis of semantic relations. In spite of being used in English, this model can also be applied in other languages. The procedure of concept analysis is based upon a "title-like phrase" which is the result of an indexer having examined a document and finally summarized its subject content. Although the PRECIS string favours the passive voice which derived from the deep-structure analysis of various index
entries, the phrase is cast in the active voice (Sørensen and Austin, 1976a). For example, a subject statement, such as "Research on the deterioration of 16th century manuscripts". The concept is analysed according to the above three stages, as shown in Figure 2.5. Stage 1 separates the sentence into a NP '16th century manuscripts deteriorate', and a V 'research'. The NP is further analysed at Stage 2, dividing a NP '16th century manuscripts' and a V 'deteriorate'. The NP '16th century manuscripts' is analysed by its semantic relations at Stage 3, where '16th century' represents a time.

After the three stages analysis, the string complete with its role operators, with a case marker, '$won' is added in order to explicate the relationship between the action 'deterioration'.
and the noun 'manuscripts', as seen below:

String:
*1) manuscripts d1500-1599
*2)deterioration w151
*2) research

The string is allocated a SIN (Subject Indicator Number), which consists of a seven-digit number struck from a computer-generated list (Austin, 1984a).

2.2.2 Semantics

The previous sections were concerned with syntactic features, entry format, and entries produced. This section moves on to the semantic side of the PRECIS; namely establishing terms which might function as a user's access points to the alphabetical index (Austin, 1977). An open-ended thesaurus is adopted which allows new terms into the thesaurus when they are encountered in the documents (Bakewell, 1975). This thesaurus is machine-held, controlled by the Reference Indicator Number (RIN). The RIN is a seven-figure number. Three types of relationships can be identified amongst terms, that are:

(a) Equivalence relationship (code $m)
   (i) Synonyms, e.g. EEC - European Economic Community
   (ii) Quasi-synonyms, e.g. imports - exports

(b) Hierarchical relationship (code $0)
   (i) Generic relationship, e.g.

   Canidae
   Foxes
(ii) Hierarchical whole-part relationship, e.g.

Japan
Tokyo

(iii) Instance relationship, e.g.

Rivers
Amazon River

(c) Associative relationship (code $n, $x, $y)

(i) Same category, e.g. energy --- power

(ii) Different category, e.g. earth --- atmosphere

The above codes are used to link different relationships, and therefore to generate two kinds of references, i.e. See and See also. The relationships between the following terms, "Manpower, Personnel, Employees, Manual workers, and Foremen", are shown below.

\[
\begin{array}{c|c|c}
$y$ & $m$ \\
Manpower & Personnel & Employees \\
\hline
$0$ & $0$ & $0$ \\
Manual workers & Foremen
\end{array}
\]

Figure 2.6: Network of terms linked by thesaural relationships

According to the above relationships, the cross-reference entries produced by the computer are:

Employees
See
Personnel

Manpower
See also
Personnel

Personnel
See also
Foremen
Manpower
Manual workers
PRECIS thesaural procedures can be applied in other languages without modification. There are two main reasons proposed by Austin (1984a:295-6):

"(a) The PRECIS thesaurus is based upon rules and relationships that are recognised in international standards (ISO, 1986) as both culture- and language-independent...
(b) The relational codes used in the PRECIS thesaurus are, by their nature, language independent..."

Apart from vocabulary control, PRECIS thesaural terms can be used in online searching. For example, users can search the UK-MARC files via access to BLAISE through keywords linked by Boolean operators (Austin, 1986).

2.3 Multilingual Experiments

PRECIS has been applied experimentally in various languages other than English since the early 1970s. Both multilingual and translingual indexing were used. Multilingual indexing means "the application of a given indexing technique independently in two or more natural languages" (Austin, 1982:239). Translingual indexing concerns "indexing a document in one source language, submitting the input string to a computer, and expecting the machine to translate this string into a different version in another target language, so that this new string can be manipulated by a standard problem into acceptable entries in the target languages" (Sørensen, 1977a: 96). The British Library has experimented with a translingual project which is limited to English, German and French (British Library, 1976; Verdier, 1979). A translingual switching program has written by P. Hancox (1983) which can translate a
PRECIS string written in English into French. In this section; however, multilingual indexing is mainly considered.

As mentioned earlier, PRECIS is mainly based on deep cases and TG. Fillmore (1968) notes that deep cases are linguistic universals. This allows PRECIS to be applied in various languages besides English. In early experiments with non-English languages, a research proposal was presented to the Office for Scientific and Technical Information (this Office has since changed to: the Research and Development Department, British Library) (Kent, 1983) which intended to test PRECIS possible application to other languages. The BNB staff translated English strings term by term into a second language and manipulated them into entries. These entries were then judged by staff subjectively or otherwise. by the non-English respondent (Austin, 1982). The results of the earlier experiments described in "Appendix Seven" of the first edition of the Manual (Austin, 1974c). Two categories were divided; i.e. Category A: languages that can be expressed by prepositions, whether or not inflections were used. Five languages were included in this Category: Dutch, English, French, German and Spanish. Category B: languages which either lack prepositions or use only a limited repertory of prepositions. This Category includes five languages also: Chinese, Finnish, Norwegian, Sinhala and Swedish (Austin, 1974c). The majority of the above ten languages are Indo-European except Chinese and Finnish.
These experiments were the first to comprehensively examine the feasibility of PRECIS used in multilingual indexing. Austin (1976d:33) noted that:

"these experiments were unplanned, and almost dilettantish; the languages chosen ranged from French to Chinese, but depended on skills which happened to be available among colleagues. Subjects which had been expressed in English strings were re-expressed, not necessarily by literal translation, in terms of a second language, preferably by a native speaker."

After these experiments, some new codes were needed in the light of inflections as well as compound terms in German and other languages.

A meeting was held in London in 1975 to examine the new codes designed to handle non-English languages. These languages included French, German, Danish and Afrikaans (Sørensen and Austin, 1976b). Some new codes were devised that pointed towards generalised solutions to specific language problems.

A series of four papers discussed PRECIS used in the multilingual context entitled PRECIS in a Multilingual Context by D. Austin, J. Sørensen and G. Lambert were published in the Libri (vol. 26, no.1-4, 1976). The first paper was an overview, which described the origins of PRECIS and its use in English language indexing (Austin, 1976d). From the linguistic basis, the authors (Sørensen & Austin, 1976a) of the second paper tried to explain that this system could be applied to more than one natural language. The third paper (Sørensen & Austin, 1976b) was the major part which discussed PRECIS
experiments in which some experiments were described. The new
codes and procedures proposed in the third paper were:
'identifying non-filing and non-printing data', 'inflection',
'differences', 'explicatives', and 'downward-reading
substitutes'. Examples were mainly given in Germanic
languages. Finally the last paper (Lambert, 1976) considered
the application of PRECIS in French. Some French examples were
presented in this paper and the current PRECIS experiments in
French were discussed.

A further discussion of PRECIS in multilingual consideration
was given by Austin (1982) in his Ph D thesis - PRECIS as a
multilingual system: a search for language-independent
explanations. In his thesis, PRECIS was regarded as a whole, so
that it could be applied to other languages apart from English.
From the point of view of theoretical explanations, this system
was set into a wider context. Several linguistic features were
discussed in Austin's thesis and new codes were proposed and
tested in languages ranging from Finnish, French, German,
Hungarian and Tamil, which included inflections, prepositional
vs postpositional, explicatives and transposition of terms.
Some of them have been discussed in Sørensen and Austin's
(1976b) paper mentioned earlier.

PRECIS is mainly used in English-speaking countries, notably in
the United Kingdom, others such as Australia (Balnaves, 1973),
Canada (Robinson, 1979; Bidd, Chevigny and Marshall, 1986) and
South Africa (Maassen, 1984). Apart from English, a number of
countries were interested in PRECIS and planned pilot projects;
for example, Denmark, France, Germany, Jamaica, India,
Malaysia, and Sweden (Austin, 1976c; Maassen, 1984). Additionally, several experiments have been carried out in various languages, such as Chinese (Chor, 1986; Hsueh and White, 1989), Danish (Sørensen, 1976), Farsi (Moradi, 1977; 1978), French (Lambert, 1976; Laliberte, 1977), German (Sørensen and Austin, 1976b; Maassen, 1984), Hungarian (Horvath and Orbau, 1987), Hindi (Venkatachari, 1982), Italian (Maltese, et al, 1984), and Polish (Sanders, 1977).

Different problems were encountered when PRECIS was applied in non-English languages. The following problems need to be taken into account before PRECIS can be used in one natural language (Sørensen, 1977a; Sørensen and Austin, 1976a):

(a) Schema of role operators: whether or not they can be used in other languages.

(b) Case marking: the use of connectives ($v$ and $w$) will vary from one language to another depending upon the language used.

(c) Semantic relations: can combinations of differences and foci, i.e. the codes and procedures for constructing compound terms and phrases be applied in the same way for any language?

Although a few foreign language examples were demonstrated in the first edition of the Manual, it lacks explanations of how to extend this system to other languages. Sørensen (1977a:94) suggested that "...an English sub-grammar extracted from a more generalized macro-grammar" for the Manual. The second edition of the Manual (Austin, 1984a) has accepted this suggestion which includes a section entitled "Applications of PRECIS in
new languages" (pp. 294-303). It broadly describes thesaurus and surface features (prepositions, postpositions, inflections and adjectives). In addition, the generalized connective codes (for prepositional and postpositional) were proposed to meet requirement of other languages.

On the grounds of the syntactic and semantic logic mentioned above, the PRECIS system can be applied to a range of different natural languages. However, due to the characteristics of various languages, certain extra machine routines and codes are needed to deal with diverse problems (British Library, 1976).

2.4 Status of PRECIS Applied in Chinese

The first discussion of PRECIS in Chinese appeared in the first edition of the Manual in 1974 (Austin, 1974c). As stated above, Chinese was included in Category B in the Appendix Seven of the Manual. One Chinese example was illustrated, which was also the first example of PRECIS in Chinese. Theme interlinks (x,y) were used in this Chinese sample.

Both D. Austin and J. Sørensen mentioned Chinese in their multilingual experiments. Sørensen (1977b:308) noted that Chinese "rely upon word order". Austin (1982:259) commented that the Chinese language "lacks prepositions and inflections, and relies mainly on social context and conventions governing word order to resolve potential ambiguities".

A more comprehensive discussion of the feasibility of applying PRECIS in Chinese was given by L. Chor (1986). Chor provided a comparison of syntactic differences between Chinese
and English. A by-pass procedure using the theme interlinks was demonstrated. Four kinds of role operators were discussed, i.e. (1) the role definer - 's' - 'applications', (2) the transitive action and the agent, (3) the two-way interaction, and (4) the author-attributed association. All of the above role operators were illustrated using both theme interlinks and backward slashes for Chinese prepositions. Furthermore, the author suggested a special set of three Greek letters (α, β, γ) representing the key system (object), the agent and the action, respectively. Both Chinese and English examples were illustrated in this paper.

Chor's study mainly discussed some role operators used in Chinese which would cause problems. Although these operators created major problems in Chinese, only a few examples were discussed and therefore cannot represent the whole problem. Furthermore, only one example each of operators 's' and 't' was given. These did not reveal other problems that would be encountered as a result of these two operators, particularly operator 's', which has more than eight types in English. One would encounter different problems when using them in Chinese.

In a paper presented by L. K. Hsueh and J. D. White (1989), the authors compared the syntactic differences between Chinese and English. They noted that there was no capitalization or spacing in Chinese. In addition, major problems had arisen from the following role operators: agents of transitive actions, 's', 't', and 'u'. These problems were due to syntactic differences between English and Chinese. The authors suggested that techniques used in artificial intelligence (AI) could
possibly be applied to PRECIS in two aspects, e.g. analysis of the document's subject statement and coding of the string. A number of Chinese examples were demonstrated in this paper.

Five Chinese articles were found in the literature on PRECIS. The first introduction to PRECIS in Chinese was given by H. H. Wang (1981). This paper generally introduced the background and formats of PRECIS. L. P. Mai (1984) discussed the principles, structures and the developing trends of PRECIS. A detailed description of PRECIS in Chinese was provided by C. C. Mao (1985). In this paper, the development background of PRECIS, its theory, indexing methods and internationalization were described. In addition, the current status of Chinese subject headings was discussed in Mao's paper. The author also suggested that PRECIS could be used in Chinese to solve some problems of indexing Chinese documents. A general introduction to PRECIS was given by L. K. Hsueh (1986). The indexing stages of PRECIS were explained and one Chinese example was given in this paper. Finally, a paper was presented by K. C. Wang and C.H. Huang (1986). This paper mainly introduced the role operators of the second edition of the Manual.

According to the literature survey to date, there are currently no institutions using PRECIS in the indexing of Chinese documents. Dykstra (1986:v) mentioned that "UTLAS staff members have encountered special interest in PRECIS on the part of Chinese librarians. Therefore a capacity to offer PRECIS in Chinese,..." A letter has been sent to UTLAS to request this and a response has been received (Cain, 1989) but UTLAS has never pursued this. No experiments have been found which use
microcomputers to experimentally test the indexing of Chinese documents using PRECIS.

2.5 PRECIS used at Loughborough University of Technology

PRECIS has been taught at the Department of Library and Information Studies of Loughborough University of Technology since the early 1970s (Smith, 1986). Because PRECIS was designed for production of computer generated index entries, the Department has developed a computerised demonstration system for online validation of PRECIS strings and entries. Development of the system began in 1977, when the Department acquired a minicomputer which served sixteen terminal lines.

As there was no PRECIS software publicly available in the late 1970s, the Department's prototype software was written by F. Smith using the BASIC programming language. A revised version of the program ran on a Digico minicomputer in 1983. Later, a new version of the program, which was written in Digital Research's CBASIC, was translated to a modern microcomputer (COMART) in 1985. This new version can run all routines for input and manipulation of strings. With the new software, a wider range of functions was also available. A new BASIC language - TURBO BASIC was used to update this system and in 1988 to expand it for use on IBM microcomputers.

The PRECIS course is taught at the Department of Library and Information Studies of Loughborough University of Technology at both undergraduate and postgraduate level. The course includes the syntactical component and the thesaural element of the system. The role operators are introduced first at lectures,
and followed by practical sessions with small groups of students. This system allows students access to terminals which enable them to input strings and produce entries.

Based on the English version PRECIS system, F. Smith has experimented with this system in other languages, such as German (Deutschebibliothek), Italian (SINTESI), and has now assisted in a development for Chinese.

In summary, PRECIS has two linguistic features, i.e. syntax and semantics. The semantics is related to vocabulary control, i.e. the creation of a thesaurus, for which an international standard is available. In addition, the relational codes used in the PRECIS thesaurus can be used in natural languages. On the other hand, the syntactic rules are different from one language to another language. Therefore, the application of PRECIS in other languages would encounter some problems.

The next chapter will discuss the characteristics of the Chinese language and compare the syntactic differences between the Chinese and the English languages.
REFERENCES


Indexer. 10(2):54-59.

BRUCE, B. (1975) Case systems for natural language. Artificial 
Intelligence. 6(4):317-360.


Books/Cole Publishing Co.

Chicago: Univ. of Chicago Press.

CHAN, L. M. (1981) PRECIS. In: Chan, L. M. Cataloguing and 
pp.187-205.


Mouton.

Dykstra, M. (ed.) PRECIS: recent applications. Halifax: 
Dalhousie University, pp. 20-42.

Librarianship. 17(4):244-266.

Library.

(ed.) PRECIS: recent applications. Halifax: Dalhousie Univ. 
School of Library Service, pp. i-ix.

FILLMORE, C. J. (1968) The case for case. In: Bach, E. and 
Harms, R. T. (eds.) Universals in linguistic theory. 


Huddersfield: British Computer Society, Information
Retrieval Specialist Group, pp.154-176.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION (1986)
Documentation - guidelines for the establishment and
development of monolingual thesauri. 2nd ed. (ISO
2788-1986) Genever: ISO.

JACOBS, R. A. AND ROSENBAUM, P. S. (1968) English
transformational grammar. London: Ginn and Company Ltd.

KEEN, E. M. (1977a) On the generation and searching of entries

KEEN, E.M. (1977b) On the processing of printed subject index

Encyclopedia of Library and Information Science. v.3,

science. v.36 suppl. 1 New York: Marcel Dekker, Inc.

LALIBERTE, M. A. (1977) Selected grammatical and syntactical
problems in adapting PRECIS to the French language.
Unpublished Ph D. dissertation. Cleveland: Case Western
Reserve Univ.

LAMBERT, G. (1976) The application of PRECIS in French. Libri

LONGACRE, R. E. (1976) An anatomy of speech notions. Lisse:
Peter de Ridder Press.


SANDERS, J. (1977) PRECIS goes to Poland. *The British Library*


3.1 Introduction to the Chinese Language

The Chinese language differs from European languages in some fundamental respects. Some basic elements are indispensable to a general understanding of this language. This section includes the Chinese official language—Mandarin, the writing system and romanization methods.

3.1.1 Mandarin (Modern Standard Chinese)

There are many dialects in China, particularly those used in southern provinces, such as Cantonese, Hakka, and Hokkien. Mandarin or Modern Standard Chinese is an official language, which represents the speech of Peking. The city of Peking has been China’s capital for a long period in Chinese history. The language spoken in this city forms the national language. The official language has been known as 普通话 'P'utunghua' in the People’s Republic of China (mainland China), which means the 'common language'. In the Republic of China (Taiwan), it is called 国语 'Kuoyu', which means 'national language'. Both P’utunghua and Kuoyu are called Mandarin or Modern Standard Chinese (Kratochvil, 1968; Li and Thompson, 1981).

3.1.2 Writing system

Perhaps the most striking typological feature of spoken Chinese is that its dialects (of which there are over 200 major varieties) are not mutually intelligible. For instance, the person speaking a Shanghai dialect will find Cantonese as
difficult to understand as an entirely foreign language. Given this lack of mutual intelligibility, the question arises if Chinese dialects can be said to constitute a single language at all. Several typological features make clear that there are indeed varieties of one language. One is a common syntax; also there is a common morphology; and there is a common writing system.

Although Mandarin and various dialects exist as virtually separate spoken languages in China, the common writing system has been using since the Ch'in Dynasty (221 B.C.) (Wang, 1977). According to the *K'anghsi Dictionary* which is the most authoritative source, there are more than 49,000 different Chinese characters in the vocabulary, but about 4,800 characters can satisfy 95% of general use, such as teaching, writing, newspaper printing and word processing (Ting, 1985). Business data-processing and general library usage need a 9,000 - 15,000 character set (Cheng, 1989).

The form of writing in Chinese is known as Hanzi 漢字. Each character is an ideogram which is made up of two types of smaller units, i.e. strokes and the radical. A stroke is a line either straight or curved (Wang, 1973; Zheng and He, 1986). The 214 radicals have been adopted in present dictionaries (Chen, 1987). Examples such as 山 'mountain' and 日 'sun', are built up of three strokes of the former and four strokes of the latter. These two characters are also radicals. They can made up of more characters. For example, the radical 木 'wood' can constitute characters, such as 李 'plum', 杏 'apricot', 村 'village', and so on.
The basic characteristic of the Chinese language is that each character is unique and can be isolated respectively. Most characters are monosyllabic. Each character has three associated features: orthography, phonology, and meaning.

The smallest unit of meaning of grammar, the morpheme, is invariably also a word in Chinese, i.e. all morphemes in Chinese are free, not bound. However, it is not the case that all Chinese words, as many Chinese words are compounds, made up of two or more morphemes (or words). This is the situation if we understand 'morpheme' as traditionally. However, because of the special characteristic just mentioned, Chinese linguists have further defined morphemes that are derived from the particular characteristics of Chinese. Lu (1979:14) defined morphemes as "the smallest units of phonetic-semantic integration". Morphemes can be separated into free and bound morphemes (Chao, 1968; Henne, Rongen and Hansen, 1977).

The free morphemes have their own meanings and they can occur as free words, such as 人 'man', 馬 'horse', 月 'moon', and 錢 'money'. On the contrary, the bound morphemes can only occur as constituents of words. They should be connected with other morphemes, for instance 葡萄 'grape', 玻璃 'glass', and 巧克力 'chocolate' are all bound morphemes.

Chinese writing belongs to a morphemic writing system which is written morpheme by morpheme (Hockett, 1958). The arrangement of the traditional Chinese writing system is in columns. The sequence is from top to bottom and right to left. However, some journals and newspapers have changed into Western style, i.e. arranged in rows. Reading sequence is from left to right within
a row (Yu & Xiao, 1983). Amongst the column or row, every Chinese character is equally without character boundaries except punctuation. This last feature poses a very severe methodological difficulty for any Chinese automatic retrieval system.

The Chinese writing system was divided into two types in 1956. The government of the People's Republic of China simplified the Chinese characters in that year (Wang, 1973). Simplified characters are being used in mainland China as well as in Singapore. Traditional full Chinese characters are still used in Taiwan and overseas. Both writing systems are being used in Hong Kong (Cheng, 1989). Traditional full Chinese characters are chosen in this research because of its usage in Taiwan.

3.1.3 Romanization

Romanization is a method for converting a nonroman script into a word that sounds like the original but is written in the roman alphabet (Miller, 1982). There are four methods of romanization which have used for transcribe Chinese characters.

(1) Gwoyeu Romatzyh (G.R.) or National Romanization

The full standardization of Mandarin includes two parts: the Mandarin Phonetic Symbols (MPS) and G.R. The MPS, a non-Latin alphabet, was adopted by the Ministry of Education of China in 1918. The main purpose of MPS was to promote the standardized pronunciation to facilitate communication between people of different races and origins (Tse, 1986). MPS is being used in Taiwan. G.R. was designed by Chinese scholars in 1926, and was adopted by the Ministry of Education in 1928.
used to represent Chinese names or words that are mentioned in a foreign text or in public signs for foreigners. The distinctive feature of this system is that it spells syllables in different tones with different letters (Chao, 1967). A new system was released by the Ministry of Education of the Republic of China in 1984, which is intended to replace the old system (Jordan, 1985).

(2) Wade-Giles

The system was invented by T. F. Wade (1818-1895), the ambassador of United Kingdom at Peking during 1871-1883 and revised by H. A. Giles (1845-1935). This system is more widely used in Taiwan and many Western language libraries for transcription of Chinese titles, authors, and place names (Ting, 1985; Wood, 1988). This system is chosen in this research.

(3) Yale

During World War II, the United States Government taught oriental languages to air force pilots. The Yale Institute of Far Eastern Language was in charge of this training. This system was designed to get quick results with home-based students learning spoken Chinese for service in the Far East (Newnham, 1973).

(4) Pinyin

This system has had a long evolution. The original system was devised by A. Dragunov for teaching Chinese in Russia under the name of Latinxua or Latinization in 1931. In 1952 a
research committee was formed to consider all aspects of the written language. The committee decided on Pinyin after some experiments. The Pinyin system uses Latin letters to spell Chinese characters and has been in use since 1958. In 1979, it became the official romanization system for Chinese characters in mainland China (Newnham, 1973; Ting, 1985).

3.2 Chinese Syntactic Features

There are several syntactic features in Chinese. Four significant features are described in this section, i.e. word order, topic-prominent, non-inflectional and coverbs.

(1) Word order

Chinese linguists emphasize that word order is the major characteristic feature of Chinese syntax. Herdan (1964:11) has said that "the whole of Chinese grammar depends on position". Chao (1968:260) further commented that "all Chinese grammar is syntax, all Chinese syntax is word order, and therefore all Chinese grammar is word order". As a result, grammatical relations and the function of words in a sentence are determined by word order (Miao, 1981). Different word order forms different meanings, even though the same characters are used. For example, 水冷, means "the water is cold", while with the word order reversed to 冷水, the meaning is changed to "cold water".

(2) Topic-prominent

In addition to word order, the most important syntactical factor of Chinese syntax is topic-prominence. A subject is traditionally assumed to be essential in a sentence of English
and related European languages. On the contrary, the subject sometimes can be omitted in the Chinese language. For example, 下雨了 'It is raining', the subject 天 'it' is omitted when the speaker and the listener both know the subject (Wang, 1947 v.2). Li and Thompson (1981:15) explained the importance of the 'topic':

"...the description of Mandarin must also include the element "topic." Because of the importance of "topic" in the grammar of Mandarin, it can be termed a topic-prominent language. Basically, the topic of a sentence is what the sentence is about. It always comes first in the sentence..."

Topic is an important element in the Chinese language. The first noun phrase in a sentence is usually associated with the topic, as is seen in Example 3.1 and Example 3.2.

Example 3.1 這朵花你喜歡嗎? (* see note below)

```
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>this CL flower you like Q</td>
</tr>
<tr>
<td>Topic</td>
</tr>
<tr>
<td>Subject</td>
</tr>
</tbody>
</table>
```

(Do you like this flower?)

Example 3.2 那本書我讀過

```
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>that CL book I read Exp</td>
</tr>
<tr>
<td>Topic</td>
</tr>
<tr>
<td>Subject</td>
</tr>
</tbody>
</table>
```

(I have read that book.)

* Note: In order to explain clearly, spaces are placed between Chinese characters in this and the following examples.

The above examples, 這朵花 'this flower' and 那本書 'that book' are all 'topic' and come first in Chinese sentences.
Consequently, Chao (1968:69) suggested that 'topic' and 'comment' rather than 'subject' and 'predicate' be the primary syntactic division in Chinese. The 'comment' is a well-formed clause which relates to the 'topic' (Liejiong and Langendowu, 1985). Two clauses are shown in the above examples, '你喜歡' 'you like' and 我讀過 'I have read', are all 'comment'.

(3) Non-inflectional

The inflectional languages have a high degree of morphological complexity with respect to the types of grammatical morphemes. By contrast, Chinese is a non-inflectional language, which has restrictions for distinguishing grammatical categories (Ross, 1984). Many syntactic features are primarily due to the differences between inflectional and non-inflectional languages. As mentioned in section 3.1.2, Chinese is an isolating language. Each free word consists of just one morpheme, therefore it cannot be further analysed into component parts (Li and Thompson, 1981). As a result, Chinese lacks complexity in word formation, such as: the declension of nouns for case, and for number (singular and plural); the conjugation of verbs for tense (present, future, past) and for aspect (perfect, imperfect, durative, etc.); the formation of adjectives from nouns by suffixation; and the formation of adverbs from adjectives (Hutchins, 1975: 18).

(4) Coverbs

The equivalent in English to Chinese coverbs are prepositions. The earlier linguists (Chao, 1968; Lu, 1979) tend to endorse the preposition category in Chinese syntax, but recently the notion of coverbs has been widely accepted by linguists (Li &
Thompson, 1981; Tai, 1982; Ross, 1984). These different viewpoints arose from the characteristics of coverbs. Chao (1968:751) stated that one of the features of Chinese prepositions was that it "does not usually omit its object, as an ordinary transitive verb when the object is in the near context." Because it is the case that coverbs have the function of verbs in Chinese, some Chinese grammarians prefer to use coverbs rather than prepositions. A comprehensive explanation of coverbs may be found in Ross (1984:14) who says "coverbs in Mandarin are thus semantically equivalent to English prepositions, but syntactically equivalent to verbs."

Dow (1983:73) separated coverbs into seven categories as shown in Table 3.1.

Table 3.1 Category of coverbs

<table>
<thead>
<tr>
<th>Categories</th>
<th>Coverbs</th>
<th>English equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Causes, Reasons</td>
<td>由於, 因為</td>
<td>because</td>
</tr>
<tr>
<td>II. Objectives Motives</td>
<td>為了, 為著</td>
<td>for</td>
</tr>
<tr>
<td>III. According to Dependent on</td>
<td>根據, 據, 親, 按照, 擬, 依照, 依, 以</td>
<td>according to depend on</td>
</tr>
<tr>
<td>IV. To hint Related to</td>
<td>把, 運</td>
<td>take, hold</td>
</tr>
<tr>
<td>V. Object</td>
<td>跟, 同, 和, 與</td>
<td>and</td>
</tr>
<tr>
<td>Comparison</td>
<td>比</td>
<td>compare with</td>
</tr>
<tr>
<td>VI. Passive</td>
<td>被, 叫, 達, 給</td>
<td>by</td>
</tr>
<tr>
<td>VII. Locative constructions</td>
<td>願(著), 沿(著)</td>
<td>along facing</td>
</tr>
<tr>
<td></td>
<td>朝, 同</td>
<td>from</td>
</tr>
<tr>
<td></td>
<td>往, 從, 由, 目</td>
<td>at</td>
</tr>
<tr>
<td></td>
<td>往, 於, 盡</td>
<td></td>
</tr>
</tbody>
</table>
3.3 Case Grammar in Chinese

The theory of case grammar, as mentioned in the previous chapter, is proposed for language universals, and not particularly just for some languages like English. Because of this, case grammar can be applied in Chinese. Investigations of cases in Chinese have been made by Chinese linguists (Li, 1971; Teng, 1975; Starosta, 1985). Evidence has shown that cases exist in Chinese, as seen in the following examples. In order to be consistent, cases are chosen in Chinese that are similar to those given in section 2.2.1.2.3, namely those based upon Fillmore's (1968, 1971) cases but the definitions will not be repeated.

(1) Agent (Agt)

Example 3.3 王先生在跑步
Mr Wang tsai jogging
Agt DUR Action
(Mr Wang is jogging.)

Example 3.4 树被风吹倒了
tree by wind blow down le
Pat Cov Agt Action CRS
(The tree was blown down by wind.)

(2) Instrument (Inst)

Example 3.5 電子信件使用在圖書館
electronic mail use in libraries
Inst Action Cov Pat
(Electronic mail is used in libraries.)
(3) Experiencer (Exper)

Example 3.6 張三 喜歡 跑車
Chang San like sports cars
Exper Action Pat
(Chang San likes sports cars.)

(4) Patient (Pat)

Example 3.7 門破了
door break le
Pat Action CRS
(The door was broken.)

Example 3.8 橋被洪水沖垮了
bridge by flood crush down le
Pat Cov Agt Action CRS
(The bridge was crushed down by flood.)

(5) Goal

Example 3.9 他怕考試
he frighten examination
Exper Action Goal
(He is frightened of examinations.)

(6) Location

Example 3.10 英國的羊毛生產
Great Britain de wool produce
Location GEN Object Action
(The production of wool in Great Britain.)
(7) Benefactive (Ben)

Example 3.11 他 賺錢 為 了 孩子
he make money for le children
Agt Action Cov CRS Ben
(He has made money for his children.)

(8) Comitative (Com)

Example 3.12 空氣 和 水
air and water
Com Cov Com
(Air and water.)

(9) Time

Example 3.13 二次 世界 大 战, 1939-1945
Word War II Time
(World War II, 1939-1945.)

3.4 Syntactic Differences Between Chinese and English

The components of English vocabulary are 26 letters of the alphabet. In Chinese, on the other hand, characters are made up of strokes and radicals. The major difference between English and Chinese is the alphabetic of the former and ideogram combinations of the latter (Herdan, 1964). Additionally, the syntactic differences between Chinese and English that are involved in PRECIS index entries are: word order, spacing, passive construction, coverbs versus prepositions, comparative construction, plural, and capitalization.
(1) Word Order

The typology of the word order can be separated into three types according to Greenberg's (1961) analysis, namely VSO (verb-subject-object), SVO (subject-verb-object), and SOV (subject-object-verb). The structure of an English sentence belongs to SVO because the verb normally follows the subject and precedes the object. It is still a controversial aspect whether the dominant word order of Mandarin is SVO or SOV. Three reasons were given by Li and Thompson (1981) why Mandarin is difficult to classify regarding word order. These are:

(1) The notion of subject is not a structurally well-defined in Mandarin. As mentioned above, Chinese is a topic-prominent language.

(2) The order in which basic words and phrases occur is governed to a large extent by considerations of meaning (i.e. semantic factors) rather than of grammatical functions. This means that a Chinese sentence with verbs can be found at the beginning, in the middle, and at the end, as seen two examples below:

Example 3.14

(a) 巴士 來 了
   bus come CRS
   (The bus has come.)

(b) 來 了 巴士 了
   come PFV bus CRS
   (One bus has come.)
The example (a) can be interpreted as 'The bus(es) which you and I are expecting has/have come'. In other words, this example is interpreted as definite (known to both the speaker and the hearer), while the example (b) is interpreted as indefinite (not known to the hearer at least; hence 'one bus(es)').

(3) Mandarin is inconsistent that as to whether VO or OV order is dominant. For example, sample texts reveal a greater number of VO than OV sentences, yet ordinarily, object may precede verbs (OV).

Miao's (1981:116) experiment proved Li and Thompson's viewpoint:

"The results of the present study showed that Chinese subjects rely more heavily on a semantic factor - lexical meaning - than on word order in interpreting Chinese utterances... Therefore, Chinese allows considerable variation in word order and SVO order is not so strict as it has been assumed."

In summary, English is easier to classify with regard to the typology of the word order, i.e. SVO. On the contrary, Chinese allows considerable variation in word order. The prevalence of SVO or SOV in Chinese will depend upon the semantic considerations in a particular sentence.

(2) Spacing

The smallest unit in English writing is a letter, then a word, and finally a sentence. Two kind of spaces are provided in a sentence, i.e. "narrow spaces between letters (at least in print), and wider spaces between words" (Halliday, 1985: 2).
The latter are used to mark the separate words. The relationship of these units can be diagrammed as below in Figure 3.1.

```
\[ \text{John wrote a letter} \]
```

Figure 3.1 The units of written English

The units of Chinese language are different from English. The smallest unit in Chinese is a stroke, then a character, and finally a sentence. Each Chinese character is equivalent to a word in English, and it can be separated without character boundaries. Chinese characters are shown in a sentence, as seen in Figure 3.2.

```
\[ \text{約翰寫了一封信} \]
```

Figure 3.2 The units of written Chinese

(3) Passive construction

In Chinese, the original usage of passive is limited and not all active forms can be changed to passive forms (Wang, 1957). The passive is regarded as unfavourable in meaning in Chinese. For example, 'damaged by', 'scolded by', and 'stolen by'.
'stolen by'. The coverbs 被 'bei', 受 'shou', 给 'kei', 叫 'chiao', and 讓 'jang' are used in the passive sentences but the most commonly used is 被 'bei' (Li & Thompson, 1981). However, the usage of passive in Modern Chinese has widened because it has been influenced by Western languages, notably the impact of European languages through translation since early this century (Wang, 1947,v.2).

The structure of English passive (as mentioned in section 2.2.1.1) is shown below:

\[ NP2 \rightarrow Aux + be + en - V - by + NP1 \]  
(Chomsky, 1957:43)

Example 3.15 The window was broken by the thief.

NP2 Aux V by NP1

Chinese passive structure is different from English, in that VP is moved to the end, and the coverb 被 'bei' is placed between NP2 and NP1 as seen below:

\[ NP2 \rightarrow 被 + NP1 + VP \]  
(Hashimoto, 1969)

bei

Example 3.16 窗户 被 小偷 打破

NP2 bei NP1 V

window by thief break

(The window was broken by the thief.)

The differences in the passive between English and Chinese can be summarized in two respects. Firstly, the scope of usage in English is wider than Chinese. Chinese passive sentences normally have unfavourable meaning, but passive sentences are
commonly used in English (Zhou and Feng, 1983). Secondly, the structure of passive is different between these two languages, as shown above.

As stated in the previous chapter, a PRECIS' string tends to favour the passive voice. The construction of syntactic order in PRECIS is Patient-Action-by-Agent. On the other hand, the structure of Chinese passive is Patient-被(bei)-Agent-Action (Chor, 1986).

(4) Coverbs versus Prepositions

Chinese coverbs semantically are regarded as English prepositions, but syntactically as verbs (as mentioned in 3.2(4)). Tai (1982:500-501) showed in a detailed way the differences between coverbs and prepositions. He analysed four characteristics of English prepositions that do not occur in Chinese coverbs:

1. English prepositions are clearly distinguished from verbs, whereas Chinese coverbs are not.

2. When an English preposition is derived from an underlying predicate, it changes its morphological form and categorial membership. In contrast, Chinese coverbs are morphologically identical to corresponding verbs. For example, the instrumental expressions uses 'to use' and 'with' in English, while Chinese has only one morpheme 用'to use' for both English equivalents.

3. An English preposition can have different meanings which carry different constituent relations within a sentence, but a Chinese coverb cannot, as seen in Example 3.17.
Example 3.17

a. I ate steak with William = 我跟威廉一起吃牛排
   (here 'with' is locative-prepositional)

b. I ate steak with chopsticks = 我用筷子吃牛排
   (here 'with' is instrumental)

(Tai, 1982)

As seen in the above examples, the preposition 'with' can be used in two ways in English, but Chinese must use two different coverbs (跟, 用) instead.

4. English prepositional phrases can be defined with regard to transformations such as preposing and others, there seems little justification in Chinese. Thus, in this language, both verbal phrases and co-verbal phrases can be topicalized, as seen in Example 3.18 (Tai, 1982).

Example 3.18

a. 我在圖書館念書 = I am studying in the library.

b. 在圖書館我念書 = Whenever I am in the library, I study.

   c. 念書我在圖書館 = Whenever I study, I do it in the library.

The above Chinese examples can be interpreted as a dialogue as follows:

Questions                                            Answers

a. What are you doing in the library?                I am studying in the library.

b. What do you do when you are in the library:
   I study.

c. When you study, where do you do it?              In the library.
This also supports the earlier statement that the word order is governed by semantics rather than syntax in the Chinese language.

(5) Comparative Construction

In English, a pattern can be used for comparing two equal things as seen below. The preposition, 'compared with' can be placed in two terms X and Y.

\[
X \text{ compared with } Y
\]

In Chinese, one pattern is used for comparing two equal things as seen below (Chao, 1968; Henne, Rongen and Hansen, 1977:157; Li and Thompson, 1981:565):

\[
\text{NP1 凭 } \text{NP2 一様 } \text{dimension with same}
\]

Example 3.19 李先生跟王先生一様高
(Mr Lee with Mr Wang same tall)

In Chinese comparative construction only subjects can be compared (Li and Thompson, 1980). Therefore, NP1 and NP2 are limited to the subject. The coverb 比 'bi' was a verb in Ancient Chinese which means 'compare'. This coverb is still used as a verb in Modern Chinese (Li and Thompson, 1980). Chao (1968:341) stated that "the verb 比 'compare' is often used as a main verb in the second position." For instance, 這個跟那個比 'This compared with that'. When the verb 比 or 比較 'compare' is used at the end of a sentence, a comparative construction
can be applied as below (Arlotto 1975:90; Li and Thompson, 1980: 238):

\[ \text{NP1} + \text{coverbs} + \text{NP2} + \text{比(比較)} \]

(compare)

Example 3.20  張先生 跟 王先生 比
Mr Chang with Mr Wang compare
(Mr Chang compared with Mr Wang.)

As seen in the above pattern, the Chinese comparative structure is different from English. The connective coverbs 與, 和, 跟 and 同 'and' are used to link two terms rather than a prepositional phrase 'compared with' in English. The verb 比(較) 'compare' is placed at the end of a sentence.

In addition to subjects, Li and Thompson (1980) observed that certain topics can be compared, but this is restricted to topics which are in some sense "possessors" of the subjects, e.g. 'noses' with respect to 'elephants', and 'political system' with respect to a country as seen in Example 3.20.

Example 3.21
日本政治制度與韓國政治制度比較

------------------

Japan political system with Korea political system compare
Subject Topic Cov Subject Topic V
(Japan's political system compared with Korea's political system.)
(6) Plural

English is an inflectional language as mentioned above. Part of its morphology involves inflectional affixes, for instance, the noun 'boy' affix -s forms plural form 'boys'. Furthermore, inflection of the plural can be divided into two types, i.e. regular and irregular plural. For example, 'girl' changes to 'girls' in the regular inflection of plural, and 'child' changes to 'children' in the irregular inflection of plural.

In English-speaking countries, indexing languages tend to prefer plural forms whenever possible (Hutchins, 1975:22). The choice of singular or plural depends upon the characteristics of nouns, e.g. "the plural form is used for terms denoting discrete objects (countable), the singular for abstract and collective terms and verbal nouns (non-countable)" (British Standards Institution, 1988:4). A detailed distinction between singular and plural is seen in Table 3.2.
Table 3.2: Guidelines to singular versus plural forms in English usage

<table>
<thead>
<tr>
<th>Class of concept</th>
<th>Singular (Examples)</th>
<th>Plural (Examples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>abstract entities</td>
<td>responsibility</td>
<td></td>
</tr>
<tr>
<td>activities</td>
<td>management, teaching</td>
<td></td>
</tr>
<tr>
<td>collective terms</td>
<td>food, weather</td>
<td></td>
</tr>
<tr>
<td>disciplines</td>
<td>biology, chemistry</td>
<td></td>
</tr>
<tr>
<td>physical non-countable things</td>
<td>cooper, silver, gold, steam</td>
<td></td>
</tr>
<tr>
<td>process terms</td>
<td>controlling</td>
<td>installing</td>
</tr>
<tr>
<td>proper names</td>
<td>Newton, Poll Tax</td>
<td></td>
</tr>
<tr>
<td>systems of abstract entities</td>
<td>Buddhism, Christianity</td>
<td></td>
</tr>
<tr>
<td>organs of the body</td>
<td>head, nose</td>
<td>ears, hands</td>
</tr>
<tr>
<td>properties</td>
<td>(specific terms)</td>
<td>(generic terms)</td>
</tr>
<tr>
<td>conditions</td>
<td>density</td>
<td>physical properties</td>
</tr>
<tr>
<td>characteristics</td>
<td>temperature</td>
<td>process conditions</td>
</tr>
<tr>
<td>events</td>
<td>discharges</td>
<td></td>
</tr>
<tr>
<td>occurrences</td>
<td>explosions</td>
<td></td>
</tr>
<tr>
<td>physical countable things</td>
<td>apples, books, chairs</td>
<td></td>
</tr>
</tbody>
</table>

(Hutchins, 1975; ISO, 1986; British Standards Institution, 1988)

By contrast, Chinese nouns are all uninflected (Hockett, 1958). Therefore, they have no singular or plural distinction, though the meanings of singular or plural are expressed by the numeral and classifier added before the noun. For example, 两个苹果 (two apples), 三张桌子 (three tables), the 两 'two' and 三 'three' are numerals representing the number. Both 个 'ke' and 张 'chang' are classifiers, which represent the unit of things.
Although the numeral and classifier can be used to express plural forms, Chinese indexing terms do not consider the singular or plural forms of nouns.

(7) Capitalization

Both upper and lower cases are used in English writing systems. By contrast, Chinese has no upper or lower case difference. For example, in the English name, Charles Robert Darwin, translation of this name into Chinese is, 查理 曼伯 达尔 文, there are no upper or lower case differences.

In summary, four Chinese syntactic features were emphasized in this chapter, i.e. word order, topic-prominent, non-inflectional, and coverbs. In addition, it has been proved that the case grammar can be applied in the Chinese language.

Finally, the syntactic differences between Chinese and English are found in: (1) word order, (2) spacing, (3) passive construction, (4) coverbs versus prepositions, (5) comparative construction, (6) plural, and (7) capitalization.
REFERENCES


CHAPTER 4
COMPUTER PROCESSING OF CHINESE DOCUMENTS

The previous two chapters have described the development of PRECIS and the characteristics of Chinese syntax. This chapter moves on to the current development of the computer processing of Chinese documents.

The Chinese language originated from China; there are many other countries who use this language as well. These countries include mainland China, Taiwan, Hong Kong, Macao, and Singapore. Also, two neighbouring countries, Japan and Korea, use thousands of Chinese characters that were borrowed from China. Chinese characters in Japanese are known as Kanji and include 6,349 Kanji characters used in the Japanese Information Standard (Chow, 1983). In Korean, Chinese characters are called Hanja and include 5,000 characters currently in use (Hyeon, 1987).

In addition, there are many collections of Chinese documents in North America and Western Europe (Ma, 1981).

4.1 Input Methods

Chinese characters have special features, such as a large number and great diversity of characters. Therefore, various input methods have been devised.

There are about 700 input methods available (Meng, 1990), but only about 50 methods have been developed commercially by
computer vendors (Yang, 1987). In mainland China, nearly 50 input methods are commonly used and those can be divided into three categories: those which (1) use the Pinyin system, (2) divide up the form of the character, and (3) combine the use of phonetics and form (Meng, 1990). At the present, the most popularly used input method in mainland China is 'Wu Pi Tzu Hsing' (五筆字型) which comes under category (2). Of the Chinese computer users in mainland China, 83% of users using this input method (Meng, 1990). In Taiwan, there are five methods that are popularly used, i.e. (1) the Tsang Chieh Input Method (倉頡輸入法), (2) the Mandarin Phonetic Symbol (注音), (3) the Simplified Method (簡易), (4) the Three-corner Coding Method (三角號碼), and (5) the Telegraph Input Method (電信輸入法) (Central Daily News, 1988). Amongst these five input methods, the Tsang Chieh Input Method is the most popular and also the fastest method. It can achieve an input of 57.1 Chinese characters per minute (Kung and Chen, 1983).

Various input methods need different keyboards for inputting Chinese characters. So far, there are three kinds of keyboard used (Chen, 1983; Fung, 1983):

(1) the QWERTY keyboard: the standard keyboard which is also called ASCII (American Standard Code for Information Interchange) keyboard;

(2) medium keyboard: the number of keys is 60 - 100;

(3) large keyboard: the number of keys are over 100, and is mainly used for inputting individual characters by a touch-sensitive tablet. Each key is represented by a small rectangular area which can be touched with a pen or a finger (Suchenwirth, et al., 1989).
Both the QWERTY and the medium keyboards are commonly used. However, the large keyboards is only used extensively in Japan.

In addition to the traditional input method, more advanced input methods are being developed, such as OCR (Optical Character Recognition) and voice recognition. An experiment recently conducted in Japan with characters recognition, showed a recognition rate 99% with a character set of 4,000 Kanji (Hsieh, 1986). In mainland China, an OCR system, 'Printed Chinese Characters Recognition System', has been developed by the Peking Information Technology College in 1988; it can recognize nearly all the characters of mainland China's national standards. Recognition rate has reached over 99%. The speed of recognition is 11 characters per second (Hsin, 1988).

In West Germany, a research team including both Chinese and German researchers at the Technical University of Berlin has completed an OCR project on Chinese characters. The total recognition rate of this system is 98.21% (Suchenwirth, et al., 1989). In Taiwan, a new OCR system has been developed in 1989. This system can recognize various fonts of Chinese characters, including hand-writing characters. This system has machine learning capability. The speed of recognition is over 400 characters per minute (Central Daily News, 1989b; 1989c).

In mainland China, the Ching Hua University has successfully implemented a voice input system which has a recognition speed of 20 characters per minute. The recognition rate of this system has reached 90% (Liu, 1987). The 'BIT Voice Processing System', designed by the Peking Information Technology College in 1988, which has achieved approval at government level in mainland China (Hsin, 1988).
In Taiwan, the National Cheng Kung University has developed a speech recognition system with the learning ability for Mandarin syllables. The recognition rate of this system has reached to 96.4% (Wang, et al., 1988). In addition, there are some other institutions that have also implemented speech recognition, such as the National Ching Hua University (Chen, et al., 1988), the National Taiwan University as well as the Academic Sinica (Hsieh, Tseng and Lee, 1988), and the Telecommunication Laboratories of the Ministry of Communications (Chang and Jeng, 1988).

The input methods, both conventional and advanced input methods, can be grouped into two categories: by shape and by sound, as seen in Table 4.1 (Suen, 1984).

<table>
<thead>
<tr>
<th>Input method</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td></td>
</tr>
<tr>
<td>numerical corner system</td>
<td>Three-corner Coding System (三角號碼)</td>
</tr>
<tr>
<td>radical-based coding system</td>
<td>Tsang Chieh (倉頡) Simplified Method (簡易)</td>
</tr>
<tr>
<td>characters</td>
<td>OCR systems</td>
</tr>
<tr>
<td>Sound</td>
<td></td>
</tr>
<tr>
<td>phonetics</td>
<td>Pinyin, MPS (注音), Wade-Giles</td>
</tr>
<tr>
<td>voice recognition system</td>
<td>BIT Voice Processing</td>
</tr>
</tbody>
</table>

The table shows that three kinds of shaped input method are used: (1) the numerical corner coding system, (2) the radical-based coding system, and (3) characters. For example, the Three-corner Coding Method belongs to the numerical corner coding system, the Tsang Chieh Input Method as well as the
Simplified Method belong to the radical-based coding system, and OCR systems belong to characters.

There are two phonetic input methods that are popularly used, i.e. Mandarin Phonetic Symbol used in Taiwan, and Pinyin system used in mainland China. In addition, the Wade-Giles system is used in the Library of Congress of the United States (Cheng, 1989). Both the shaped input system and the sound input system can be combined to form other systems. The majority of commercial Chinese systems normally offer four or more input methods, such as radical input methods, phonetic methods, and so on.

4.2 Internal Codes

The character code used for the alphanumeric writing system is one-byte code, for example, one seven-bit byte of ASCII allows $2^7 = 128$ unique combinations. Similarly, one 8-bit byte used in EBCDIC (Extended Binary Coded Decimal Interchange Code) can generate $2^8 = 256$ unique combinations. One-byte code is inadequate to represent Chinese characters due to the multiplicity of Chinese characters. As a result, at least two-byte code needs to be used for Chinese characters. Two 7-bit bytes can generate $2^7 \times 2^7 = 16,384$ characters, and two 8-bit bytes generate $2^8 \times 2^8 = 65,536$ characters.

Many countries were involved in the development of Chinese internal code. These include mainland China, Canada, Japan, Korea, Taiwan and the United States. Japan was the first country to develop a set of Chinese internal code using two bytes in 1978 (Clews, 1988; Takahashi and Kanaka, 1981). The other countries mostly followed the Japanese two-byte standard.
but two particular systems: CCCII (Chinese Character Code for Information Interchange) and REACC (RLIN East Asian Character Code) adopted a three-byte standard.

The structure of two-byte internal code is constituted by the first byte (high byte) and the second byte (low byte) (as shown in Appendix 4.1). Both CCCII and REACC use three-byte code. CCCII is intended to include all existing Chinese characters, i.e. an estimate of more than 80,000 characters (Hsieh, et al., 1981). The number of Chinese characters exceed the capacity of two-byte code (65,536). Therefore, a three-byte seven-bit code has been designed in accordance with ISO (International Organization for Standardization) standards (646 and 2022), which is a three-dimensional 94x94x94 structure, i.e. 94 positions in 94 sections in 94 planes (as shown in Appendix 4.2) (Fung, 1983). Total coding positions of this code are 830,584, which includes: planes 1 to 8 for more than 50,000 Chinese characters; planes 9 to 64 for variations and simplified forms; and planes 65 to 94 for users's option, e.g. Manchurian, Mongolian, Tibetan, Japanese Kanji or Korean Hanja, and so on (Wang and Dunn, 1982).

REACC was developed by the RLG (the Research Libraries Group, Inc.) for the RLIN (Research Libraries Information Network) and was regarded as "a database management oriented approach to the East Asian character set problem" (Haeger, 1987:159). This code integrated standards from four countries, i.e. mainland China: GB2313-80; Japan: JIS C6226; Korea: KIPS (Korean Information Processing System); and Taiwan: CCCII.
In Canada, UTLAS has created a Chinese characters database, i.e. CCRF (Chinese Character Cross Reference File), which includes various code values, filing values and phonetic forms for each character. This database includes three different character sets: the BIG-5, IBM 5550 Chinese character set and CCCII (Cain, 1990).

The development of internal codes in Taiwan has been somewhat different. Here three separate internal codes have been developed respectively, i.e. CISCII (Chinese Industrial Standard Code for Information Interchange), BIG-5, and CCCII. CCCII has been mentioned above, CISCII was developed in 1981, and includes approximately 16,000 characters. This code was superseded by SICGCC (Standard Interchange Code for Generally-used Chinese Characters) in 1983 and revised in 1986. Meanwhile, this code was made a national standard in Taiwan (Clews, 1988). BIG-5 Code was developed by the Institute for Information Industry, which is mainly used in five applications packages (that are developed by local computer vendors). Within this code, the 'commonly used code' (5,401 characters) is exactly similar to SICGCC, whereas two characters more than CISCII in 'less commonly used code', i.e. 7,652 vs 7,650 (ETien Information Company, 1987). This code is popularly used by several microcomputer software systems (Tai, 1987).

In addition to those three internal codes, some vendors have developed their own Chinese internal codes, e.g. IBM 5550 Chinese character set and WANG's Three-corner Code. A detailed picture of these internal codes is seen in Table 4.2.
Although several internal codes exist in various countries, Chinese data cannot be communicated through those codes because there is no international standard for Chinese information interchange. The ISO provides a standard for control codes of alphanumeric text information; however, CJK (Chinese, Japanese, Korean)
and Korean) data transmission requires additional control codes for a more complex character set (Yamada, 1984). Adoption of CCCII might be a choice to communicate those internal codes because version three of this code includes nearly all of Chinese characters and those standards used in mainland China, Japan, Korea and Taiwan. In addition, this code provides a mechanism of interchange amongst those character sets (Tseng, et al., 1987). Therefore, CCCII is suitable for multilingual collections: Chinese (both full traditional characters and simplified characters), Japanese (Kana and Kanji) and Korean (Hangul and Hanja).

CCCII was adopted as a national standard of the United States in 1989. Nevertheless, this code is used by the Library of Congress, OCLC (Online Computer Library Center), WLN (Western Library Network), and RLIN (Central Daily News, 1989a). In mainland China, a new version of internal code has been developed which is known as GB2901. This code has completed the mapping between GB code and CCCII (Cheng, 1989). Moreover, UTLAS' CCRF is based on CCCII to mapping other codes, e.g. BIG-5 and IBM 5550 Chinese Character Set. Consequently, adoption of CCCII might be a solution for mapping internal codes amongst those countries.

4.3 Output Devices

There are two basic types of output, i.e. character and voice. Character is mainly considered in this section, which will be further divided into printout and display (Suen, 1984). Normally Western characters dot matrix, 7 x 9, is used for printout. Because of the complexity of Chinese characters, this
configuration is not sufficient (Chen, 1983). The density of characters is determined by the resolution of dot matrix, i.e. the number of dots per inch, such as 16 x 16, 24 x 24, 28 x 28, 32 x 32, and 48 x 48. These are used in Chinese characters (Fung, 1983). The greater the numbers of dots, the greater the density of characters. For example, if the number of dot matrix is over 24 x 24, complicated strokes of Chinese characters can be displayed clearly. If the number of dot matrix is over 32 x 32, different fonts of characters can be displayed, e.g. Ming font (明體), Sung font (宋體), or Kai shu (楷書) (Chen, 1983). With a non-dot matrix printer, i.e. laser printer, the result is even more satisfactory. For example, laser printers have high resolution which includes over 300 dots per inch, compared with other systems that have about 200 dots per inch (Suen, 1984).

The principle of display for Chinese characters on CRT (Cathode-ray tube) is similar to the printout. The display unit must be addressable, i.e. it should have graphic capability. The number of characters that can be displayed on a screen depends on both the resolution of the screen and the size of the characters' dot matrix. Different kinds of dot screens are used for displaying Chinese characters, including 640 x 390, 640 x 408, 640 x 416, 720 x 348 and 720 x 352 dots (ETien Information Company, 1987).

4.4 Hardware and Software

The hardware used in the Chinese system in the early stages were minicomputers and mainframes, e.g. HP, IBM and WANG. With the development of more powerful microprocessors in 1980s,
Chinese systems have been installed in microcomputers. For the purposes of this thesis, microcomputer applications are chiefly considered.

Four methods can be used for storage of Chinese characters in microcomputers: (1) floppy disks; (2) RAM (Random Access Memory); (3) RAM/floppy disks; and (4) ROM (Read-Only Memory), as seen in Table 4.3 (Tai, 1987:39).

Table 4.3: Four types of storage for Chinese characters

<table>
<thead>
<tr>
<th>Types of storage</th>
<th>Descriptions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floppy disks</td>
<td>- needs a disk drive;</td>
<td>Haio Shent'ung 2100 (小神通 2100)</td>
</tr>
<tr>
<td></td>
<td>- access Chinese characters from floppies</td>
<td></td>
</tr>
<tr>
<td>RAM</td>
<td>- all Chinese characters are stored on floppies;</td>
<td>ET 2416F</td>
</tr>
<tr>
<td></td>
<td>- boot the system, then load in RAM</td>
<td>Kuoch'iao Hantieh 500 (國喬漢碟 500)</td>
</tr>
<tr>
<td>RAM/floppy disks</td>
<td>- store all of Chinese characters on floppies or on the hard disk;</td>
<td>IBM 5550</td>
</tr>
<tr>
<td></td>
<td>- boot the system, commonly used characters (about 2,000) are loaded to RAM;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- if characters cannot be found in RAM, then access to floppies or the hard disk</td>
<td></td>
</tr>
<tr>
<td>ROM</td>
<td>- all of Chinese characters are stored in ROM</td>
<td>T'ienlung 570</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ET 2416</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chingyeh 2100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(精業 2100)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ts'angchien 2025-G</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(倉頡 2125-G)</td>
</tr>
</tbody>
</table>

Installation of the Chinese system requires a Hercules card in a monochrome monitor, or an EGA card in a colour monitor. For the software, a microcomputer operating system, CCDOS (Chinese Character Disk Operating System), has been developed in
mainland China, which is mainly based upon MS-DOS. This software incorporates MS-DOS, and thus only one disk is necessary for its operation (Wu and White, 1990). Different Chinese systems have been developed in Taiwan, where most Chinese operating systems are under MS-DOS, i.e. it is necessary first to load MS-DOS, then load the Chinese system (as seen in Appendix 4.3). However, a new attempt to standardize Taiwan's software industry has been carried out by the Microsoft Corporation in November 1988. An agreement was signed by the Microsoft Corporation with five leading hardware companies: Mitac, Acer, Copam, Tatung and Datatech to develop a Chinese version of DOS which was known as C-DOS (Besher, 1989). Most Chinese systems can be used on IBM PC and compatible microcomputers.

Although commonly used software packages can also be used in Chinese systems, such as dBASE III, LOTUS 1-2-3 and PE II (Personal Editor II), these systems are not 100 per cent compatible with Chinese systems because of the different nature of languages. Some programs have to be adjusted when Chinese data are processed. For example, two or three bytes are used to represent the Chinese character, therefore, the program should avoid splitting Chinese characters at the end of a line (Cheng, 1989).

Additionally, some Chinese desk top publishing packages have been developed in Taiwan, e.g. Pot'ai (柏泰), Pisheng (華昇), Hanyun (翰芸), and Fujungfang Yamo (芙蓉坊飛墨) (Central Daily News, 1990).
4.5 Computer Processing of Chinese Documents in Libraries and Information Services

Chronologically, the National Diet Library of Japan had initially started its library automation project using a computer system to process Japanese bibliographic data in 1971, in this case Kanji (Takahashi and Kanaka, 1981; Maruyama, 1987). Computerization of the "Union List of Scientific Serials in Libraries of the Republic of China," was carried out by the Science and Technology Information Center of Taiwan in 1973. Since then, developments have occurred in various institutions of Taiwan, such as the National Chungshan Institute of Science in 1974, the National Taiwan Normal University in 1978, the National Central Library as well as the Agricultural Science Information Center in 1979 (Hu, 1983). The first edition of the Chinese MARC format was published in 1981. Based on the Chinese MARC format, the National Central Library has been developing a computerized library system since 1981 (Wang, 1987). In 1989, the National Central Library successfully developed the Chinese MARC Database on CD-ROM (Chou, Chiang and Wang, 1989).

The development of automation in mainland China started in 1978, with the establishment of "automation of key research centers" and "national networking" as top priority goals of China's modernization of science and technology. Significant progress had been made in the processing of Chinese documents by 1985 (Maier, 1986). A Chinese MARC format, the PUL MARC
format, was published by the Peking University in 1985 (Cheng, 1989). The National Library of China has currently undergone automation (Maier, 1987). Finally, the Central National Library of Korea had started automation in 1981 with a computer which was capable of processing 5,000 Chinese characters (Hyeon, 1987).

In addition to the above four countries, computer processing of CJK documents has developed in the United States, particularly at the Library of Congress, RLG (RLIN) and OCLC. The first CJK production had been installed and connected to RLIN in 1983 (Haeger, 1983). In the same year, the Library of Congress entered the first CJK record into the RLIN database. Twenty two institutions, where collections form approximately 70 per cent of the East Asian holdings in North America, used RLIN for processing CJK documents by 1986. CJK terminals have been installed in many countries, including the United States, Canada and the United Kingdom (Haeger, 1987). OCLC started the automation of CJK documents in 1983. The special OCLC M300 Workstation is used for input/output of CJK documents. Cooperation between OCLC and the National Central Library (Taiwan) was set up in which the Library's Chinese records could be entered into OCLC's database (Wang, 1985; Lee, 1988). In addition, UTLAS has installed a Chinese CATSS (Cataloguing Support Service) at the National Central Library (Taiwan) in 1989 (Cain, 1990). Details of these databases are seen in Table 4.4.
### Table 4.4: Databases of Chinese documents

<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Database</th>
<th>Format</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainland China</td>
<td>National Library of China</td>
<td>Not yet available</td>
<td>PUL MARC</td>
<td></td>
</tr>
<tr>
<td>(Cheng, 1989)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>National Diet Library</td>
<td>946,353 records</td>
<td>Japan MARC</td>
<td>February 1990</td>
</tr>
<tr>
<td>(Imon, 1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Korea</td>
<td>Central National Library</td>
<td>70,151 titles</td>
<td>KOR MARC</td>
<td>1987</td>
</tr>
<tr>
<td>(Hyun, 1987)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>National Central Library</td>
<td>223,718 titles of books 17,819 titles of rare books</td>
<td>Chinese MARC</td>
<td>30th May 1990</td>
</tr>
<tr>
<td>(Chiang, 1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S.A.</td>
<td>OCLC</td>
<td>169,603 Chinese records</td>
<td>US MARC</td>
<td>8th April 1990</td>
</tr>
<tr>
<td>(Hurley, 1990; Washburn, 1990)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RLIN</td>
<td>598,005 CJK records ca. 260,000 Chinese records</td>
<td>US MARC</td>
<td>30th April 1990</td>
</tr>
</tbody>
</table>

In total, over 2,000,000 CJK records have been created by various countries and institutions. Most of the databases are created in MARC format, for example Japan MARC in Japan, Chinese MARC in Taiwan, PUL MARC in mainland China, and KOR MARC in Korea. Most of these MARC formats are based upon UNIMARC, therefore, the representations of tags and fields are correspondent. Although they are not 100 per cent compatible due to the requirements of different countries, techniques can be used to transfer data between MARC formats (Cheng, 1989). For example, the National Central Library has successfully transferred data from the Library of Congress' US MARC to Chinese MARC. Furthermore, the Library of Congress has successfully completed the conversion from US MARC into UNIMARC in 1985 (Chou, 1985).
The computerization of Chinese documents has increasingly developed in the past decade. Although many countries have endeavoured to computerize Chinese records and over 2,000,000 records have been created, there still remain some obstacles that prevent full interchange of Chinese records in those countries, i.e. different Chinese internal codes and different Chinese MARC formats. In order to solve these problems, some kinds of standards or protocols have to be set up. Then, Chinese records could be fully communicated amongst those countries.

In summary, this chapter has mainly described current computer processing in Chinese, including various input methods, different Chinese internal codes, output devices, and hardware and software. This study is based on an experiment which was carried out on a microcomputer. All requirements related to the experiment, such as input method, the internal code, the output device, and hardware and software will be described in the next chapter.
REFERENCES


NATIONAL SCIENCE COMMISSION, et al. (eds.) (1986) *Standard*
interchange code for generally-used Chinese characters.


CHAPTER 5

EXPERIMENTAL METHODOLOGY

This chapter presents the various stages adopted during the experiment, the sampling procedure, and the hardware and software. Finally the limitations of the study are discussed.

5.1 Methodology

An experimental approach has been used in this study. It consisted of two stages: the pilot and the main experiments. The study experimented with a sample of 492 titles that were randomly chosen from a Chinese periodical index, and six Chinese strings that were chosen and translated from English examples recorded in the Manual (Austin, 1984) (see section 5.2). Using this sample, a comprehensive test of all the PRECIS's role operators was carried out.

Since the programs involved were mainly designed for English, a pilot experiment was needed in order to realize the process of the programs. Preparation of the experiment was the major task in the pilot experiment. It included testing the programs and setting up the formats for input and output. Some strings were chosen from the sample titles for testing. Since it was not possible to input and output the Chinese characters in the early stage of the experiment (due to the unavailability of a Chinese computer system), a romanization of Chinese strings was performed instead. The Wade-Giles system was chosen for the romanization, as this system is popularly used in Taiwan (see section 3.1.3(2)). Chinese strings were then transcribed into roman characters and were input to the computer.
It was found that the romanization was not sufficient to help the input and output process of this study. The reasons were:

1. Chinese script has three features, i.e. sounds (phonetic aspect), shapes, and meanings that are derived from the combination of the sounds and shapes;
2. the Wade-Giles system, like the other romanization systems, transcribes only the sounds of the Chinese character, but not the shapes; consequently,
3. the romanized Chinese strings and entries produced were meaningless; and it was impractical to retrace the meanings to the original Chinese subject statements (see Appendix 5.1).

Fortunately, six months after the pilot experiment was completed, the ETien Chinese System was installed, thus the main experiment could be carried out properly.

During the main experiment, the English version PRECIS's role operators were tested in indexing 492 Chinese subject statements with necessary translations, i.e. to produce Chinese strings as well as entries (see Appendix 5.2). It was found that all but three role operators and two codes had chances to be tested with the Chinese strings. The role operators that had not been tested were the operators 's' - 'participation', 't' - 'A' expounded by 'B', 'r' (aggregates and associates), and codes '$n' as well as '$0'. To test the application of the remaining role operators in Chinese texts, six example strings were chosen from the relevant examples available in the Manual (see section 5.2), were translated into Chinese, and input to the computer.
Analysis of the experimental results was based on the Chinese syntactic features and syntactic differences between Chinese and English that were presented in sections 3.2 and 3.4. Since the PRECIS programs used in the experiment were originally designed to produce English entries, and there are syntactic differences between English and Chinese; the application of the English programs in Chinese texts has lead to several problems. In Chapter 6 problems encountered are analysed. During the analysis, problems were grouped according to the subject categories, that is, to define whether or not problems are subject specific. The solutions to the problems are presented in Chapter 7.

The solutions were then compared with previous research (see Chapter 8). Two research projects were chosen, i.e. those of Austin (1974) and Chor (1986), since these mainly discussed PRECIS as applied in the Chinese language and suggested some solutions to the problems encountered.

Finally, the whole procedure of Chinese index manipulation was demonstrated, including input, sorting and merging, and printing out (see Chapter 9). The demonstration used some strings from the samples of the string, including all examples that were discussed in Chapter 7 and Chapter 8.

Thus, the pilot and main experiments were carried out mainly according to the following stages: (1) analysis of subject statements; (2) coding; (3) input, computer processing, and output; (4) analysis of the output based on Chinese syntactic rules; (5) modifications of role operators and the accompanying programs; (6) comparison of the findings of the present and
previous research; and (7) demonstration of Chinese index manipulation. These stages are illustrated in Figure 5.1.

| (1) Analysis of subject statements |
| (2) Coding |
| --- Input of Chinese strings |
| (3) Computer processing |
| --- Output of Chinese entries |
| (4) Analysis of the output based on Chinese syntactic rules |
| (5) Modifications of role operators and the accompanying programs |
| (6) Comparison of the findings of the present and previous research |
| (7) Demonstration of Chinese index manipulation |

Figure 5.1 Seven stages of the experiment
5.2 Sampling

The Index to Chinese Periodical Literature was purposely selected since it has a complete subject coverage. One issue (i.e. vol.18, no.1, May 1987) was then randomly chosen from this periodical. The Index is published quarterly by the National Central Library in Taiwan. All issues of this periodical have the same format, and the number of entries ranges from 5,000 to 6,000.

A systematic sampling (Powell, 1985) procedure was executed to choose the titles to be included in the experiment. The sampling ratio was 1:10. Out of 5050 titles listed in the above issue, 505 titles were chosen as the test data. It was found that there were 492 titles in Chinese, and 13 in English. Since the present study is aimed to experiment with Chinese documents, the 13 English titles were excluded from the present study. All subjects covered in the Index were included in the experiment. The sample titles (492) represented about 10% of the titles listed in each category.

As mentioned before, in order to make a comprehensive test of all of the PRECIS's role operators, the experiment included also six strings that were taken from the Example 4 of 's' - 'participation' (Austin, 1984:149), Example 5 of 't' - 'A' expounded by 'B' (Austin, 1984:165), Example 16 of 'r' - 'aggregates' (Austin, 1984:87), Example 20 of 'r' - 'associates' (Austin, 1984:88), an example of the use of code '$0$' (Austin, 1984:68), and an example of the use of '$n$' (Austin, 1984:69). These examples were chosen randomly amongst the available examples, and were translated into Chinese. They
were considered as belonging to subject categories 'Generalities' (1 string), 'Philosophy' (2 strings), 'Natural Sciences' (1 string), and 'World History, Geography & Biography' (2 strings). Table 5.1 depicts the distribution of the sample strings according to the subject categories.

Table 5.1: Subject categories of the sample

<table>
<thead>
<tr>
<th>Subject categories</th>
<th>Number of titles in the chosen index</th>
<th>Number of titles included in the sample</th>
<th>Number of titles from the Manual (Austin, 1984)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalities</td>
<td>215</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>Philosophy</td>
<td>123</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Religion</td>
<td>101</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>230</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>Applied Sciences</td>
<td>1761</td>
<td>173</td>
<td></td>
</tr>
<tr>
<td>Social Sciences</td>
<td>1608</td>
<td>157</td>
<td></td>
</tr>
<tr>
<td>Chinese History &amp; Geography</td>
<td>115</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>World History, Geography &amp; Biography</td>
<td>396</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>Language &amp; Literature</td>
<td>225</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>The Arts</td>
<td>209</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Research of China &amp; the Soviet Union</td>
<td>157</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5050</td>
<td>492</td>
<td>6</td>
</tr>
</tbody>
</table>

5.3 Hardware and Software

A microcomputer, IBM XT, was used in the experiment. A Hercules card was installed to meet the requirement for a monochrome monitor.
The software used was the PRECIS English version software written by F. Smith, an ETien Chinese System designed by a Taiwanese company - ETien Information Company, an enhancement of MS-DOS providing for Chinese input/output system, and a word processor, GALAXY. The software used for generating entries is described in Chapter 9.

The Big-5 Chinese Internal Code of 13,053 Chinese characters are used in the ETien Chinese System. This Code has a significant feature. The first byte (high byte) of Chinese characters and punctuation in the Code has values greater than 161, and thus can be used to distinguish Chinese from English during the modifications of the programs.

The Chinese character set was stored on floppy disks. In order to access Chinese characters easily, this character set was loaded into the hard disk. A phonetic input method, i.e. the Mandarin Phonetic System, was chosen as the input method since it is easy to learn and use. The output device of Chinese characters was 16x16 dot matrix font.

The algorithms for entry construction have been described in the Manual (Austin, 1984:340-386). They are not repeated in this study, but some algorithms are discussed in Chapter 7.

5.4 Limitations

As mentioned above, Chinese strings were written mostly based on the titles of periodicals, and the present study is mainly concerned with the syntactical aspects. Because of this, it is possible that a small proportion of Chinese subject indexes
produced in this study do not represent exactly the content of related documents.

So far, there is no Chinese index produced using PRECIS. Consequently, Chinese index entries in this experiment have to be evaluated without the possibility of direct comparison with any printed source.

The next chapter presents the analysis of the computer output based on the Chinese syntactic rules.
REFERENCES


CHAPTER 6
DISCUSSION OF PROBLEMS

This chapter is about the results of the computer processing, i.e. stage 4 of the experiment (see section 5.1). The aim of this chapter is to describe the problems encountered when English version PRECIS was used in Chinese. The problems were detected through the analysis of the Chinese index entries using Chinese syntactic rules; and the causes were identified through the association of the problems with the role operators used in the related strings.

In the discussion, problems are grouped according to the subject categories and causes. Those that resulted from the same operators and codes are discussed once under the subject categories that first experienced the problems. As stated previously, eleven subject categories were covered in the experiment, i.e. 'Generalities', 'Philosophy', 'Religion', 'Natural Sciences', 'Applied Sciences', 'Social Sciences', 'Chinese History & Geography', 'World History, Geography & Biography', 'Language & Literature', 'The Arts' and 'Research of China & the Soviet Union'.

In order to show the problems encountered clearly to the non-Chinese speaking readers, the Chinese examples are accompanied by the English word order from the Chinese original examples. Thus, the English entries in the examples are not the products of computer processes. In addition, spacing (see section 3.4(2)) as well as the singular-plural noun form (see section 3.4(6)) are not considered in the English examples.
A chapter summary sums up the discussion of the problems occurring across the eleven subject categories.

6.1 Generalities

6.1.1 Role operators used

In this category, 24 strings were tested. All strings have at least one operator '1'. Half of the strings (12) used operators '2' and 'p' (50%), respectively. About one-third of the strings used downward-reading connective code '$v' (37.5%), operators '0' (29.1%) and 'g' (25%). The other operators were used by less strings, as shown in Table 6.1.

In the *Manual* (Austin, 1984), operator '3' is used to indicate five types of performers or agents, i.e. (1) direct objects, (2) internal process involving intakes, (3) agents of transitive actions, (4) instruments, and (5) factors. In this thesis, unless it is specified, operator '3' represent types (1), (2), (4) and (5).

6.1.2 Problems encountered

Problems occurred in this category were related to the following role operators: downward-reading connective code ($v$), operators 'g', 's' - 'applications' and upward-reading connective ($w$), 't' - 'A' related to 'B', 't' - 'A' compared with 'B', and 't' - 'A' expounded by 'B'. They are described in Problem 1 to 6.
Table 6.1: Distribution of 'Generalities' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cents (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 24</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>7</td>
<td>29.1</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>50.0</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>21.0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>16.7</td>
</tr>
<tr>
<td>g</td>
<td>6</td>
<td>25.0</td>
</tr>
<tr>
<td>p</td>
<td>12</td>
<td>50.0</td>
</tr>
<tr>
<td>s - applications</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>2</td>
<td>8.3</td>
</tr>
<tr>
<td>- 'A' compared with 'B'</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>- 'A' expounded by 'B'</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>$x$y$z$</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>$s$</td>
<td>3</td>
<td>12.5</td>
</tr>
<tr>
<td>$d$</td>
<td>2</td>
<td>8.3</td>
</tr>
<tr>
<td>$v$</td>
<td>9</td>
<td>37.5</td>
</tr>
<tr>
<td>$w$</td>
<td>2</td>
<td>8.3</td>
</tr>
</tbody>
</table>

* Note: Many role operators were difficult to count because they were used more than once in the same string. Therefore, the above figures were counted for each string as a unit, e.g. if two operator '2's or three operator 'p's were used in the same string, they were all counted only once.

Problem 1: Downward-reading connective ($v$)

Applying the English program in Chinese strings produced unnecessary spaces within an entry when a downward-reading connective ($v$) was used. The problem is illustrated as seen in Example 6.1.1.
Example 6.1.1

Subject: 三民主義的次序和結構

(Order and structure of the Three Principles of the People)

Chinese string:

*1)*三民主義
*p)*次序$\&$
*g)*結構

Entries:

三民主義
次序 $\&$ 結構 (Problems 1 and 2)

次序. 三民主義
結構. 三民主義

English word order from Chinese original string:

*1)*Three Principles of the People
*p)*order$\&$
*g)*structure

English word order from Chinese original entries:

Three Principles of the People
Order & structure (Problems 1 and 2)

Order. Three Principles of the People
Structure. Three Principles of the People

The example shows that the use of downward-reading connective code ($v$) created two unexpected spaces in the Chinese phrase '次序 $\&$ 結構' (order $\&$ structure), and separated the phrase into three parts: '次序', '&', and '結構'. As a result, the meaning of this phrase is destroyed.

Problem 2: Operator 'g'

The ampersand '&' is used in operator 'g' to link two or more than two terms in English strings. Since the ampersand is
never used in the Chinese text, the use of this sign may destroy the meaning of the produced entry, e.g. '次序 & 結構' (order & structure).

Problem 3: Operator 's' - 'applications' and upward-reading connective ($w$)

The role defining term 'applications' is used in English as follows: 's)applications$vof$win'. The term 'applications' was translated into a Chinese term '應用'; and the preposition 'in' into '在'. Since in Chinese there is no such term as 'of', this preposition was translated into the nearest Chinese particle '的', which means apostrophe ('). The problem encountered by operator 's' - 'applications' and upward-reading connective ($w$) is shown below:

Example 6.1.2

Subject: CD-ROM 應用在CARIM TM 電腦輔助資訊檢索與導向系統

(Applications of CD-ROM in CARIM TM computer assisted information retrieval and guiding systems)

Chinese string:

*1)資訊 L0
*2)資訊檢索S$31電腦輔助$32CARIM TMSv&
*3)導向系統
s)應用$的$Sw在
*3)CD-ROM
Entries:

Information retrieval
CARIM TM™ computer assisted information retrieval. Applications' CD-ROM (Problem 3)

Computer assisted information retrieval
CARIM TM™ computer assisted information retrieval. Applications' CD-ROM (Problem 3)

CARIM TM™ Computer assisted information retrieval
Applications' CD-ROM (Problem 3)

Guiding systems
Applications' CD-ROM (Problem 3)

CD-ROM
Applications in CARIM TM™ computer assisted information retrieval & guiding systems (Problem 3)

According to the above example, the use of English algorithm 'application-of-in' in Chinese produced two phrases: (1) 應用的 CD-ROM (applications' CD-ROM), and (2) 應用在 CARIM
However, the use of the role defining term 應用 (applications) in Chinese should follow these patterns:

(i) 應用 - - (applications - -)
(ii) 應用在 - - (applications in - -)

Thus, the phrase (1) is grammatically wrong according to the above pattern, since neither a Chinese particle 的 ('') nor spaces are needed between the role defining term 應用 (applications) and another term. The problem of the second phrase is caused by the spaces produced before and after the coverb 在 (in).

In Example 6.1.2, the differencing code 'S3' was used, instead of 'S2'. The decision to use this code was taken in order to meet the requirement of Chinese compound terms, i.e. no space is needed within a compound term.

Problem 4: Operator 't' - 'A' related to 'B'

One problem was found which was caused by the operator 't' - 'A' related to 'B'. The English phrase 'related to' was translated into a Chinese term 關係. The problem is illustrated by one example below:

Example 6.1.3
Subject: 呂氏春秋與名家關係
(Lushih Chunchiu related to Minchia)
In the English program, the punctuation mark '%' is used to mark terms that will be printed in italics. Since Chinese script does not have italic form, this mark has no effect at all, but appeared in the produced entries.

In addition, the algorithm of operator 't' produced phrases that were ungrammatical, such as '呂氏春秋 關係 名家' (Lushih
Chunchiu %related to% Minchia, '關係 名家' (%related to% Minchia), and '關係 呂氏春秋' (%related to% Lushih Chunchiu). The grammatical errors were caused by the differences between English and Chinese patterns of relating two equal things. In English, the pattern is 'A related to B', whereas in Chinese it is A和B關係 (A with B relate). Consequently, the above three phrases should be changed to '呂氏春秋和名家關係' (Lushih Chunchiu with Minchia relate), '和名家關係' (with Minchia relate), and '和呂氏春秋關係' (with Lushih Chunchiu relate), respectively; otherwise they will be meaningless.

Problem 5: Operator 't' - 'A' compared with 'B'

Another pattern of operator 't' is 'A' compared with 'B'. The phrase 'compared with' is translated into a Chinese term '比較'. The problem of this operator is illustrated below:

Example 6.1.4

Subject: 三民主義和各種主義的比較  
(Three Principles of the People compared with various principles)

Chinese string:

*1)三民主義  
t)比較  
1)各種主義

Entry:

三民主義  
%比較%各種主義 (Problem 5)

English word order from Chinese original string:

*1)Three Principles of the People  
t)compared with  
1)various principles

English word order from Chinese original entry:

Three Principles of the People  
%compared with% various principles (Problem 5)
A phrase shown in the above example, i.e. \%比較\% 各種主義 (%compared with% various principles), is ungrammatical according to Chinese syntax. This is caused by the differences between English and Chinese pattern of comparing two equal things. As stated in section 3.4 (5), the Chinese patterns for comparing two equal things are:

A 和 B 比較  (A with B compare)
B 和 A 比較  (B with A compare)

Thus, the above phrase 各種主義 (compare with various principles) should be written as follows: 各種主義比較 (with various principles compare).

Problem 6: Operator 't' - 'A' expounded by 'B'

Example 6.1.5

Subject: 意義被資訊理論說明
(Meaning expounded by information theory)

Chinese string:

*1)意義
 t)$v說明被$w說明
*3)資訊理論

Entries:

意義
%說明被% 資訊理論  (Problem 6)

資訊理論
%說明% 意義  (Problem 6)

English string: (Austin, 1984:165)

*1) meaning
 t)$v expounded by$w expounding
*3) information theory

English word order from Chinese original entries:

Meaning
%expounded by% information theory  (Problem 6)

Information theory
%expounding% meaning  (Problem 6)
The first entry of the above example, i.e. '％說明被％資訊理論' (%expounded by% information theory) is unacceptable according to the Chinese passive construction (see section 3.4(3)). The Chinese passive phrase should be written like this: 被資訊理論說明 (by information theory expound). The second entry is acceptable, but one space produced by the upward-reading connective code ($w$) has to be removed between two terms 說明 (expounding) and 意義 (meaning). In addition, the punctuation mark '％' has to be removed. Two forms of the term 'expound' are used in the above example, i.e. 'expounded' and 'expounding'. As stated in section 3.2 (3), the Chinese language lacks complexity in word formation, therefore only one form of this term used in Chinese, i.e. 說明 (expound).

The summary of six problems is presented in Table 6.2.

Table 6.2 Problems encountered in 'Generalities'

<table>
<thead>
<tr>
<th>No of problems</th>
<th>Causes of problems</th>
<th>Problem descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 $v$</td>
<td>unexpected spaces in the produced Chinese entries separated a phrase into three pieces, and thus destroyed its meaning</td>
<td></td>
</tr>
<tr>
<td>2 $g$</td>
<td>the ampersand '&amp;' is never used in Chinese text, and thus the use of this sign may destroy the meaning of the related Chinese entry</td>
<td></td>
</tr>
<tr>
<td>3 $s$ - applications and $w$</td>
<td>1. Chinese phrases produced are ungrammatical as the Chinese patterns of operator 's' - 'applications' are different from the English ones 2. two unnecessary spaces were produced when the upward-reading connective was used in Chinese strings, i.e. before and after the term 佐 (in)</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
The table shows that the uses of the above operators lead to various problems, ranging from unnecessary spaces that do not cause any change of meaning to those that cause the destruction of meaning. The more serious problems are grammatical ones. Fortunately, the role operators that caused these problems were only used by a small percentage of 'Generalities' strings.

### 6.2 Philosophy

#### 6.2.1 Role operators used

Thirteen strings were tested in this category. Operators '2' and 'p' were used by 84.6% of strings, respectively. Operator
'p' was heavily used since some strings were related to 'Chinese philosophers' that were parts of 'Chinese philosophy'. The close-up differencing code '$3' was used by 61.5% of the sample strings (see Table 6.3).

Table 6.3 Distribution of 'Philosophy' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>46.2</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>84.6</td>
</tr>
<tr>
<td>3 - agents of transitive actions</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>15.4</td>
</tr>
<tr>
<td>$</td>
<td>2</td>
<td>15.4</td>
</tr>
<tr>
<td>g</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>p</td>
<td>11</td>
<td>84.6</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>$x$y$z$</td>
<td>2</td>
<td>15.4</td>
</tr>
<tr>
<td>$3$</td>
<td>8</td>
<td>61.5</td>
</tr>
<tr>
<td>$d$</td>
<td>4</td>
<td>30.8</td>
</tr>
<tr>
<td>$n$</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>$o$</td>
<td>1</td>
<td>7.7</td>
</tr>
<tr>
<td>$v$</td>
<td>3</td>
<td>23.1</td>
</tr>
</tbody>
</table>

6.2.2 Problems encountered

Two new types of problems were found in this category that resulted from the use of operator 'f' (Problem 7) and agents of transitive actions (Problem 15), respectively. Although the problem related to agents of transitive actions was firstly
encountered in 'Philosophy', it will be discussed in the subject category 'Social Sciences', since the suitable example to explain this problem is available in 'Social Sciences'. The other types of problems are the same as those problems 1, 2 and 4 mentioned above.

**Problem 7: Operator 'f'**

The use of operator 'f' brought about the same problem as that of operator 'g' (Problem 2), as seen in Table 6.4.

<table>
<thead>
<tr>
<th>No of problem</th>
<th>Cause of problem</th>
<th>Problem description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>'f'</td>
<td>the ampersand '&amp;' is never used in Chinese text, and thus the use of this sign may destroy the meaning of the related Chinese entry</td>
</tr>
</tbody>
</table>

6.3 Religion

6.3.1 Role operators used

Ten strings were tested in this category. Both operator '1' and operator 'p' were used by most strings, i.e. 70% and 60% strings, respectively. Forty percent of strings used operators '0' and '2', respectively. The other operators were used by less than 4 strings each. A complete picture is shown in Table 6.5.
### Table 6.5 Distribution of 'Religion' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>g</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>p</td>
<td>6</td>
<td>60</td>
</tr>
<tr>
<td>q</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>s - influence</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>$x,y$</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>$1$</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>$3$</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>$v$</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>$w$</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

#### 6.3.2 Problems encountered

Five types of problems were met in this category, i.e. those related to operators 'f', 'g', 't' - 'A' related to 'B', 's' - 'influence' and downward-reading connective code ($v$). All types of problem have been discussed except the problem that related to operator 's' - 'influence' which is explained in Problem 8.

**Problem 8: operator 's' - 'influence'**

The role defining term 'influence' is used in English as
follows: 's)influence$vof$won'. The preposition 'on' is translated into a Chinese coverb '在'. One example that illustrates the problem is presented below:

Example 6.3.1

Subject: 基督教内在教義對西方世界影響

(Influence of Christian spiritual doctrines on the Western world)

Chinese string:

*1)西方世界
  s)影響$v的$w在
  3)基督教內在教義  NU2
*3)基督教
*p)內在教義

Entries:

西方世界
影響的基督教內在教義 (Problem 8)

基督教
內在教義. 影響在西方世界 (Problem 8)

內在教義. 基督教
影響在西方世界 (Problem 8)

English word order from Chinese original string:

*1)Western world
  s)influence$vof$won
  3)Christian spiritual doctrines  NU2
*3)Christianity
*p)spiritual doctrines

English word order from Chinese original entries:

Western world
Influence's Christian spiritual doctrines (Problem 8)

Christianity
Spiritual doctrines. Influence on Western world (Problem 8)

Spiritual doctrines. Christianity
Influence on Western world (Problem 8)

In the above example, two Chinese phrases are ungrammatical, i.e. '影響的基督教內在教義' (influence's Christian spiritual doctrines) and '影響在西方世界' (influence on
Western world). These errors are due to the differences between Chinese and English passive construction. The patterns of operator 's' - 'influence' in Chinese are:

(i) 影響 (by - - influence)
(ii) 影響 - - (influence - - )

Thus, those phrases should be changed to:
'影響西方世界' (influence Western world) and '影響西方世界' (influence Western world), respectively.

To sum up, a new type of problem was met in this category, which is caused by operator 's' - 'influence'.

Table 6.6 New type of problem encountered in 'Religion'

<table>
<thead>
<tr>
<th>No of problem</th>
<th>Causes of problem</th>
<th>Problem description</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>s - influence</td>
<td>the passive form of Chinese entries produced are unacceptable according to Chinese syntax</td>
</tr>
</tbody>
</table>

6.4 Natural Sciences

6.4.1 Role operators used

Twenty Chinese strings were tested in this category. According to the number of the strings, operators '1' (80%) and '2' (75%) were the most used operators. The close-up differencing code '$3' (50%) and operator 'p' (45%) came afterwards. The other operators were used by less than nine strings each (see Table 6.7).
Table 6.7 Distribution of 'Natural Sciences' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>123</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.4.2 Problems encountered

Four types of problems were encountered in this category, i.e. operators 'f', 'g', 't' - 'A' related to 'B', and downward-reading connective code ($v). All of these problems have been discussed above.

6.5 Applied Sciences

6.5.1 Role operators used

In this category, 173 Chinese strings were tested. According to the numbers of strings, the most frequently used operators were operators '2' (87.3%) and '1' (83.8%). Operator 'p' (34.7%).
close-up differencing code '$3' (42.2%), and downward-reading connective code '$v$' (32.4%) came second. About one-fifth of the strings used operators 'g' (24.3%) and '3' (23.1%). The other operators were used by less numbers of strings. A complete picture is shown in Table 6.8.

6.5.2 Problems encountered

The 'Applied Sciences' category had the greatest number of strings to test. Consequently, many problems were encountered, i.e. those related to operators 's' (10.9%), 't' (7.5%), agents of transitive actions (1.7%), and following differences (postpositions) (3.5%). Some problems have been discussed earlier, such as problems resulted from the use of operators 's' - 'applications' (as Problem 3), 't' - 'A' related to 'B' (as Problem 4) and 's' - 'influence' (as Problem 8). The problem of agents of transitive actions will be discussed in the next subject category, i.e. 'Social Sciences'. The other problems are described in Problem 9 to 14.

Problem 9: Operator 's' - 'effects'

In English, the role defining term 'effects' is used as follows: 's)effects$vo$won'. As mentioned in previous sections, the two English prepositions, 'of' and 'on' were translated into a Chinese particle '的' and a coverb '在', respectively. Problems related to operator 's' - 'effects' are illustrated in Example 6.5.1.
Table 6.8 Distribution of 'Applied Sciences' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 173</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>26</td>
<td>15.0</td>
</tr>
<tr>
<td>1</td>
<td>145</td>
<td>83.8</td>
</tr>
<tr>
<td>2</td>
<td>151</td>
<td>87.3</td>
</tr>
<tr>
<td>3</td>
<td>37</td>
<td>21.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- agents of transitive actions</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>5</td>
<td>13</td>
<td>7.5</td>
</tr>
<tr>
<td>f</td>
<td>4</td>
<td>2.3</td>
</tr>
<tr>
<td>g</td>
<td>42</td>
<td>24.3</td>
</tr>
<tr>
<td>p</td>
<td>60</td>
<td>34.7</td>
</tr>
<tr>
<td>q</td>
<td>13</td>
<td>7.5</td>
</tr>
<tr>
<td>s - influence</td>
<td>7</td>
<td>4.0</td>
</tr>
<tr>
<td>- applications</td>
<td>5</td>
<td>2.9</td>
</tr>
<tr>
<td>- effects</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>- use</td>
<td>2</td>
<td>1.1</td>
</tr>
<tr>
<td>- attitudes</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>- role</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>- policies</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>12</td>
<td>6.9</td>
</tr>
<tr>
<td>t - 'A' compared with 'B'</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>x$y$z</td>
<td>10</td>
<td>5.8</td>
</tr>
<tr>
<td>$1</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>$3</td>
<td>73</td>
<td>42.2</td>
</tr>
<tr>
<td>$d</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>$v</td>
<td>56</td>
<td>32.4</td>
</tr>
<tr>
<td>$w</td>
<td>18</td>
<td>10.4</td>
</tr>
<tr>
<td>following differences (postpositions)</td>
<td>6</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Example 6.5.1

Subject: 精神科護理在職訓練的效果

(Effects of in-service training on nursing of patients with mental disorders)

Chinese string:

*1)精神病症
*2)護理
  2)精神病症護理 ND2
  s)效果svw的$w在
*3)訓練$31在職

Entries:

精神病症
  護理. 效果 的 在職訓練 (Problem 9)

護理. 精神病症
  效果 的 在職訓練 (Problem 9)

訓練
  在職訓練. 效果 在 精神病患護理 (Problem 9)

在職訓練
  效果 在 精神病患護理 (Problem 9)

English word order from Chinese original string:

*1)patients with mental disorders
*2)nursing
  2)patients with mental disorders nursing ND2
  s)effects$vof$won
*3)training$31in-service

English word order from Chinese original entries:

Patients with mental disorders
  Nursing. Effects' in-service training (Problem 9)

Nursing. Patients with mental disorders
  Effects' in-service training (Problem 9)

Training
  In-service training. Effects on patients with mental disorders nursing (Problem 9)

In-service training
  Effects on patients with mental disorders nursing (Problem 9)

Two phrases '效果 的 在職訓練' (Effects' in-service training) and '效果 在 精神病患護理' (Effects on patients with mental disorders nursing) are unacceptable according to Chinese
syntax, since the patterns of operator 's' - 'effects' in Chinese are:

(i) - - 的效果 ( - - 's effects)
(ii) 對 - - 的效果 ( to - - 's effects)

Based on these patterns, the above two phrases should be changed to 在職訓練的效果 (in-service training's effects) and 對精神病患護理的效果 (to patients with mental disorders nursing's effects), respectively.

An upward-reading substitute was decided to be used in the above Chinese example, as 精神病患護理 (patients with mental disorders nursing). The English equivalent phrase of Chinese substitute is not correct idiomatic usage, which was caused by the different phrase formation between English and Chinese. In English, both substitutes and connectives have the same general purpose: "they link together closely bound concepts that should be presented as a set when some other term is in the lead" (Austin, 1984:38). On the other hand, connectives are more economical than substitutes because "the required phrases consist of terms selected from the string and linked by prepositions without changing their order or spellings" (Austin, 1984:39). Therefore, connectives are more preferred by indexers than substitutes.

As a result, connectives are commonly used to form an English phrase. For instance, two English terms in the above example, i.e. 'patients with mental disorders' and 'nursing', can be linked using a preposition 'of' following the upward-reading connective. The formation of an English phrase will be: 'nursing of patients with mental disorders'.
By contrast, the two Chinese terms, 護理 (nursing) and 精神病患 (patients with mental disorders), cannot be linked using a preposition. The formation of a Chinese phrase using these two terms should be 精神病患的護理 (patients with mental disorders' nursing). The action term 護理 (nursing) is placed at the end of the phrase, and a particle 的 (') is used to link this two terms. Consequently, the English algorithm of the upward-reading connective cannot be used for this. The upward-reading substitute is used in the above example instead of the upward-reading connective. In the experiment, both the downward-reading and upward-reading substitutes were represented by 'NU number' and 'ND number', respectively.

Problem 10: Operator 's' - 'use'

The role defining term 'use' is used in English as follows: 's)use$v$of$win'. The English preposition 'in' is translated into a Chinese coverb '在'. The role defining term 'use' can be translated into two Chinese terms, either 使用 or 利用. The term 使用 is chosen in the following example.

Example 6.5.2

Subject: 使用Nifedipine, Propranolol和Hydralazine治療高血壓
(Use of Nifedipine, Propranolol and Hydralazine in treatment of hypertension)

Chinese string:

*1)人
*2)高血壓
*2)治療
2)治療人的高血壓 ND3
s)使用$v$的$w$在
*3)Nifedipine
*g)Propranolol$v$&
*g)Hydralazine
Entries:

人
高血压，使用于的Nifedipine，Propranolol & Hydralazine
(Problem 10)

高血压，人
治疗，使用于的Nifedipine，Propranolol & Hydralazine
(Problem 10)

治疗，高血压，人
使用于的Nifedipine，Propranolol & Hydralazine (Problem 10)

Nifedipine
使用在治疗人的高血压 (Problem 10)

Propranolol
使用在治疗人的高血压 (Problem 10)

Hydralazine
使用在治疗人的高血压 (Problem 10)

English word order from Chinese original string:

*1) man
*2) hypertension
*2) treatment
  2) treatment human's hypertension ND3
 s) use $v of$ win
*3) Nifedipine
*g) Propranolol $v &$
*g) Hydralazine

English word order from Chinese original entries:

Man
Hypertension. Treatment. Use's Nifedipine, Propranolol & Hydralazine (Problem 10).

Hypertension. Man
Treatment. Use's Nifedipine, Propranolol & Hydralazine (Problem 10)

Treatment. Hypertension. Man
Use's Nifedipine, Propranolol & Hydralazine (Problem 10)

Nifedipine
Use in treatment human's hypertension (Problem 10)

Propranolol
Use in treatment human's hypertension (Problem 10)

Hydralazine
Use in treatment human's hypertension (Problem 10)

The English algorithm of the pattern 'use-of-in' was used in the above strings and produced two phrases: (1) '使用的Nifedipine, Propranolol & Hydralazine' (Use's Nifedipine,
Propranolol & Hydralazine) and (2) '使用在治癒人的高血壓' (Use in treatment human's hypertension). However, the use of the term 使用 (use) in Chinese should follow these patterns:

使用 -- (use -- )

使用在 -- (use in -- )

Consequently, the phrase (1) should be put right by avoiding the use of the particle 的 ('). The phrase (2) is acceptable as long as the two unexpected spaces before and after the coverb 在 (in) are removed.

As in the Example 6.5.1, an upward-reading substitute (ND3) was decided to be used in the above Chinese example, i.e. 治癒人的高血壓 (treatment human's hypertension), since the Chinese terms, 治療, 人, and 高血壓 cannot be linked using the prepositions.

Problem 11: Operator 's' - 'attitudes'

In English, the role defining term 'attitudes' is used as follows: 's)attitudes$vof$wto'. The term 'attitudes' was translated into a Chinese term 態度, and the preposition, 'to' was translated into a Chinese coverb 對. One example involving this operator is shown below:

Example 6.5.3

Subject: 護士對精神病患的態度

(Attitudes of nurses to patients with mental disorders)
In the above example, two Chinese phrases, '態度 的 護士' (Attitudes' nurses) and '態度 對 精神病患' (Attitudes to patients with mental disorders) are ungrammatical according to Chinese syntax, since the patterns of operator 's' - 'attitudes' in Chinese should be:

(i) - - 的態度 ( - - 's attitudes)
(ii) 對 - - 的態度 (to - - 's attitudes)

Thus, the two Chinese phrases should be changed to 護士的態度 (nurses' attitudes) and 對精神病患的態度 (to patients with mental disorders' attitudes), respectively.

Problem 12: Operator 's' - 'role'

The role defining term 'role' is used in English as follows: 's)role$vo$swin'. The role defining term 'role' was translated
into a Chinese term 角色. Problems related to this operator is illustrated below:

Example 6.5.4

Subject: 護士在輸血治療中扮演的角色

(Role of nurses in blood transfusion)

Chinese string:

*1)人
*p)血
*2)輸送
  2)人的輸血 ND3
 s)角色$v的$w在
*3)護士

Entries:

人 血. 輸送. 角色 的 護士 (Problem 12)
血. 人
 輸送. 角色 的 護士 (Problem 12)
輸送. 血. 人
 角色 的 護士 (Problem 12)

護士 規則 在 人的輸血 (Problem 12)

English word order from Chinese original string:

*1)man
*p)blood
*2)transfusion
  2)human's blood transfusion ND3
 s)role$vof$win
*3)nurses

English word order from Chinese original entries:

Man
 Blood. Transfusion. Role's nurses (Problem 12)

Blood. Man
 Transfusion. Role's nurses (Problem 12)

Transfusion. Blood. Man
 Role's nurses (Problem 12)

Nurses
 Role in human's blood transfusion (Problem 12)
In the above example, phrases '角色的護士' (Role's nurses) and '角色在人的輸血' (Role in human's blood transfusion) were contrary to the rules of the related Chinese syntax. The patterns of 'role-of-in' in Chinese are:

(i) - - 的角色   ( - - 's role)  
(ii) 在 - - 的角色 (in - - 's role)

According to this pattern, the above two phrases should be changed to 護士的角色 (nurses' role) and 在人的輸血的角色 (in man's blood transfusion's role) respectively.

Problem 13: Operator 's' - 'policies'

In English, the role defining term 'policies' is used as follows: 's)policies$vofof$won'. The term 'policies' was translated into a Chinese term 政策. Problems related to this operator are illustrated below:

Example 6.5.5

Subject: 臺灣政府對稻米價格的政策

(Taiwanese government policies on rice costs)

Chinese string:

*0) 臺灣  
*1) 稻米  
*p) 價格  
1) 稻米價格 ND2  
s) 政策$的$往  
*3) 政府
Entries:

台湾．rice．政策的．政府 (Problem 13)

rice．台湾
价格．政策的．政府 (Problem 13)

价格．rice．台湾
政策的．政府 (Problem 13)

政府．台湾
政策在．rice价格 (Problem 13)

English word order from Chinese original string:

*0) Taiwan
*1) rice
*p) costs
1) rice costs ND2
s) policies$wof$won
*3) government

English word order from Chinese original entries:

Taiwan
Rice. Costs. Policies' government (Problem 13)

Rice. Taiwan
Costs. Policies' government (Problem 13)

Costs. Rice. Taiwan
Policies' government (Problem 13)

Government. Taiwan
Policies on rice costs (Problem 13)

Two Chinese entries shown in the above example are ungrammatical, i.e. 政策的政府 (Policies' government) and 政策在rice价格 (Policies on rice costs). According to Chinese syntax, there are two patterns of using the role defining term 政策 (policies), i.e.

(i) --的政策 (--'s policies)
(ii) 對--的政策 (to--'s policies)

As a result, the above two ungrammatical phrases should be changed to: 政府的政策 (government's policies) and 對rice costs' policies), respectively.
Problem 14: Following differences (postpositions)

In English the term 'after' is a preposition, while in Chinese it is considered as a postposition. A postposition should follow a related action term. The formation of the action term and postposition can act as lead as well as non-lead.

During the experiment, six compound terms were found using the same postposition 后 (after). They were 檢查後 (examination after), 移植後 (transplant after), 手術後 (operation after), 出院後 (hospital after), 收穫後 (harvest after), and 地震後 (earthquake after). In order to show the problem related to the Chinese postpositions, an example using one of these compound terms, i.e. 地震後 (earthquake after) and the preceding difference '$3', is presented below:

Example 6.5.6

Subject: 日本地震後建築物重建

(Reconstruction of buildings after earthquakes in Japan)

Chinese string

*0)日本
*1)建築物
*2)地震後
*2)重建

Entries

日本 建築物. 後地震. 重建 (Problem 14)

建築物. 日本 後地震. 重建 (Problem 14)

地震. 建築物. 日本 後地震. 重建 (Problem 14)

後地震. 建築物. 日本 (Problem 14)

重建. 後地震. 建築物. 日本 (Problem 14)
The above example shows that the wrong compound term 同地震 (after earthquake) is produced instead of the right one 地震後 (earthquakes after). To overcome this, two new role operators, i.e. for lead and non-lead respectively, are needed for the Chinese postpositions.

To conclude, six problems were encountered in this category, i.e. those that were caused by operator 's': 'effects', 'use', 'attitudes', 'policies', as well as 'role'; and the lack of the role operator for the Chinese postposition 後 (after) in the English version program (see Table 6.9).
Table 6.9 New types of problems encountered in 'Applied Sciences'

<table>
<thead>
<tr>
<th>No of problems</th>
<th>Causes of problems</th>
<th>Problem descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 s - effects</td>
<td>Chinese entries produced by the patterns are unacceptable, since the Chinese patterns of 's' - 'effects' are different from the English ones</td>
<td></td>
</tr>
</tbody>
</table>
| 10 s - use     | 1. Chinese phrases produced are ungrammatical as the Chinese patterns of operator 's' - 'use' are different from the English ones  
                 2. two unnecessary spaces were produced before and after (in) when the upward-reading connective ($w$) was used |
| 11 s - attitudes| the Chinese entries produced by this operator are ungrammatical, because the Chinese patterns of 's' - 'attitudes' are different from the English ones |
| 12 s - role    | the Chinese entries caused by this operator are unacceptable, since the Chinese patterns of 's' - 'role' are different from the English ones |
| 13 s - policies| the Chinese entries caused by this pattern are unacceptable, since the Chinese patterns of 's' - 'policies' are different from the English ones |
| 14 following differences (postpositions) | according to Chinese syntax, the term 後 (after) should follow the related term. This cannot be done yet due to the lack of the relevant role operators |

6.6 Social Sciences

6.6.1 Role operators used

The number of example titles tested in the category 'Social Sciences' was 157. The most used operator was operator '2'
(83.4%), rather than '1' (68.8%). A complete picture is shown below:

Table 6.10 Distribution of 'Social Sciences' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>43</td>
<td>27.4</td>
</tr>
<tr>
<td>1</td>
<td>108</td>
<td>68.8</td>
</tr>
<tr>
<td>2</td>
<td>131</td>
<td>83.4</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>- agents of transitive actions</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0.6</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>3.2</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>f</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>g</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>p</td>
<td>20.4</td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>s - use</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>- applications</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>- effects</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>- influence</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>t - 'A' related to 'B'</td>
<td>6.4</td>
</tr>
<tr>
<td></td>
<td>- 'A' compared with 'B'</td>
<td>2.5</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>$x$y$z$</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>$3$</td>
<td>27.4</td>
</tr>
<tr>
<td></td>
<td>$d$</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>$v$</td>
<td>32.5</td>
</tr>
<tr>
<td></td>
<td>$w$</td>
<td>5.1</td>
</tr>
</tbody>
</table>
6.6.2 Problems encountered

This category is the second largest sample group, i.e. after 'Applied Sciences'. Consequently, problems encountered in this category are almost as various as in the 'Applied Sciences'. They are, amongst other things, related to operators 't', 's', 'u', and agents of transitive actions. Most of the problems have been described before, except two problems that are caused by agents of transitive actions (Problem 15), and operator 'u' (Problem 16).

Problem 15: Agents of transitive actions

As mentioned in section 6.1, operator '3' is used to indicate five types of performers or agents. A problem related to the type (3), i.e. agents of transitive actions, is shown below:

Example 6.6.1

Subject: 南非被日本经济制裁

(Economic sanctions of South Africa by Japan)

Chinese string:
*1)南非
*2)经济制裁$v被$w的
*3)日本

Entries:
南非
 經济制裁 被 日本 (Problem 15)

經濟制裁，南非
 被 日本

日本
 經济制裁 的 南非 (Problem 15)
English word order from Chinese original string:

*1) South Africa
*2) economic sanctions $v$ by $w$ of
*3) Japan

English word order from Chinese original entries:

South Africa
   Economic sanctions by Japan (Problem 15)

Economic sanctions. South Africa
   By Japan

Japan
   Economic sanctions' South Africa (Problem 15)

The above two Chinese entries, i.e. (1) 经济制裁 被日本 (Economic sanctions by Japan) and (2) 经济制裁 的 南非 (Economic sanctions' South Africa), are unacceptable according to Chinese syntax. The problem of entry (1) is caused by the use of Chinese coverb 被 (by) in the downward-reading connective ($v$), which is related to the Chinese passive construction. The problem of entry (2) resulted from the use of the particle 的 (') in the upward-reading connective. Instead, they should be written like these 被日本经济制裁 (by Japan economic sanctions) and 经济制裁南非 (economic sanctions South Africa), respectively.

Problem 16 Operator 'u'

Two strings were found that included operator 'u'. Both of them used two-way interaction. In the English pattern of operator 'u', the English preposition 'with' is used following the two connectives ($v$, $w$). This preposition 'with' was translated into a Chinese coverb 和.
Example 6.6.2

Subject: 美國與發展中國家經濟能關係
(Economic relations between United States and developing countries)

Chinese string:

*1)美國
*u)經濟關係$v和$w和
*1)發展中國家

Entries:

美國
經濟關係 和 發展中國家 (Problem 16)

經濟關係. 美國
和 發展中國家

經濟關係. 發展中國家
和 美國

發展中國家
經濟關係 和 美國 (Problem 16)

English word order from Chinese original string:

*1)United States
*u)economic relations$with$with
*1)developing countries

English word order from Chinese original entries:

United States
Economic relations with developing countries (Problem 16)

Economic relations. United States
With developing countries

Economic relations. Developing countries
With United States

Developing countries
Economic relations with United States (Problem 16)

In the above example, two phrases, i.e. '經濟關係 和 發展中國家' (Economic relations with developing countries) and '經濟關係 和 美國' (Economic relations with United States) are unacceptable
according to Chinese syntax. The patterns of two-way interaction in Chinese are:

(i) \[ A \text{ 和B - - 関係} \quad \text{(with B - - relations)} \]

(ii) \[ B \text{ 和A - - 関係} \quad \text{(with A - - relations)} \]

Thus, according to Chinese patterns of two-way interaction, the above two phrases should be changed to：和發展中國家經濟關係 (with developing countries economic relations) and 和美國經濟關係 (with United States economic relations), respectively.

To sum up, eleven problems were encountered in the category 'Social Sciences'. All of them have been discussed before, except two that are related to agents of transitive actions and operator 'u'.

Table 6.11 New types of problems encountered in 'Social Sciences'

<table>
<thead>
<tr>
<th>No of problems</th>
<th>Causes of problems</th>
<th>Problem descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>15 agents of transitive actions</td>
<td>the entries produced are unacceptable according to the Chinese passive construction</td>
<td></td>
</tr>
<tr>
<td>16 u</td>
<td>the Chinese entries caused by the pattern of operator 'u' are ungrammatical according to the patterns of two-way interaction in Chinese</td>
<td></td>
</tr>
</tbody>
</table>

6.7 Chinese History & Geography

6.7.1 Role operators used

Twelve Chinese sample titles were tested in this category. Similarly to the previous and most other subject categories,
the most frequently used operator was operator '1' (91.7%). About fifty per cent of the strings used operator '2' (50%) and downward-reading connective code '$v$' (41.7%). A complete picture is shown in Table 6.12.

Table 6.12 Distribution of 'Chinese History & Geography' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings ( N = 12 )</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>16.7</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>91.7</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>50.0</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>16.7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>16.7</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>g</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>p</td>
<td>3</td>
<td>25.0</td>
</tr>
<tr>
<td>u</td>
<td>3</td>
<td>25.0</td>
</tr>
<tr>
<td>$x$ $y$ $z$</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>8.3</td>
</tr>
<tr>
<td>$v$</td>
<td>5</td>
<td>41.7</td>
</tr>
<tr>
<td>$w$</td>
<td>3</td>
<td>25.0</td>
</tr>
</tbody>
</table>

6.7.2 Problems encountered

Five types of problems, i.e. those that were caused by operators 'f', 'g', 'u' and connective codes ($v$, $w$), were found in this category and have been discussed before.
6.8 World History, Geography & Biography

6.8.1 Role operators used

In this category, 33 strings were tested. According to the number of the strings, the most frequently used operator was operator '1' (94%). Operator '6' (57.6%) came second, since this operator was used for 'Biography'. A complete picture is depicted in Table 6.13.

Table 6.13 Distribution of 'World History, Geography & Biography' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N = 33 )</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>12.1</td>
</tr>
<tr>
<td>1</td>
<td>31</td>
<td>94.0</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>39.4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td>- agents of transitive actions</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>19</td>
<td>57.6</td>
</tr>
<tr>
<td>p</td>
<td>8</td>
<td>24.2</td>
</tr>
<tr>
<td>q</td>
<td>6</td>
<td>18.2</td>
</tr>
<tr>
<td>r</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>s - participation</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>- policies</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>( $x$y$z )</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>$3 )</td>
<td>4</td>
<td>12.1</td>
</tr>
<tr>
<td>$d )</td>
<td>3</td>
<td>12.1</td>
</tr>
<tr>
<td>$v )</td>
<td>2</td>
<td>9.1</td>
</tr>
<tr>
<td>$w )</td>
<td>1</td>
<td>3.0</td>
</tr>
</tbody>
</table>
6.8.2 Problems encountered

One new problem was encountered in this category which was related to the role defining term 'participation', as seen in the example below:

Problem 17 Operator 's' - 'participation'

Example 6.8.1

Subject: 教士參與義大利的抵抗運動
(Participation of clergy in Italian resistances movements, 1940-1945)

Chinese string:

*0)義大利
*1)抵抗運動$d1940-1945
s)參與$s的$w在
*3)教士

Entries:

義大利
抵抗運動. 1940-1945. 參與 的 教士 (Problem 17)

抵禦運動. 義大利
1940-1945. 參與 的 教士 (Problem 17)

教士. 義大利
參與 在 抵抗運動. 1940-1945 (Problem 17)

English string (Austin, 1984:149):

*0)Italy
*1)resistance movements$d1940-1945
s)participation$vo$win
*3)clergy

English word order from Chinese original entries:

Italy
Resistance movements. 1940-1945. Participation's clergy (Problem 17)

Resistence movements. Italy
1940-1945. Participation's clergy (Problem 17)

Clergy. Italy
Participation in resistance movements, 1940-1945 (Problem 17)
Two entries of the above example are unacceptable according to Chinese syntax, i.e. (1) '參與的教士' (Participation's clergy) and (2) '參與在抵抗運動' (Participation in resistance movements). These are caused by the patterns of 參與 (participation) in Chinese are:

(i) 參與 (participate)

(ii) 參與 (participate)

According to these patterns, the entry (1) should be changed to 敎士的參與 (clergy's participate), and the entry (2) to 參與抵抗運動 (participate resistance movements).

Table 6.14 New type of problem encountered in 'World History, Geography & Biography'

<table>
<thead>
<tr>
<th>No of problem</th>
<th>Cause of problem</th>
<th>Problem description</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>s - participation</td>
<td>the Chinese entries produced by this operator are ungrammatical, because the Chinese patterns of 's' - 'participation' are different from the English ones</td>
</tr>
</tbody>
</table>

6.9 Language & Literature

6.9.1 Role operators used

Twenty-one strings were tested in this category. According to the number of strings, the most heavily used operator was '1' (85.7%). Operator 'p' (76.2%) came second, since there were 16 strings that were related to 'Chinese Literature', such as poems, essays and fiction. A complete picture is shown in Table 6.15.
Table 6.15 Distribution of 'Language & Literature' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>85.7</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>61.9</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>9.5</td>
</tr>
<tr>
<td>- agents of transitive actions</td>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>g</td>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>p</td>
<td>16</td>
<td>76.2</td>
</tr>
<tr>
<td>q</td>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>s - role</td>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>$x$y$z$</td>
<td>3</td>
<td>14.3</td>
</tr>
<tr>
<td>$3$</td>
<td>2</td>
<td>9.5</td>
</tr>
<tr>
<td>$d$</td>
<td>1</td>
<td>4.8</td>
</tr>
<tr>
<td>$v$</td>
<td>6</td>
<td>28.6</td>
</tr>
<tr>
<td>$w$</td>
<td>1</td>
<td>4.8</td>
</tr>
</tbody>
</table>

6.9.2 Problems encountered

Five types of problems were found in this category, i.e. those that were caused by '$v', operators 'f', 'g', 's' - 'role', and agents of transitive actions, respectively. These problems have been discussed before.
6.10 The Arts

6.10.1 Role operators used

In this category, 20 Chinese strings were tested. Two operators, '1' and '2', were used by more than 64% strings each. Operator 'p' was used by over half of the strings, since these strings were about Chinese art works which were parts of 'Chinese Arts'.

Table 6.16 Distribution of 'The Arts' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>70</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>65</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>g</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>p</td>
<td>11</td>
<td>55</td>
</tr>
<tr>
<td>q</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>s - applications</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>$3</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>$d</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>$v</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>$w</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

6.10.2 Problems encountered

In this category, four types of problems were encountered. They resulted from the use of '$v$', operators 'g', 's' -
'applications', and operator 't' - 'A' related to 'B', respectively. All of these problems have been discussed before.

6.11 Research on China & the Soviet Union

6.11.1 Role operators used

Fifteen strings were tested in this category. Operators '1' and '2' were used by more than 85% strings. Operator '0' was used by over 45% of the strings, since operator '0' is used for 'location' which, in this category, referred to China and the Soviet Union.

Table 6.17 Distribution of 'Research on China & the Soviet Union' strings by role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Number of strings N = 15</th>
<th>Per cent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7</td>
<td>46.7</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>93.3</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>86.7</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>6.7</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>13.3</td>
</tr>
<tr>
<td>g</td>
<td>3</td>
<td>20.0</td>
</tr>
<tr>
<td>p</td>
<td>6</td>
<td>40.0</td>
</tr>
<tr>
<td>t - 'A' related to 'B'</td>
<td>2</td>
<td>13.3</td>
</tr>
<tr>
<td>$3</td>
<td>2</td>
<td>13.3</td>
</tr>
<tr>
<td>$d</td>
<td>3</td>
<td>20.0</td>
</tr>
<tr>
<td>$v</td>
<td>3</td>
<td>20.0</td>
</tr>
</tbody>
</table>

6.11.2 Problems encountered

There was no new type of problem encountered in this subject category.
Chapter Summary

The problems encountered by the eleven categories have been described above. They are grouped according to the types of role operators that caused the problems, and subject categories of the strings.

6.12.1 Role operators used in eleven subject categories

A total number of 498 Chinese strings were tested in the experiment. According to the numbers of strings, operators '1' and '2' were the most heavily used operators across the subject categories, i.e. 79.1% and 77.5%, respectively (see Table 6.18). Similar situation could be found also in each subject category, i.e. most of the strings of each subject category used either role operator '1' (e.g. 'Generalities' and 'Research on China & the Soviet Union') or '2' (e.g. 'Applied Sciences' and 'Research on China & the Soviet Union').

In addition to the following differences (postpositions) (0.1%), the least used codes were $n$ and $o$, i.e. 0.2% each, as they were used by only one subject category, 'Philosophy'. However, '$n$' and '$o$' are not the codes which are designed especially for this one category. It might be used also in other subject categories.

Operators '1', '2', '3', 'p' and downward-reading connective ($v$) were used by all subject categories. Operator 'g' was used by all categories except 'World History, Geography & Biography'. However, operator 'g' might be used also in this category.
There were more strings that made use of operator '6' than those that used the other extra-core concepts operators (i.e. operator '4' and '5'). Amongst the sample that made use of operators of special classes of action (i.e. 's', 't', and 'u'), those that used operator 't' were the biggest group, i.e. 7.4% compared with 6.4% ('s') and 1% ('u').

Table 6.18 also shows that some role operators were used by a large percentage of strings in particular subject categories. For example, operator '0' was used heavily in category 'Research on China & the Soviet Union', as two countries were involved in this category. The category 'World History, Geography & Biography' used operator '6' in 57.6% of its strings, that is, especially for 'Biography'. Operator 'u' was found only in two categories, i.e. 'Social Sciences' and 'Chinese History & Geography', since some relationships between two countries appeared in these two categories.
Table 6.18 Percentage of strings according to subject categories and role operators

<table>
<thead>
<tr>
<th>Role \ Subject \ categories</th>
<th>*Gen \ Phil \ Rel</th>
<th>Nat \ Sci</th>
<th>Appl \ Sci</th>
<th>Soc \ Sci</th>
<th>Chi \ His</th>
<th>World \ His</th>
<th>Lang \ Arts</th>
<th>Chi \ &amp; Sov</th>
<th>[All subject] \ categories</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(100%) (100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
<td>(100%)</td>
</tr>
<tr>
<td>0</td>
<td>29.1</td>
<td>40</td>
<td>15</td>
<td>15.0</td>
<td>27.4</td>
<td>16.7</td>
<td>12.1</td>
<td>14.3</td>
<td>15.0</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>46.2</td>
<td>70</td>
<td>60</td>
<td>63.8</td>
<td>68.8</td>
<td>91.7</td>
<td>94.0</td>
<td>85.7</td>
</tr>
<tr>
<td>2</td>
<td>50.0</td>
<td>84.6</td>
<td>40</td>
<td>75</td>
<td>57.3</td>
<td>83.4</td>
<td>50</td>
<td>39.4</td>
<td>61.9</td>
</tr>
<tr>
<td>3</td>
<td>21.0</td>
<td>7.7</td>
<td>20</td>
<td>20</td>
<td>23.1</td>
<td>21.7</td>
<td>16.7</td>
<td>12.1</td>
<td>14.3</td>
</tr>
<tr>
<td>4</td>
<td>4.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6</td>
<td>0.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>16.7</td>
<td>15.4</td>
<td>-</td>
<td>10</td>
<td>7.5</td>
<td>6.4</td>
<td>16.7</td>
<td>57.6</td>
<td>-</td>
</tr>
<tr>
<td>f</td>
<td>-</td>
<td>15.4</td>
<td>10</td>
<td>5</td>
<td>2.3</td>
<td>2.5</td>
<td>8.3</td>
<td>-</td>
<td>4.8</td>
</tr>
<tr>
<td>g</td>
<td>25.0</td>
<td>7.7</td>
<td>10</td>
<td>5</td>
<td>24.3</td>
<td>22.3</td>
<td>8.3</td>
<td>-</td>
<td>14.3</td>
</tr>
<tr>
<td>p</td>
<td>50.0</td>
<td>84.6</td>
<td>60</td>
<td>45</td>
<td>34.7</td>
<td>20.4</td>
<td>25.0</td>
<td>24.2</td>
<td>76.2</td>
</tr>
<tr>
<td>q</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>5</td>
<td>7.5</td>
<td>12.1</td>
<td>-</td>
<td>18.2</td>
<td>14.3</td>
</tr>
<tr>
<td>r</td>
<td>-</td>
<td>-</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>s</td>
<td>4.2</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>10.9</td>
<td>4.5</td>
<td>-</td>
<td>6.0</td>
<td>4.8</td>
</tr>
<tr>
<td>t</td>
<td>16.7</td>
<td>7.7</td>
<td>20</td>
<td>5</td>
<td>7.5</td>
<td>8.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>u</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.3</td>
<td>25.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$x^y$z^c</td>
<td>4.2</td>
<td>15.4</td>
<td>10</td>
<td>-</td>
<td>5.8</td>
<td>4.5</td>
<td>8.3</td>
<td>3.0</td>
<td>14.3</td>
</tr>
<tr>
<td>$l$</td>
<td>-</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>0.6</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$s$</td>
<td>12.5</td>
<td>61.5</td>
<td>20</td>
<td>50</td>
<td>42.2</td>
<td>27.4</td>
<td>-</td>
<td>12.1</td>
<td>9.5</td>
</tr>
<tr>
<td>$d$</td>
<td>8.3</td>
<td>30.8</td>
<td>-</td>
<td>-</td>
<td>1.7</td>
<td>8.9</td>
<td>8.3</td>
<td>12.1</td>
<td>4.8</td>
</tr>
<tr>
<td>$n$</td>
<td>-</td>
<td>7.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$o$</td>
<td>-</td>
<td>7.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$v$</td>
<td>37.5</td>
<td>23.1</td>
<td>20</td>
<td>5</td>
<td>32.4</td>
<td>32.5</td>
<td>41.7</td>
<td>9.1</td>
<td>28.6</td>
</tr>
<tr>
<td>$w$</td>
<td>8.3</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>10.4</td>
<td>5.1</td>
<td>25</td>
<td>3.0</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Following differences

*Gen: Generalities  
Phil: Philosophy  
Rel: Religion  
Nat Sci: Pure Sciences  
App: Applied Sciences  
Soc Sci: Social Sciences  
Chi His: Chinese History & Geography  
World His: World History, Geography  
Chi & Sov: Research of China & the Soviet Union  
Lang & Lit: Language & Literature
6.12.2 Problems encountered in eleven subject categories

Some types of problems occurred more than once across the subject categories, but discussed only once. Table 6.19 gives a complete picture of the problems encountered by the subject categories.

Table 6.19 shows that the number of subject categories affected by problems ranged from eleven (Problem 1) to one (Problem 6, 11, 14 and 17). However, in reality, all subject categories are prone to all types of problems.

The fundamental problems encountered in these subject categories were brought about by the differences between Chinese and English syntactic rules, including passive construction, patterns of comparing and relating two things, special patterns used in operator 's', and the formation of a phrase using a preposition and postposition. The minor problems were related to spacing and the ampersand '&'.
<table>
<thead>
<tr>
<th>No of problem problems</th>
<th>Causes of</th>
<th>Problem description</th>
<th>Subject categories affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$v$</td>
<td>unexpected spaces in the produced Chinese entries separated a phrase into three pieces, and thus destroyed its meaning</td>
<td>11 categories the percentages of strings of each category that were affected by the problems are ranging from 5% to 41.7%</td>
</tr>
<tr>
<td>2</td>
<td>$g$</td>
<td>the ampersand (&amp;) is never used in Chinese text, and thus the use of this sign may destroy the meaning of the related Chinese entry</td>
<td>10 categories except 'World History, Geography &amp; Biography' the percentages of strings each category that used this operator ranging from 5% to 26.1%</td>
</tr>
<tr>
<td>3</td>
<td>$s$ - applications and $$m$</td>
<td>1. Chinese phrases produced are ungrammatical as the Chinese patterns of operator 's' - &quot;application&quot; are different from the English ones 2. two unnecessary spaces were produced when the upward-reading connective was used in Chinese strings, i.e. before and after the term (\in)</td>
<td>4 categories Generalities (4.3%) Applied Sciences (2.9%) Social Sciences (1.3%) The Arts (5%)</td>
</tr>
<tr>
<td>4</td>
<td>'t' related to 'b'</td>
<td>1. the use of punctuation %' for italicizing the related terms had no effect in Chinese, as Chinese script did not have italic form, instead the %' appeared in the produced entries 2. the phrases produced are grammatically wrong due to the differences between Chinese and English patterns of relating two things</td>
<td>8 categories Generalities (8.7%) Philosophy (7.7%) Religion (20%) Natural Sciences (5%) Applied Sciences (6.9%) Social Sciences (6.4%) The Arts (5%) Research of China &amp; the Soviet Union (13.3%)</td>
</tr>
<tr>
<td>5</td>
<td>'t' compared with 'b'</td>
<td>1. the use of punctuation %' for italicizing the related terms has no effect in Chinese, as Chinese script does not have italic form, instead the %' appeared in the produced entries 2. the phrases produced are grammatically wrong due to the differences between Chinese and English patterns of comparing two things</td>
<td>3 categories Generalities (4.3%) Applied Sciences (0.6%) Social Sciences (2%)</td>
</tr>
</tbody>
</table>

(continued)
(Table 6.19 continued)

| 6 | t - 'A' expounded by 'e' | 1. the use of punctuation 't' for italicizing the related terms has no effect in Chinese, as Chinese script does not have italic form, instead the 't' appeared in the produced entries |
| 7 | f | 2. the entry produced is unacceptable according to the Chinese passive construction |
| 8 | s - influence | 3. the use of the upward-reading connective ($\w$) code has led to unnecessary spaces |
| 9 | s - effects | the ampersand 'e' is never used in Chinese text, and thus the use of this sign may destroy the meaning of the related Chinese entry |
| 10 | s - use | the Chinese entries produced are unacceptable according to the produced entries |
| 11 | s - attitudes | the Chinese patterns of operator 's' - 'effects' are different from the English ones |

1 category Generalities (4.2%)

7 categories
- Philosophy (15.4%)
- Religion (10%)
- Natural Sciences (5%)
- Applied Sciences (2.3%)
- Social Sciences (2.3%)
- Chinese History & Geography (8.3%)
- Language & Literature (4.8%)

3 categories
- Religion (10%)
- Applied Sciences (4%)
- Social Sciences (0.6%)

2 categories
- Applied Sciences (1.1%)
- Social Sciences (2%)

2 categories
- Applied Sciences (1.1%)
- Social Sciences (2%)

1 category
- Applied Sciences (0.6%)

(continued)
(Table 6.19 continued)

| 12 | s - role | the Chinese entries caused by this operator are unacceptable, since the Chinese patterns of 's' - 'role' are different from the English ones | 2 categories | Applied Sciences (0.6%) | Language & Literature (4.0%) |
| 13 | s - policies | the Chinese entries caused by this patterns are unacceptable, since the Chinese patterns of 's' - 'policies' are different from the English ones | 2 categories | Applied Sciences (0.6%) | World History, Geography & Biography (3%) |
| 14 | following differences (postpositions) | according to Chinese syntax, the term (after) should follow the related term. This cannot be done yet due to the lack of the relevant role operators and codes | 1 category | Applied Sciences (3.5%) |
| 15 | agents of transitive actions | the entries produced are unacceptable according to the Chinese passive construction | 5 categories | Philosophy (7.7%) | Applied Sciences (1.1%) | Social Sciences (3.2%) | World History, Geography & Biography (1%) | Language & Literature (4.0%) |
| 16 | operator 'u' | the Chinese entries caused by the patterns of operator 'u' are ungrammatical according to the patterns of two-way interaction in Chinese | 2 categories | Social Sciences (1.3%) | Chinese History & Geography (25%) |
| 17 | s - participation | the Chinese entries produced by this operator are ungrammatical, because the Chinese patterns of 's - participation' are different from the English ones | 1 category | World History, Geography & Biography (3%) |

In addition to the above problems, typographic problems were also encountered, such as boldface, capitalization and italics. These problems will all be discussed in the next chapter.
The above problems can be classified into eight groups mostly according to the types of role operators, i.e.

1. **connective codes:**
   - - $v$ (Problem 1)
   - - $w$ (Problem 3)

2. **coordinate concepts**
   - - operator 'g' (Problem 2)
   - - operator 'f' (Problem 7)

3. **agents of transitive actions** (Problem 15)

4. **role definers - operator 's'**
   - - applications (Problem 3)
   - - influence (Problem 8)
   - - effects (Problem 9)
   - - use (Problem 10)
   - - attitudes (Problem 11)
   - - role (Problem 12)
   - - policies (Problem 13)
   - - participation (Problem 17)

5. **two-way interaction - operator 'u'** (Problem 16)

6. **author-attributed associations - operator 't'**
   - - 'A' related to 'B' (Problem 4)
   - - 'A' compared with 'B' (Problem 5)
   - - 'A' expounded by 'B' (Problem 6)

7. **Following differences** (postpositions) (Problem 14)

8. **typography**

These findings are in line with Sørenson and Austin's findings (mentioned in section 2.3) where they observed problems that may arise when PRECIS is used in natural languages other than
English, i.e. problems related to:

(a) schema of role operators: operators 's', 't', 'u', and agents of transitive actions;
(b) case marking: connective codes ($v$ and $w$); and
(c) semantic relations: coordinate concepts (operators 'f' and 'g'), and following differences.

In summary, during the experiment it was found that the role operators which caused problems (especially major problems) were used by only a small percentage of strings. Operators that were used the most did not result in any problems. This fact proved that PRECIS is a multilingual indexing system, and thus can be used in the Chinese language with some modifications to the English version programs. The solutions to those problem areas will be given in the next chapter.
Problems discovered during the experiment have been presented in Chapter 6. The aim of this chapter is to discuss those problems and to suggest solutions according to eight groups of problem areas. These are: (1) connective codes, (2) coordinate concepts, (3) agents of transitive actions, (4) role definers - operator 's', (5) two-way interaction - operator 'u' (6) author-attributed associations - operator 't', (7) following differences (postpositions), and (8) typography.

7.1 Connective Codes

As stated earlier (see section 3.4), all Chinese characters are equal and separated from each other without character boundaries. Consequently, spaces resulting from the use of connective codes ($v$ and $w$) should be removed.

The Chinese internal code was used to modify the programs. As mentioned in section 5.4, the first byte (high byte) of Chinese characters and punctuation in the Chinese internal code, BIG-5, has a value greater than 161. This significant feature was used in the modification, i.e. by adding the following decision statement to the programs in order to identify the internal codes: if the value of the first byte (high byte) is greater than 161, i.e. Chinese data, there are no spaces between terms for connective codes ($v$, $w$). In contrast, if the value of the first byte is less than 161, namely English data, spaces are needed between the connected term (see Appendix 7.1).
Thus, two entries presented in Problems 1 and 3 will then appear like this: 次序和結構 (order & structure) and '應用在 CARIM TM 電腦輔助資訊檢索和導向系統' (apply in CARIM TM computer assisted information retrieval & guiding system). Spaces have been removed in these entries. In addition, the ampersand '&' is changed to a Chinese coverb 和 (and), which is discussed in the following section.

7.2 Coordinate Concepts

Both Problems 2 and 7, which were brought about by operators 'g' and 'f' respectively, suggested that the ampersand '&' had to be changed to the Chinese equivalence. In the Chinese language the ampersand '&' is equal to coverbs, 與, 和, 或 or 同, which means 'and'. The function of these coverbs is to act as conjunctions which serve to join terms.

The coverb 與 (and) is a remnant from the ancient Chinese language. In modern Chinese writing 和 (and) is the more widely used (Wang, 1947). The coverb 或 is the most commonly used conjunction in the spoken language; and the coverb 同 is used in the Central and Southern dialect of China (Chao, 1968). For the purposes of this thesis, the coverb 和 (and) is suggested to replace the ampersand when the codes of coordinate concepts ('f', 'g') are used in Chinese strings. In addition, unnecessary spaces have to be removed before and after the coverb 和 (and) in the entry which has been discussed in section 7.1. As a result, the modified entry of Problem 2 is 次序和結構 (order & structure).
7.3 Agents of Transitive Actions

As mentioned in Chapter 6, the problem related to agents of transitive actions was caused by the structural differences which exist between the Chinese and English passive. To show the differences, the example in section 3.4 is restated here.

Example 7.3.1

The window was broken by the thief

NP2  Aux  V  by  NP1

Pat  Action  by  Agt

role  |--- (1)   (2)   (3)
operators -

* Note: 'Pat' is the abbreviation of 'Patient'. The term 'Patient' is used for the object of transitive action.

In English, the construction of syntactic order in agents of transitive actions is:

Pat - Action - by - Agt

role  |--- (1)   (2)   (3)
operators -

The algorithm to produce an English entry which is related to the agents of transitive actions is shown in Table 7.1 (Austin, 1984:352-3).
Table 7.1: The algorithm of the English downward-reading connective ($v$)

<table>
<thead>
<tr>
<th>Steps</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>assemble a phrase consisting of:</td>
<td>String:</td>
</tr>
<tr>
<td>(i) the gated term containing $v$</td>
<td>*1)windows</td>
</tr>
<tr>
<td>(ii) the $v$ connective</td>
<td>2)breaking by thieves</td>
</tr>
<tr>
<td>(iii) the next permitted term later in the string</td>
<td>*3)thieves</td>
</tr>
<tr>
<td>String:</td>
<td></td>
</tr>
<tr>
<td>*1) windows</td>
<td></td>
</tr>
<tr>
<td>2) breaking by thieves</td>
<td></td>
</tr>
<tr>
<td>*3) thieves</td>
<td></td>
</tr>
<tr>
<td>Entries:</td>
<td></td>
</tr>
<tr>
<td>Windows</td>
<td></td>
</tr>
<tr>
<td>Breaking by thieves</td>
<td></td>
</tr>
<tr>
<td>(i) (ii) (iii)</td>
<td></td>
</tr>
<tr>
<td>Thieves</td>
<td></td>
</tr>
<tr>
<td>Breaking of windows</td>
<td></td>
</tr>
</tbody>
</table>

On the other hand, the structure of the Chinese passive is as follows:

```
<table>
<thead>
<tr>
<th>NP2</th>
<th>NP1</th>
<th>VP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pat</td>
<td>被</td>
<td>Agt</td>
</tr>
</tbody>
</table>
```

role operators - |--- (1) 被 (3) (2)

window by thief break

This diagram shows that the Chinese syntactic order in agents of transitive actions is:

```
| Pat | 被  | Agt | Action |
```

role operators - |--- (1) 被(by) (3) (2)

If the above Chinese syntactic order is to be followed, the related Chinese strings and entries should be written as follows:

**Chinese string:**

*1) 窗户 (1)
2) 打破 $v$ 被 $w$ (2) (3)
*3) 小偷 (4)

**English word order from Chinese original string:**

*1) windows (1)
2) break $v$ by $w$ (2) (3)
*3) thieves (4)
As seen in the entry (ii) of the above example, there is no coverb needed between a verb 打破 (break) and a noun 窗户 (windows) in the Chinese entry. Therefore, a blank connective was used in the upward-reading connective.

The algorithm for producing the entry (i) starts with the downward-reading connective term, a coverb 被 (by), then the agent 小偷 (thieves), and ends with the action term 打破 (break). These steps of the algorithm are shown in Table 7.2.

Table 7.2: The algorithm of the Chinese downward-reading connective ($v$)

<table>
<thead>
<tr>
<th>Steps</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>assemble a phrase consisting of:</td>
<td>String:</td>
</tr>
<tr>
<td>(i) the $v$ connective</td>
<td>*1) 窗户</td>
</tr>
<tr>
<td>(ii) the next permitted term later in the string</td>
<td>2) 打破 $v$被 $w$</td>
</tr>
<tr>
<td>(iii) the gated term containing $v$</td>
<td>*3) 小偷 (ii)</td>
</tr>
<tr>
<td></td>
<td>Entry:</td>
</tr>
<tr>
<td></td>
<td>窗户 (windows)被小偷打破 by thieves break</td>
</tr>
</tbody>
</table>

The modification of the English version programs to meet the above algorithm for a Chinese downward-reading connective can be seen in Appendix 7.2. This algorithm can be used for all entries related to the Chinese passive
construction. Thus, Problem 15 is resolved using the new
algorithm as follows:

Example 7.3.2
Subject: 南非被日本経済制裁

Chinese string: English word order from Chinese
*1) 南非 order string:
*2) 経済制裁$w$被$w$
*3) 日本

Entries:

南非
被日本経済制裁
経済制裁，南非
被日本
日本
経済制裁南非

English word order from Chinese original entries:

South Africa
By Japan economic sanctions

Economic sanctions. South Africa
By Japan

Japan
Economic sanctions South Africa

7.4 Role Definers - Operator 's'

Some role operators, notably role definers - operator 's', are
used in accordance with English syntax. The role defining
function of operator 's' is regularly handled by standard terms
and connectives:

s) applications$vo$w$w$in  s) participation$vo$w$w$in
s) attitudes$vo$w$w$to  s) policies$vo$w$w$on
s) effects$vo$w$w$on  s) role$vo$w$w$in
s) influence$vo$w$w$on  s) use$vo$w$w$in
All the terms listed above are not exhaustive but the most commonly used terms in the Chinese language. To facilitate the solutions to the problems related to these terms, terms are grouped according to characteristics of the problems they brought about, i.e. (1) 'applications' and 'use', (2) 'influence', and (3) 'effects', 'attitudes', 'role', 'participation', and 'policies'.

7.4.1 'Applications' and 'use'

The patterns of the uses of 'applications' and 'use' in Chinese are similar. Therefore, the problems brought about by these two role defining terms are also similar, i.e.

(1) Chinese phrases produced are ungrammatical as the Chinese patterns of operator 's' - 'use' and 's' - 'applications' are different from the English ones.

(2) Two unnecessary spaces were produced before and after the coverb 在 (in) when the upward-reading connective ($w$) was used.

Discussing role defining terms, Austin (1984:151) stated that "English allows distinctions that cannot be made in all other languages - for example, ... between 'use' and 'applications'." This is true in the case of Chinese language. In Chinese both 'applications' and 'use' can be translated into 用, 使用, 應用 and 利用, which are basically synonymous and are regarded as instrumental verbs (Li, 1971; Light, 1979).

As seen in Examples 6.1.2 and 6.5.2, the coverb 在 (in) was used with the instrumental verbs 應用 (applications) and 使用 (use). This coverb is often used with instrumental verbs
because it is a locative marker. The position of this coverb is normally after the verbs and preceded the nouns (Li, 1971; Light, 1979).

Example 3.5 in section 3.3 (2) shows that the uses of both the verb 使用 (use) and the coverb 在 (in) in the case of 'instrument' as seen in Figure 7.1 (Teng, 1975:89).

![Diagram](https://example.com/diagram.png)

(Electronic mail is used in libraries.)

Figure 7.1: The structure of instrumental verbs and the coverb '在'

Consequently, in order to produce the right Chinese entries using the role defining term 使用 (use), the related English pattern 'use$vof$win' should be changed to '使用$v$在' (use$v$win). The following example illustrates the effect of the new pattern of operator 's' - 'use', and hence, 'applications' on the Chinese entries.
Example 7.4.1

Subject: 電子信件使用在圖書館

(Use of electronic mail in libraries)

<table>
<thead>
<tr>
<th>Chinese string</th>
<th>English word order from Chinese original string:</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1)圖書館</td>
<td>*1)libraries</td>
</tr>
<tr>
<td>s)使用電子信件</td>
<td>s)use$v$win</td>
</tr>
<tr>
<td>*3)電子信件</td>
<td>*3)electronic mail</td>
</tr>
</tbody>
</table>

Entries:

<table>
<thead>
<tr>
<th>圖書館</th>
<th>Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>使用電子信件</td>
<td>Use electronic mail</td>
</tr>
<tr>
<td>電子信件</td>
<td>Electronic mail</td>
</tr>
<tr>
<td>使用在圖書館</td>
<td>Use in libraries</td>
</tr>
</tbody>
</table>

7.4.2 'Influence'

As stated in 6.3.2, the patterns of operator 's' - 'influence' in Chinese are:

(i) 受（彼）--- 影響 (by --- influence)

(ii) 影響 --- (influence --- )

The coverb 受 or 被 which means 'by' is used in the first pattern and forms a Chinese passive phrase. Since the Chinese passive construction differs from English one, the related algorithm has to be modified. Following the Chinese passive construction mentioned in section 7.3, the problem with this role defining term should be resolved by the Chinese pattern '2)影響$v$受$w$' (2)influence$by$w), as seen in Example 7.4.2.

Example 7.4.2

Subject: 基督教内在教義對西方世界影響

(Influence of Christian spiritual doctrines on the Western world)
Chinese string:

*1)西方世界
2)影響$\ast w$
3)基督教內在教義  NU2
*3)基督教
*p)內在教義

Entries:

西方世界
受基督教內在教義影響

基督教
內在教義，影響西方世界

內在教義，基督教
影響西方世界

English word order from Chinese original string:

*1)Western world
2)influence$\ast w$
3)Christian spiritual doctrines  NU2
*3)Christianity
*p)spiritual doctrines

English word order from Chinese original entries:

Western world
By Christian spiritual doctrines influence

Christianity
Spiritual doctrines. Influence Western world

Spiritual doctrines. Christianity
Influence Western world

7.4.3 'Effects, attitudes, role, policies, and participation'

Five similar problems of operator 's found in the experiment
are related to 'effects' (Problem 9), 'attitudes' (Problem 11),
'role' (Problem 12), 'policies' (Problem 13), and
'participation' (Problem 17). These role defining terms should
be used according to the following patterns:
(1) -- the attitude (effects, policies, role, participate)

(- -'s attitudes (effects, policies, role, participate))

(2)

(2.1) to - -'s attitude (effects, policies)

(to - -'s attitudes (effects, policies))

(2.2) in - -'s role

(2.3) participate --

(participate --)

In order to show the problems related to these five patterns of operator 's', Example 6.5.3 is restated below:

Example 7.4.3

Subject: 护士對精神患者的態度

(Attitudes of nurses to patients with mental disorders)

Chinese string: 

*1)精神患者 (1) 
s)態度$的$對 
(2) (3) (4)

*3)護士 (5)

English word order from Chinese original string:

*1)patients with mental disorders (1) 
s)attitudes$of$to 
(2) (3) (4)

*3)nurses (5)

According to the patterns (1) and (2.1), the above string should produce the right entries which are as follows:

Entries:

(i) 精神病患 (1)

護士的態度

--- --- ---

(5) (3) (2)

(ii) 護士 (5)

對精神病患的態度

-- --- --- ---

(4) (1) (3) (2)
Using the existing role operators and algorithms, it is not possible to produce the right Chinese entries presented above. Basically, there are three methods to overcome this problem, i.e. by:

I. adding new role operators;
II. modifying the existing algorithms; or
III. using theme interlinks.

The first two ways are too complicated, and are not reasonable to follow only for solving a problem that does not occur very often. According to Table 6.19, these five role defining terms were used only in one to two subject categories of the sample, i.e. by 4.8% to 0.6% strings of the related categories. Consequently, the method III is chosen. The following example illustrates the application of theme interlinks.

Example 7.4.4

Chinese string:

\[ x^* 1) \text{精神} \ \text{病患} \]
\[ y \ 3) \text{護士的態度} \]
\[ x^* 1) \text{護士} \]
\[ y \ P) \text{對精神} \ \text{病患的態度} \]

Entries:

精神病患
護士的態度
護士
對精神病患的態度
7.5 Two-way Interaction - Operator 'u'

Problem 16 has shown that there are two entries that should be produced according to the following patterns:

(i) \[ A \quad \text{和} \quad B \quad \text{--}\quad \text{關係} \quad (\text{簽約，合作，協商}) \]
with \[ B \quad \text{relations} \quad (\text{contracts，cooperation，negotiations}) \]

(ii) \[ B \quad \text{和} \quad A \quad \text{--}\quad \text{關係} \quad (\text{簽約，合作，協商}) \]
with \[ A \quad \text{relations} \quad (\text{contracts，cooperation，negotiations}) \quad (\text{Austin, 1984}) \]

To illustrate the use of these patterns, Example 6.6.2 is represented as follows:

Example 7.5.1

Subject: 美國和發展中國家經濟關係

(Economic relations between United States and developing countries)

Chinese string:  

*1) 美國 (1)  
*u) 經濟關係 (2)  
*1) 發展中國家 (5)

English word order from Chinese original string:

*1) United States (1)  
*u) economic relations (2)  
*1) developing countries (5)
According to the patterns of Chinese two-way interaction, the correct entries should be written as follows:

**Entries:**

(i) 美國 (1)
    和發展中國家經濟關係
    ------
    (3)  (5)  (2)

(ii) 經濟關係．美國
     ------
     (2)  (1)
     和發展中國家
     ------
     (3)  (5)

(iii) 經濟關係．發展中國家
      ------
      (2)  (5)
      和美國
      ------
      (4)(1)

(iv) 發展中國家 (5)
    和美國經濟關係
    ------
    (4)  (1)  (2)

**English word order from Chinese original entries:**

(i) United States (1)
    With developing countries economic relations
    ------
    (3)  (5)  (2)

(ii) Economic relations. United States
     ------
     (2)  (1)
     With developing countries
     ------
     (3)  (5)

(iii) Economic relations. Developing countries
      ------
      (2)  (5)
      With United States
      ------
      (4)  (1)

(iv) Developing countries (5)
    With United States economic relations
    ------
    (4)  (1)  (2)
To produce the entry (i), the algorithm for Chinese passive construction which has been developed in section 7.3 should be used.

Since the term 經濟關係 (economic relations) acts as a lead in the above string, the consequent two entries (entries (ii) and (iii)) are all acceptable according to Chinese syntax. In other words, the original algorithms of operator 'u' is only used for a term prefixed by the operator 'u' and marked as a lead.

To produce entry (iv), a modification of the following algorithm shown in Table 7.3 (Austin, 1984:353) is needed.

Table 7.3 The algorithm of the English upward-reading connective ($w$)

<table>
<thead>
<tr>
<th>Steps</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>assemble a phrase consisting of:</td>
<td></td>
</tr>
<tr>
<td>(i) the gated term containing $w$</td>
<td>String:</td>
</tr>
<tr>
<td>(ii) the connective ($w$)</td>
<td>*1) buildings (iii)</td>
</tr>
<tr>
<td>(iii) the next permitted term earlier in the string</td>
<td>*2) damage$w$by$w$to</td>
</tr>
<tr>
<td></td>
<td>(i) (ii)</td>
</tr>
<tr>
<td></td>
<td>*3)storms</td>
</tr>
<tr>
<td></td>
<td>The third entry:</td>
</tr>
<tr>
<td></td>
<td>Storms</td>
</tr>
<tr>
<td></td>
<td>Damage to buildings</td>
</tr>
<tr>
<td></td>
<td>(i) (ii) (iii)</td>
</tr>
</tbody>
</table>

According to the requirement for the entry (iv), the above English algorithm has to be modified to the following new algorithm as seen in Table 7.4.
Table 7.4 The algorithm of the Chinese upward-reading connective ($w$)

<table>
<thead>
<tr>
<th>Steps</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>assemble a phrase consisting of:</td>
<td>String:</td>
</tr>
<tr>
<td>(i) the connective ($w$)</td>
<td>*1) 類國 (ii)</td>
</tr>
<tr>
<td>(ii) the next permitted term</td>
<td>*u) 經濟關係$w$和$w$和</td>
</tr>
<tr>
<td>earlier in the string</td>
<td>(iii) (i)</td>
</tr>
<tr>
<td>(iii) the gated term containing $w$</td>
<td>*1) 發展中國家</td>
</tr>
<tr>
<td></td>
<td>The last entry:</td>
</tr>
<tr>
<td></td>
<td>發展中國家</td>
</tr>
<tr>
<td></td>
<td>和美國經濟關係</td>
</tr>
<tr>
<td></td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>(i) (ii) (iii)</td>
</tr>
</tbody>
</table>

The modification of the related programs according to this new algorithm is presented in Appendix 7.3.

7.6 Author-attributed Associations - Operator 't'

There are three types of author-attributed associations - operator 't', i.e. 'A' compared with 'B', 'A' related to 'B', and 'A' expounded by 'B' (Austin, 1984). All of them were met in the experiment, i.e. Problem 4 'A' related to 'B', Problem 5 'A' compared with 'B', and Problem 6 'A' expounded by 'B'.

7.6.1 'A' compared with 'B'

In English, the phrase 'compared with' is placed between two terms being compared in order to form two entries, e.g.

A compared with B

B compared with A

The structure of the Chinese comparative is different from that of English. As stated in section 3.4 (5), the Chinese verbs 比 and 比较 are used for comparing. These two verbs can act as the main verbs and are placed at the end of a
sentence. Thus, the pattern of Chinese comparative construction is as follows:

\[
\text{NP1 + verb (和) + NP2 + V (比，比較)}
\]

In the above pattern, the NP1 and NP2 are limited to the subject or the topic (Li and Thompson, 1981). The coverb 和 is used in the Chinese comparative construction to link two terms, and the verb 比 (比較) (compare) is placed at the end of the entry. Thus, the comparison is written like this:

A 和 B 比較 (A with B compare)
B 和 A 比較 (B with A compare)

Consequently, the coverb 和 (with) should be added in the string, i.e. following downward-reading connective ($v$) and upward-reading connective ($w$), respectively as follows:

比較$\text{sv}$和$\text{sw}$ (compare$\text{sv}$with$\text{sw}$with)

In order to produce the entry, '和B比較' (with B compare), the algorithm of the Chinese downward-reading connective should be used.

In addition, the entry produced by the upward-reading connective should be: '和A比較' (with A compare). The algorithm of the Chinese upward-reading connective should be used for this.

It was found that all the required new algorithms (Chinese downward-reading and upward-reading connectives) for 'A compared with B' are similar to the Chinese algorithms of operator 'u' developed in section 7.5. As a result, operator 'u' will be used instead of operator 't'. However, in the
application related to this problem, operator 'u' will not be used for a lead. Thus, Example 6.1.4 of Problem 5 can be corrected using operator 'u' as follows:

Example 7.6.1

Subject: 三民主義與各種主義比較

(Three Principles of the People compared with various principles)

Chinese string:

*1) Three Principles of the People
u) compare v with w with
1) various principles

Entry:

三民主義
與各種主義比較

English word order from Chinese original string:

*1) Three Principles of the People
u) compare v with w with
1) various principles

English word order from Chinese original entry:

Three Principles of the People
With various principles compare

7.6.2 'A' related to 'B'

In Chinese, the pattern of 'A' related to 'B' is similar to 'A' compared with 'B', i.e.

A 和 B 関係 (A with B relate)
B 和 A 関係 (B with A relate)

The solution to this pattern is the same as section 7.6.1, i.e. the Chinese algorithm of operator 'u' is used rather than operator 't'. Therefore, Problem 4 in section 6.1.2 is solved as shown in Example 7.6.2.
Example 7.6.2

Subject: 呂氏春秋和名家關係

(Lushih Chunchiu related to Minchia)

Chinese string:

*1) Lushih Chunciu related to Minchia

Entries:

Chinese classics, Lushih Chunchiu with Minchia relate

Chinese classics

Lushih Chunchiu with Minchia relate

Lushih Chunchiu

With Minchia relate

Minchia

With Lushih Chunchiu relate

English word order from Chinese original strings:

*1) Classics $31 Chinese

English word order from Chinese original entries:

Classics

Chinese classics, Lushih Chunchiu with Minchia relate

Lushih Chunchiu

With Minchia relate

Minchia

With Lushih Chunchiu relate

7.6.3 'A' expounded by 'B'

In English, the pattern of operator 't' - 'A' expounded by 'B' is: 't)$vexpounded by$wexpounding'. Two forms of the term
'expound' are used here, i.e. 'expounded' and 'expounding'. In Chinese the verb 說明 (expound) has only one form as the Chinese language is a non-inflectional language.

In the pattern of operator 't' - 'A' expounded by 'B', the Chinese passive construction is involved. Consequently, the algorithm for the Chinese passive voice that has been developed in section 7.3 should be used to solve the problem related to operator 't' - 'A' expounded by 'B'. In addition, a blank connective should be used in the upward-reading connective since the action term 說明 (expound) and the noun 意義 (meaning) should be linked without a coverb. As a result, operator '2' is used instead of operator 't'.

Thus, the new pattern to be used in Chinese is '2)説明$v被$w' (2)expound$vby$w) and it is shown in Example 7.6.3.

Example 7.6.3

Subject: 意義被資訊理論説明

(Meaning expounded by information theory)

Chinese string:

*1)意義
2)説明$v被$w
*3)資訊理論

Entries:

意義
被資訊理論説明
資訊理論
説明意義

English word order from Chinese original string:

*1)meaning
2)expound$vby$w
*3)information theory
English word order from Chinese original entries:

Meaning
  By information theory expound

Information theory
  Expound meaning

7.7 Following Differences (Postpositions)

In the Chinese language, some morphemes can be placed following a subordinated substantive, for example 前 (front, before), 後 (behind, after), 上 (on, above) and 下 (below, under). They are not free words, and they should follow a noun or a verb to form a compound term, for example 生產前 (childbirth before), 移植後 (transplant after), and 手術後 (operation after). As stated in 6.5.2, these morphemes are called postpositions. This special feature is not found in English.

The following differences that were discussed in the Manual are different from the Chinese ones. In English, the following differences are related to compound concepts and are "sometimes expressed in the form of prepositional phrases, such as 'Management by objectives' or 'Games for children'" (Austin, 1984:63). Thus, the following differences consider only the formation of English prepositional phrases. The prepositional phrases are expressed by preceding differences codes. For example, the prepositional phrase 'hospitals for children' is coded as:

*1)children$21hospitals for (Austin, 1984:63)

In Austin's (1982: 253-257) thesis, both prepositions and postpositions were discussed in some languages, such as Hebrew, Finnish and Hungarian. In the Manual, a grid of connectives
($v$ and $w$) was proposed to extend the original functions for using in multilingual texts, including prepositions and postpositions as seen in Table 7.5 (Austin, 1984:300).

Table 7.5: Prepositional and postpositional codes

<table>
<thead>
<tr>
<th></th>
<th>Space-generating</th>
<th>Close-up</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prepositional</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Postpositional</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The postpositions are included in the above grid, but they are used for connectives ($v$, $w$) and not for differencing.

Consequently, none of the available codes can be used for the Chinese postpositions. Some other role operators or codes are needed for this. Two codes, '$4'$ and '$5$', were suggested by F. Smith for the following differences for the Italian language, but the algorithms are not totally suitable for Chinese. Two examples below illustrate the unsuitability of '$4'$ for Chinese non-lead and '$5'$ for Chinese lead, respectively.

Example 7.7.1
Subject: 日本地震後建築物重建
(Reconstruction of buildings after earthquakes in Japan)

English string:

*0)Japan
*1)buildings
*2)earthquakes$4$after
*2)reconstruction
Entries:

(1) Japan

(2) Buildings. Japan
   Earthquakes after. Reconstruction

(3) Earthquakes. Buildings. Japan
   Earthquakes after. Reconstruction

(4) Earthquakes after. Buildings. Japan
   Reconstruction

(5) Reconstruction. Buildings. Japan

Two entries, i.e. entries (4) and (5) in the above example are not suitable for Chinese. The entry (4) results from the use of $4 for non-lead is not needed in Chinese. The term 'earthquakes after' does not appear in entry (5).

Example 7.7.2

English string:

*0) Japan
*2) earthquakes$51 after
*2) reconstruction
*3) buildings

Entries:

(1) Japan
   Buildings. Earthquakes after. Reconstruction

(2) Buildings. Japan
   Earthquakes after. Reconstruction

(3) Earthquakes. Buildings. Japan
   Earthquakes after. Reconstruction

(4) Earthquakes after. Buildings. Japan
   Reconstruction

(5) After. Buildings. Japan
   Earthquakes after. Reconstruction

(6) Reconstruction. Buildings. Japan

Two entries in the above example are unacceptable in Chinese, i.e. entries (5), and (6). As mentioned before, the morpheme
'後' (after) cannot be used independently but should follow other verbs or nouns. In other words, a postposition alone cannot be a lead except combined with another term. Consequently, entry (5) has to be removed. The problem of entry (6) is similar to the problem of entry (5) of Example 7.7.1, i.e., one term 'earthquakes after' does not appear.

As a result, the algorithms of following differences for the Italian language suggested by F. Smith should be modified to suit Chinese requirements. In addition, the related programs should also be modified to remove the unnecessary spaces resulting from the use of the codes of the following differences in Chinese entries. The whole modification can be seen in Appendix 7.4.

Applying the new algorithms of following differencing, problems encountered in Examples 7.7.1 and 7.7.2 have been resolved as seen below:

Example 7.7.3

Chinese string: English word order from Chinese original string:

*0)日本
*1)建築物
*2)地震$$41後
*2)重建

Entries:

日本 建築物 地震後 重建

建築物 日本 地震後 重建

地震 建築物 日本 地震後 重建

重建 地震後 建築物 日本
English word order from Chinese original entries:

Japan
Buildings. Earthquakes after. Reconstruction

Buildings. Japan
Earthquakes after. Reconstruction

Earthquakes. Buildings. Japan
Earthquakes after. Reconstruction


Example 7.7.4

Chinese string:

*0) 日本
*1) 建築物
*2) 地震後

English word order from Chinese original string:

*0) Japan
*1) buildings
*2) earthquakes after

Entries:

日本
建築物。地震後。重建

建築物。日本
地震後。重建

地震。建築物。日本
地震後。重建

地震後。建築物。日本
重建

重建。地震後。建築物。日本

English word order from Chinese original entries:

Japan
Buildings. Earthquakes after. Reconstruction

Buildings. Japan
Earthquakes after. Reconstruction

Earthquakes. Buildings. Japan
Earthquakes after. Reconstruction

Earthquakes after. Buildings. Japan
Reconstruction

The codes of preceding differences are used in English and other languages to make up a compound term as shown in Table 7.6 (Austin, 1984:48).

### Table 7.6 The first number in standard differencing codes

<table>
<thead>
<tr>
<th>Non-lead</th>
<th>Space-generating</th>
<th></th>
<th>Close-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Lead</td>
<td>2</td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

In the above grid, figures '0', '2' are used to generate spaces, while figures '1', '3' are used for close-up, i.e. no space will be generated between two terms. Two figures '1' and '3' are used in Chinese because codes of space-generating ('0', '2') are not applicable for Chinese strings. Thus, in Chinese, figures '0' and '2' are redundant.

In order to produce a new grid which is particularly used for Chinese differences, both preceding differences and following differences are all included in this new grid. As a result, a new grid is suggested for Chinese differences, including preceding differences and following differences as follows:

### Table 7.7: Chinese differencing codes

<table>
<thead>
<tr>
<th></th>
<th>Non-lead</th>
<th>Lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preceding differences</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Following differences</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
7.8 Typography

PRECIS is used for producing printed subject indexes, therefore some typographic matters are preset, such as boldface, capitalization and italics. Owing to the characteristics of the Chinese language, these matters have to be discussed for producing printed Chinese subject indexes.

7.8.1 Boldface

A distinctive typeface is selected for the lead, e.g. boldface or upper case (Austin, 1984). Boldface is used in the British National Bibliography. In Chinese, boldface cannot be produced because, at present, Chinese characters are regarded as graphics by computers. As a result, the Chinese National Bibliography and the Index to Chinese Periodical Literature cannot be printed in boldface.

7.8.2 Capitalization

Both upper and lower cases are used in English. By contrast, Chinese has no upper and lower cases differences. In English names, for example, 'William Shakespeare' the first letter of each name should use upper case. Translation of this name into Chinese is 威廉·莎士比亚 and thus there are no upper and lower cases differences.

In English entries, both upper and lower cases are used, i.e. upper case is used in the first letter of the lead, qualifier and display. The term coded 't' uses a lower-case initial in the entry.
Application of the English version programs in Chinese would produce the following results: 三 was changed to 両, and 中 to 奶. Thus, the first Chinese characters of the entries were changed to other different characters due to the capitalization effect of the English programs.

Modification of the related part of the programs is the same as that of spacing mentioned in the previous section, i.e. the first byte of the internal code is used (as seen in Appendix 7.5).

7.8.3 Italics

Italics are used in some operators, such as operators 't', '4' to '6', '$d' and typographic codes ('$e', '$f', '$h' '$i') (Austin, 1984:40). Italics are not used in the Chinese language. Some other fonts are used instead, such as 'Ming font' (明體) and 'Sung font' (宋體). However, in the experiment these fonts were not considered for producing Chinese entries due to the limitations of the hardware and software used for the experiment.

Problems 4 and 5 mentioned in Chapter 6 showed that the use of operator 't' led to the appearance of the punctuation '%' in the produced Chinese entries. Since the present study has decided to use operator 'u' in Chinese instead of operator 't', the problem with '%' no longer exists.

7.8.4 Punctuation

Punctuation that are available in the experiment programs are full stop '.', comma ',', colon ':' and hyphen '-'. These
punctuation codes use ASCII code and hence each punctuation is represented by one-byte. However, every Chinese character or punctuation is represented by two-byte or three-byte depending on which kind of internal code is used as mentioned in 4.2. The two-byte code, BIG-5, was used in this experiment. In order to manipulate Chinese records consistently, the ASCII code for punctuation mentioned above have been changed to Chinese punctuation (as seen in Appendix 7.6), i.e. full stop ' . ', comma ',', colon ':', and hyphen '－'. In addition, the space in Chinese is also represented by two-byte code '·'. In this study, each Chinese punctuation requires a two-byte representation. The punctuation differences between English and Chinese are shown in Table 7.8.

Table 7.8 The punctuation differences between English and Chinese

<table>
<thead>
<tr>
<th>Punctuation</th>
<th>English (one-byte)</th>
<th>Chinese (two-byte)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full stop</td>
<td>' . '</td>
<td>'·'</td>
</tr>
<tr>
<td>Comma</td>
<td>','</td>
<td>','</td>
</tr>
<tr>
<td>Colon</td>
<td>':'</td>
<td>':'</td>
</tr>
<tr>
<td>Hyphen</td>
<td>'－'</td>
<td>'－'</td>
</tr>
<tr>
<td>Space</td>
<td>'·'</td>
<td>'·'</td>
</tr>
</tbody>
</table>

Consequently, the format of Chinese entries have to be modified as follows:

1) full stop:
   (1) following a lead term

   Chinese full stop ' . ' and Chinese space '·' are used between lead and qualifier, e.g.

   終濟關係。美國
   和發展中國家
(2) within a qualifier or a display

Chinese full stop '、' is used, e.g.

輸血。 人
治療。博士的角色

2) comma: Chinese comma '、' and Chinese space ' ' are used in a date, coordinate concepts, $e$, and $f$, e.g.

臺灣
公共圖書館，1895-1945。研究

3) hyphen: two Chinese spaces, ' ' are placed before and after Chinese hyphen '—', since Chinese hyphen is similar to the Chinese figure, one '一'. They are used in operators '4' to '6', e.g.

交通設施。福建和浙江。中國
1986。整建 — 調查

4) colon: Chinese colon '：' and Chinese space ' ' are used in operator 'q', e.g.

臺灣
墾丁。國家公園：墾丁國家公園。史前文化

7.9 Chapter Summary

Problems found are mainly related to agents of transitive actions, coordinate concepts ('f', 'g'), connectives ($v$, $w$), special classes of action ('s', 't', 'u'), and following differences (postpositions). The other operators are all acceptable in Chinese, i.e.

- core concepts ('0' to '2')
- extra-core concepts ('4' to '6')
- dependent elements ('p', 'q', 'r')
- theme interlinks ('$x$', '$y$', '$z$')
A summary of all discussion is shown in Table 7.9.

Table 7.9 Solutions to the problems

<table>
<thead>
<tr>
<th>Sections</th>
<th>Role operators</th>
<th>Problems' Location</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Connective codes</td>
<td>1, 3</td>
<td>the related programs have been modified to remove unnecessary spaces</td>
</tr>
<tr>
<td>7.2</td>
<td>Coordinate concepts</td>
<td>2, 7</td>
<td>the ampersand '&amp;' has been changed to Chinese connective '和' (and), and the programs developed in section 7.1 are used to remove unnecessary spaces before and after the connective '和' (and)</td>
</tr>
<tr>
<td>7.3</td>
<td>Agents of transitive actions</td>
<td>15</td>
<td>1. a new algorithm for the Chinese passive construction, i.e. Chinese downward-reading connective algorithm, has been developed 2. a blank connective is used in the upward-reading connective to link two terms</td>
</tr>
<tr>
<td>7.4</td>
<td>Role definers - operator 's'</td>
<td>3, 10</td>
<td>the English original pattern 's)applications$of$win' has been changed to Chinese pattern 's)應用$於$in' (s)applications$of$win</td>
</tr>
<tr>
<td>7.4.1</td>
<td>'Applications' and 'use'</td>
<td></td>
<td>the original algorithm 's)influence$of$win' will not be used, instead a new algorithm '2)影響$於$被$w$ (2)influence$by$w' is developed</td>
</tr>
<tr>
<td>7.4.2</td>
<td>'Influence'</td>
<td>8</td>
<td>the original algorithm 's)influence$of$win' will not be used, instead a new algorithm '2)影響$於$被$w$ (2)influence$by$w' is developed</td>
</tr>
<tr>
<td>7.4.3</td>
<td>'Effects'</td>
<td>9</td>
<td>these interlinks are used instead of operator 's'</td>
</tr>
<tr>
<td>7.5</td>
<td>Two-way interactions - operator 'u'</td>
<td>16</td>
<td>1. the Chinese downward-reading connective algorithm is used to produce the first entry 2. the original algorithm of operator 'u' is only used for a term prefixed by the operator 'u' and marked as a lead 3. a new algorithm of Chinese upward-reading connective (sw) has been developed for the last entry</td>
</tr>
</tbody>
</table>

(continued)
7.6 Author-attributed associations - operator 't'

7.6.1 'A' compared with 'B' 5  
the new algorithm 'u) 比較 $v$和 $w$ and ' 
(v)compared$\text{with}$$\text{with}$ developed in section 7.5  
is used instead of the original algorithm  
't)compared with', but terms 比較 (compare)  
and 類似 (relate) will not be used as leads

7.6.2 'A' related to 'B' 4  

7.6.3 'A' expounded by 'B' 6  
using the Chinese passive algorithm developed in  
section 7.3, a new algorithm '2) 明 明 $v$和 $w$  
(2)expounded$\text{by}$' has been developed; thus the  
original algorithm 't)expounded by$\text{by}$expounding'  
will not be used in Chinese

7.7 Following Differences (Postpositions) 14  
two new codes '4' and '5' adopted from the  
PRECIS program for Italian language are  
adjusted to fit the Chinese postpositions

7.8 Typography

7.8.1 Boldface  
it is suggested to use the appropriate hardware  
and software to produce boldface

7.8.2 Capitalization  
since Chinese characters do not have upper  
and lower case differences, the related parts of  
the programs are modified

7.8.3 Italics 4, 5  
Problem 4 and 5 caused by italics have been  
resolved, since operator 't' will not be used in  
Chinese

7.8.4 Punctuation  
punctuation of ASCII codes are changed to  
Chinese punctuation, i.e. two-byte code

* Some discussions here were not mentioned in Chapter 6,  
therefore the field 'Problem' is blank.

The above solutions to the problems encountered in the Chinese  
language are compared with the English PRECIS role operators  
as seen in Table 7.10.
Table 7.10 Comparison of role operators used in English and Chinese

<table>
<thead>
<tr>
<th>Role operators and codes</th>
<th>English</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Location</td>
<td>n.c.</td>
</tr>
<tr>
<td>1</td>
<td>Key system</td>
<td>n.c.</td>
</tr>
<tr>
<td>2</td>
<td>Action</td>
<td>n.c.</td>
</tr>
<tr>
<td>3</td>
<td>'Direct object', 'Intake', 'Instrument', and 'Factor'</td>
<td>n.c.</td>
</tr>
<tr>
<td>3 - agents of transitive actions</td>
<td>the algorithm of downward-reading connective is:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a) the gated term containing $v$</td>
<td>a) the $\psi$ connective</td>
</tr>
<tr>
<td></td>
<td>b) the $\psi$ connective</td>
<td>b) the next permitted term later in the string</td>
</tr>
<tr>
<td></td>
<td>c) the next permitted term later in the string</td>
<td>c) the gated term containing $v$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. a blank connective is used in the upward-reading connective</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. the above algorithms are also used for the role defining term 影响 (influence), and author-attributed associations 'A' expounded by 'B'</td>
</tr>
<tr>
<td>4</td>
<td>Viewpoint-as-form</td>
<td>n.c.</td>
</tr>
<tr>
<td>5</td>
<td>Selected instance</td>
<td>n.c.</td>
</tr>
<tr>
<td>6</td>
<td>Form of document</td>
<td>n.c.</td>
</tr>
<tr>
<td>f</td>
<td>'Bound' coordinate concept</td>
<td>the coverb 影响 (and) is used instead of the ampersand &amp;</td>
</tr>
<tr>
<td>g</td>
<td>Standard coordinate concept</td>
<td>the coverb 影响 (and) is used instead of the ampersand &amp;</td>
</tr>
<tr>
<td>p</td>
<td>Part; Property</td>
<td>n.c.</td>
</tr>
<tr>
<td>q</td>
<td>Member of a quasi-generic group</td>
<td>n.c.</td>
</tr>
</tbody>
</table>

(continued)

* Note: n.c. = no change, i.e. there is no modification of the original English programs

n.u. = no use, i.e. the English role operators are not used in Chinese, and vice versa
(Table 7.10 continued)

<table>
<thead>
<tr>
<th></th>
<th>Assembly</th>
<th>n.c.</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>- applications</td>
<td>s)applications$vof$win</td>
</tr>
<tr>
<td></td>
<td>- use</td>
<td>s)use$vof$win</td>
</tr>
<tr>
<td></td>
<td>- influence</td>
<td>s)influence$vof$won</td>
</tr>
<tr>
<td></td>
<td>- effects</td>
<td>s)effects$vof$won</td>
</tr>
<tr>
<td>s</td>
<td>- attitudes</td>
<td>s)attitudes$vof$to</td>
</tr>
<tr>
<td></td>
<td>- role</td>
<td>s)role$vof$win</td>
</tr>
<tr>
<td></td>
<td>- policies</td>
<td>s)policies$vof$won</td>
</tr>
<tr>
<td></td>
<td>- participation</td>
<td>s)participation$vof$win</td>
</tr>
<tr>
<td></td>
<td>- related to</td>
<td>t)related to</td>
</tr>
<tr>
<td>t</td>
<td>- compared with</td>
<td>t)compared with</td>
</tr>
<tr>
<td></td>
<td>- expounded by</td>
<td>t)expounded by$vof$expounding</td>
</tr>
<tr>
<td>u</td>
<td>Two-way interaction</td>
<td></td>
</tr>
<tr>
<td>$x$</td>
<td>$y$</td>
<td>$z$</td>
</tr>
</tbody>
</table>
(Table 7.10 continued)

| $0 | Non-lead, space generating | n.u. |
| $1 | Non-lead, close-up | n.c. |
| $2 | Lead, space generating | n.u. |
| $3 | Lead, close-up | n.c. |

**Following differences**

| $4 | n.u. | Non-lead, close-up |
| $5 | n.u. | Lead, close-up |

$d | Dates | n.c. |

$g | Non-lead parenthetical difference | n.c. |
| $h | Lead parenthetical difference | n.c. |

Connectives

| $v | Downward-reading connective | the programs have been modified to remove unnecessary spaces |
| $w | Upward-reading connective | |

Typographic codes

| $e | Non-filing part in Italic preceded by comma | n.u. |
| $f | Filing part in italic preceded by comma | n.u. |
| $g | Filing part in roman, no preceding punctuation | n.u. |
| $h | Filing part in italic preceded by full point | n.u. |
| $i | Filing part in italic, no preceding punctuation | n.u. |

** Note: These codes are adopted from the PRECIS program for the Italian language and modified for the Chinese postpositions.**

The above table shows that there are six types of solutions that the present study has made regarding the English version PRECIS role operators and the related programs, i.e.

1) discarding the English unusable role operators;
2) using alternative role operators;
3) modifying some role operators and the related programs;
4) creating new algorithms and related programs;
5) suggesting the use of appropriate hardware and software; and
6) accepting the rest of the original role operators.

(1) Discarding the English unusable role operators

The preceding differences codes $'0'$ and $'2'$, and the whole typographic codes are all discarded as they do not have any use in Chinese. In addition, the present study has found that the operator 't' and some role defining terms of operator 's' (i.e. 影響 (influence), 態度 (attitudes), 效果 (effects), 參與 (participation), 政策 (policies), and 角色 (role)) cannot be used in Chinese, and hence decided to discard them. To overcome the failure of operators 's' and 't', the following solutions were carried out.

(2) Using alternative role operators

Theme interlinks are used in Chinese for the discarded role operator 's' related to role defining terms, i.e. 影響 (attitudes), 效果 (effects), 參與 (participation), 政策 (policies), and 角色 (role).

(3) Modifying some role operators and the related programs

Some role operators cannot be completely used in Chinese. They need some modifications to the accompanying programs. These role operators include, for example, agents of transitive actions, operator 's' - 'applications' and 'use', connectives ($v$, $w$), and operator 'u'. The modified operator 'u' replaced the functions of the discarded operator 't'. In addition, the
modified agents of transitive actions replaced operator 's' - 'influence'.

(4) Creating new algorithms and related programs

One problem encountered cannot be solved by using the available role operators, i.e. postpositions. To overcome this problem, two new codes ('$4' and '$5') originally devised for the Italian language were modified and suggested for the Chinese following differences.

(5) Suggesting the use of appropriate hardware and software

This suggestion is mainly for typographic problems. Some hardware and software can produce different fonts that can be used for boldface and italics.

(6) Accepting the rest of the original role operators

The rest role operators can be used in Chinese without any modification. They are, for example, operators ('1', '2', '4', '5', '6', 'p', 'q', and 'r'), primary codes ('$x$, '$y$, '$z$, '$a$, '$c', and '$d$), and secondary codes ('$1$, '$3$, '$d$, '$n', and '$0$).

Most of the above solutions have never been suggested before. In order to know the efficiency and effectiveness of these solutions, these solutions are compared with the previous research ones in the next chapter.
REFERENCES


Applications of PRECIS in indexing Chinese documents have encountered problems that were presented in Chapter 6. Solutions to those problems were suggested in Chapter 7. This chapter is an attempt to evaluate those solutions.

Only two of the eight articles discussed the potentiality of applying PRECIS in Chinese (described in section 2.4), i.e. those of Austin (1974) and Chor (1986). Therefore, the comparison here is restricted to the works of these two investigators. They only examined some problems and proposed the solutions.

8.1 Austin's Research

In the first edition of the Manual (Austin, 1974), ten languages were demonstrated. Two categories were separated, i.e.

- Category A: languages in which the grammatical role can be expressed by employing a full range of prepositions, whether or not inflections are also used.

- Category B: languages which either lack prepositions, or else use only a limited repertory of prepositions. In this group, the grammatical role is usually expressed by an adjectival phrase, whether or not inflections are also employed.
Austin grouped the Chinese language into Category B, as it has few prepositions and uses adjective phrases to express the grammatical role.

With regard to the lack of prepositions, the previous chapters have proved that there were only two kinds of coverb used in the Chinese strings, i.e.

1. 被 (by) and 受 (by): they were used in the passive construction, e.g. agents of transitive actions.
2. 在 (at, in): it was used in the instrumental verbs, 使用 (use) and 應用 (applications), e.g. operator 's'.

Using only one example, Austin suggested a solution to the problem brought about by the application of operator 's' - 'role' in Chinese. The example is shown below:

<table>
<thead>
<tr>
<th>Chinese string</th>
<th>English word order from Chinese original string:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(x)(1)都市地域 ✓</td>
<td>(x)(1)urban regions ✓</td>
</tr>
<tr>
<td>(y)(2)地域計画 ✓</td>
<td>(y)(2)regional planning ✓</td>
</tr>
<tr>
<td>(y)(3)社會學家任務</td>
<td>(y)(3)social scientists role</td>
</tr>
<tr>
<td>(x)(1)社會學家 ✓</td>
<td>(x)(1)social scientists ✓</td>
</tr>
<tr>
<td>(y)(p)關於都市計劃任務</td>
<td>(y)(p)related to urban planning</td>
</tr>
</tbody>
</table>

Entries:

都市地域
地域計画，社會學家任務

地域計画，都市地域
社會學家任務

社會學家
關於都市計劃任務
Theme interlinks were used in the above example, and the role defining term 'role' was translated into the Chinese term 任务. In the English-Chinese dictionary (Taluh Book, 1981), the English term 'role' was translated into two Chinese terms, i.e. 任务 and 角色. In this thesis, the latter was used (see section 7.4.3). Problem related to operator 's' - 'role' has been discussed in 7.4.3 and the solution here is similar to Austin's one, i.e. theme interlinks were used.

8.2 Chor's Research

Chor's (1986:20) research did not involve computer generation of entries and was based on the hypothesis:

"...the predicate transformation cannot be effected in Chinese strings so as to produce meaningful index entries. Consequently, a by-pass procedure using the theme interlinks is demonstrated."

From this viewpoint, she demonstrated the use of theme interlinks in order to solve the problems brought about by the following role operators, i.e. (1) role definers, (2) agents of transitive actions, (3) two-way interaction, and (4) author-attributed associations. In addition, she found that
spacing is not needed in written Chinese. The following paragraphs show the comparisons of the Chor's findings and the present study's ones. In the presentation of Chor's examples, the English word orders from Chinese original examples are given instead of the English equivalents provided by Chor. This is to show the problem more clearly to the non-Chinese speaking readers.

8.2.1 Role definers - operator 's'

One example of role definers was given by Chor as follows:

Example 8.1

Subject: 系統理論應用於國際關係研究

(Applications of systems theory in research on international relations)

Chinese string:

(x)(2)國際關係
(y)(2)研究
(y)(3)系統理論之應用
(x)(2)系統理論
(y)(2)於\國際關係研究之應用

Entries:

國際關係
------
研究：系統理論之應用

系統理論
------
於國際關係研究之應用

English word order from Chinese original string:

(x)(2)international relations
(y)(2)research
(y)(3)systems theory's applications
(x)(2)systems theory
(y)(2)in\international relations research's applications
International relations
Systems theory's applications

In the above example, theme interlinks were used. In addition, the Chinese coverb 蔦 (in, at) was enclosed in three backward slashes to indicate that it is to be printed in the index but has no filing value (Chor, 1986). The coverb 蔦 (in, at) was used in the ancient Chinese language as the function of locational relation in a verb-noun since Later Han (25 – 220 A.D.). However, this coverb has now been replaced by the coverb 在 (in, at) in modern Chinese language (Li, 1980).

In this thesis, the solution to the problems related to the role defining term 應用 (operator 's' - 'applications'), as suggested in section 7.4.1, is as follows: the English original pattern 's)applications$vo$win' has been changed to the Chinese pattern 's)應用$v$在' (s)applications$v$win). The following example shows the application of this solution.

Example 8.2

Chinese string:

*2)國際關係
2)研究
2)國際關係研究 ND2
s)應用$v$在
*3)系統理論

Entries:

國際關係
研究．應用系統理論

系統理論
應用在國際關係研究
The above example shows that the original algorithm can be used in the role defining term 應用 (applications) as the entries produced are acceptable in Chinese syntax. Three backward slashes are not used in this example because the coverb 在 (in) is placed within a Chinese phrase, 應用在國際關係研究 (Applications in international relations research). In addition, the coverb 在 (in, at) used in modern Chinese language is chosen in this research.

In order to compare the solutions suggested by Chor and those proposed by the present study, a comparison of Example 8.1 and Example 8.2 is shown in Table 8.1.
Table 8.1 Comparison of two proposed solutions to the problem related to operator 's' - 'applications'

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Example 8.1 (Chor)</th>
<th>Example 8.2 (The present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>theme interlinks</td>
<td>the English original pattern 's)applications$vofof$win' is changed to Chinese pattern 's)應用$v$在' (s)application$sv$win)</td>
</tr>
<tr>
<td>No. of terms in the string</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Coverbs</td>
<td>在 (in, at)</td>
<td>在 (in, at)</td>
</tr>
<tr>
<td>Entries</td>
<td>acceptable</td>
<td>acceptable</td>
</tr>
</tbody>
</table>

The above table shows that theme interlinks were used in Chor's research. The present study, however, used the original operator but with a little modification of the related pattern. The modification is that a blank connective is used in the downward-reading connective. Both solutions produce acceptable entries.

8.2.2 Agents of transitive actions

One example of agents of transitive actions given by Chor is (1986:27-28) as follows:

Example 8.3

Subject: 建築物被霜損壞

(Buildings damaged by frost)

Chinese string:

(x)(1)建築物 ✓
(y)(2)損壞
(y)(q)霜，損壞
(x)(1)霜 ✓
(y)(2)對，建築物，損壞
Entries:

建築物

損壞: 維霜之損壞

損壞. 建築物

維霜之損壞

Theme interlinks were used in the above example. Operator 'q' was used in one theme interlink which caused repetitions of the term 損壞 (damage), as seen in the first as well as the second entries.

In the present study (see section 7.3), the solutions to the problems related to agents of transitive actions are:

1. a new algorithm for the Chinese passive construction, i.e. Chinese downward-reading connective algorithm, has been developed;

2. a blank connective is used in the upward-reading connective to link two terms.
Applications of these solutions to the above example is shown in Example 8.4.

Example 8.4

<table>
<thead>
<tr>
<th>Chinese string:</th>
<th>English word order from Chinese original string:</th>
</tr>
</thead>
<tbody>
<tr>
<td>*1) 建築物</td>
<td>*1) buildings</td>
</tr>
<tr>
<td>*2) 損壞 $v$ 被 $w$</td>
<td>*2) damage $v$ by $w$</td>
</tr>
<tr>
<td>*3) 霜</td>
<td>*3) frost</td>
</tr>
</tbody>
</table>

Entries:

<table>
<thead>
<tr>
<th>建築物</th>
<th>Buildings</th>
</tr>
</thead>
<tbody>
<tr>
<td>損壞</td>
<td>By frost</td>
</tr>
<tr>
<td>霜</td>
<td>Damage</td>
</tr>
</tbody>
</table>

Comparison of Example 8.3 and Example 8.4 will find that the string of the latter is shorter than the former. In Example 8.4, the normal structure of the agent of transitive actions was used, i.e. object (1) - action (2) - agent (3). However, the algorithm has been changed according to the Chinese passive construction. In addition, the coverb 'by' was used in Chor's string, though another coverb '被' (by) was used in the present study's strings as it is commonly used in the Chinese passive. The comparison is shown in Table 8.2.
Table 8.2 Comparison of two proposed solutions to the problem related to agents of transitive actions

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Example 8.3 (Chor)</th>
<th>Example 8.4 (The present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>theme interlinks</td>
<td>1. a new algorithm for the Chinese passive construction is developed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. a blank connective is used in the upward-reading connective</td>
</tr>
<tr>
<td>No. of terms in the string</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Coverbs</td>
<td>经 (by)</td>
<td>被 (by)</td>
</tr>
<tr>
<td>Entries</td>
<td>duplication of the term in the produced entries due to the use of operator 'q'</td>
<td>no duplication of the term in the produced entries</td>
</tr>
</tbody>
</table>

8.2.3 Two-way interaction

In Chor's research (1986:31-32), one example of two-way interactions was illustrated as seen below:

Example 8.5
Subject: 中國與美國外交關係
(China's foreign relations with United States)

Chinese string:

(x)(l)China    \checkmark d
(y)(2)外交關係 \checkmark
(y)(q)與\美\國外交關係
(x)(1)美國    \checkmark d
(y)(2)外交關係 \checkmark
(y)(q)與\中\國外交關係
Entries:

中國
外交關係：與美國外交關係
外交關係，中國
---------
與美國外交關係
--

美國
外交關係：與中國外交關係
外交關係，美國
---------
與中國外交關係
--

English word order from Chinese original string:

(x)(1)China   ✓   d
(y)(2)foreign relations ✓
(y)(q)\with\United States foreign relations
(x)(1)United States ✓   d
(y)(2)foreign relations ✓
(y)(q)\with\China foreign relations

Entries:

China
Foreign relations: With United States foreign relations

Foreign relations. China
With United States foreign relations

United States
Foreign relations: With China foreign relations

Foreign relations. United States
With China foreign relations

As in the previous examples, theme interlinks were used in the above example. For the two separate theme interlinks, operator 'q' was used under operator '2' which caused repetitions of the term 外交關係 (foreign relations), as seen in each of the above entries.
In addition, two backward slashes were used for the coverb 與 (with). The coverb 與 (with) is a remnant from the ancient Chinese language as stated in section 7.2, while another coverb 和 (and) is more widely used in modern Chinese writing.

The present study's solutions to the problems related to the two-way interaction were proposed in section 7.5 as follows:

1. the Chinese downward-reading algorithm is used in the first entry;
2. the original algorithm of operator 'u' is only used for a term prefixed by the operator 'u' and marked as a lead term;
3. a new algorithm of Chinese upward-reading connective ($w$) has been developed for the last entry.

Applying these solutions to the above Chinese example will produce the results as follows:

Example 8.6

Chinese string:

*1)*中國
*u)*外交關係$\$和$\$和
*1)美國

Entries:

中國
和美國外交關係
外交關係. 中國
和美國
外交關係. 美國
和中國
美國
和中國外交關係
**English word order from Chinese original string:**

*1) China
*u) foreign relations
$v)$ with $w)$ with
*1) United States

**English word order from Chinese original entries:**

China
  With United States foreign relations

Foreign relations. China
  With United States

Foreign relations. United States
  With China

United States
  With China foreign relations

The original structure of operator 'u' is still used in Example 8.6, but two new algorithms were suggested due to the requirements of Chinese syntax. In addition, the coverb 和 (with) is chosen in this example.

The comparison of Example 8.5 and Example 8.6 will find that the number of terms used in the latter (3) is only half of the former (6). The other differences between Example 8.5 and Example 8.6 are shown below in Table 8.3.
Table 8.3 Comparison of two proposed solutions to the problem related to two-way interaction

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Example 8.5 (Chor)</th>
<th>Example 8.6 (The present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>theme interlinks</td>
<td></td>
</tr>
<tr>
<td>1. the algorithm of the</td>
<td></td>
<td>1. the algorithm of the Chinese</td>
</tr>
<tr>
<td>Chinese downward-</td>
<td></td>
<td>reading connective is used to</td>
</tr>
<tr>
<td>reading connective is</td>
<td></td>
<td>produce the first entry</td>
</tr>
<tr>
<td>used to produce the</td>
<td></td>
<td>2. the original algorithm of</td>
</tr>
<tr>
<td>first entry</td>
<td></td>
<td>operator 'u' is only used for a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>term prefixed by the operator 'u'</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and marked as a lead</td>
</tr>
<tr>
<td>2. the original algorithm</td>
<td></td>
<td>3. a new algorithm of Chinese</td>
</tr>
<tr>
<td>of operator 'u' is only</td>
<td></td>
<td>upward-reading connective ($w$)</td>
</tr>
<tr>
<td>used for a term</td>
<td></td>
<td>is developed for the last entry</td>
</tr>
<tr>
<td>prefixed by the operator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'u' and marked as a lead</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No. of terms in the string</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Connective terms</td>
<td>與 (and)</td>
<td>和 (and)</td>
</tr>
<tr>
<td>Entries</td>
<td>duplication of the term in the produced entries due to the use of operator 'q'</td>
<td>no duplication of the term in the produced entries</td>
</tr>
</tbody>
</table>

8.2.4 Author-attributed associations - 'A' related to 'B'

One example of author-attributed associations - 'A' related to 'B' given by Chor (1986:31-32) is presented below:

Example 8.7

Subject: 兒童學業成就與身體發展之關係

(Children's academic achievement related to physical development)
Chinese string:

(z) (1) 兒童 切
(x) (2) 學業成績 切
(y) (2) 學業成績之關係
(x) (2) 身體發育 切
(y) (2) 身體發育之關係

Entries:

(i) 兒童

學業成績. 與身體發育之關係

(ii) 學業成績. 兒童

攝身體發育之關係

(iii) 兒童

攝身體發育. 與學業成績之關係

(iv) 身體發育. 兒童

攝學業成績之關係

English word order from Chinese original string:

(z) (1) children 切
(x) (2) academic achievement 切
(y) (2) with physical development's relations
(x) (2) physical development 切
(y) (2) with academic achievement's relations

English word order from Chinese original entries:

(i) Children

Academic achievement. With physical development's relations

(ii) Academic achievement. Children

With physical development's relations

(iii) Children

Physical development. With academic achievement's relations

(iv) Physical development. Children

With academic achievement's relations
The use of two theme interlinks in the above example results in the separation of whole concepts in the displays into parts, as seen in entries (i) and (iii). In these two entries, two parts of the displays should be linked together without the full stop ' as: 學業成績與身體發展之關係 (academic achievement with physical development relations), and 身體發展與學業成績之關係 (physical development with academic achievement relations).

The operator 't' - 'A' related to 'B' was discussed in 7.6.2. The solutions to the problem related to this operator are as follows: the pattern of the Chinese operator 'u' - 'u)關係$V$和$W$和' (u)relate$V$with$W$with) is used instead of the original pattern of operator 't' - 't)related to', but the term 関係 (relate) will not act as a lead. These solutions are applied to Example 8.7, and the results are:

Example 8.8

Chinese string:

*1) 兒童
*2) 學業成績
  u) 関係$V$和$W$和
*2) 身體發展

Entries:

(i) 兒童
    學業成績, 兒童
    和身體發展

(ii) 學業成績, 兒童
    和身體發展

(iii) 身體發展, 兒童
    和學業成績

English word order from Chinese original string:

*1) children
*2) academic achievement
  u) relate$V$with$W$with
*2) physical development
English word order from Chinese original entries:

(i) Children
   Academic achievement with physical development
   relate

(ii) Academic achievement. Children
      With physical development relate

(iii) Physical development. Children
      With academic achievement relate

The algorithms used to produce the above entries are the product of the modifications have been done in Chapter 7 (see Appendix 7.2 and 7.3).

In the above example, the entry (i) 學童成績和身體發展關係
(Academic achievement with physical development relate) shows that there is no division in the display. The comparison between Example 8.7 and Example 8.8 is illustrated in Table 8.4.

Table 8.4 Comparison of two proposed solutions to the problem related to author-attributed associations

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Example 8.7 (Chor)</th>
<th>Example 8.8 (The present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>theme interlinks</td>
<td>the Chinese operator 'u' 'u)關係sv与$ww$with' (u)relate$ywith$ww with) is used instead of the original operator 't' 't)related to', but the term 關係(relate) will not be used as a lead</td>
<td></td>
</tr>
<tr>
<td>No. of terms in the string</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Connective terms</td>
<td>與 (and)</td>
<td>和 (and)</td>
</tr>
<tr>
<td>Entries</td>
<td>separation of whole concepts in the displays</td>
<td>no separation of whole concepts in the displays</td>
</tr>
</tbody>
</table>
8.2.5 The set of Greek letters ($\alpha$, $\beta$, $\gamma$)

So far, Chor used these interlinks to solve the above problems. In addition, Chor analysed how Chinese syntax differs from English and attempted to distinguish them in terms of the ordinary sequence of the existing role operators of 1 (Object) - 2 (Action) - 3 (Agent). For Chinese, she suggested the use of a special set of three Greek letters ($\alpha$, $\beta$, $\gamma$) representing the key system (object), the agent and the action, respectively. The new sequence suggested by Chor is as follows:

$$\alpha \quad -- \quad \beta \quad -- \quad \gamma$$

Object -- Agent -- Action

Chor (1986:39-41) demonstrated the use of those Greek letters using an example related to operator 's' - 'applications'. The demonstration is shown in below:

Example 8.9

**Subject:** 電腦程序編寫於文件索引編製上之應用

(Applications of programming of computer systems in indexing of documents)

<table>
<thead>
<tr>
<th>Chinese string:</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\alpha$) 文件 ✓</td>
<td>(Key System)</td>
</tr>
<tr>
<td>($\gamma$) 索引編製 ✓</td>
<td>(Action)</td>
</tr>
<tr>
<td>(sub 2↑) ($\gamma$) 文件索引編製上</td>
<td>(Substitute)</td>
</tr>
<tr>
<td>(β) 電腦$\beta$之_b於&quot;&quot; ✓</td>
<td>(Indirect Agent)</td>
</tr>
<tr>
<td>(s) 應用</td>
<td>(Role Definer)</td>
</tr>
<tr>
<td>(sub 5↓) ($\gamma$) 電腦於文件索引編製上之應用 (Substitute)</td>
<td></td>
</tr>
<tr>
<td>($\gamma$) 程序編寫 ✓</td>
<td>(Action)</td>
</tr>
</tbody>
</table>
Chinese entries:

(i) 文件
索引編製. 電腦之應用. 程序編寫

(ii) 索引編製. 文件
 電腦之應用. 程序編寫

(iii) 電腦
 於文件索引編製上之應用. 程序編寫

(iv) 程序編寫. 電腦於文件索引編製上之應用

English word order from Chinese original string:

(i) Documents
 Indexing. Computer systems' applications. Programming

(ii) Indexing. Documents
 Computer systems' applications. Programming

(iii) Computer systems
 In documents indexing's applications. Programming

(iv) Programming. Computer systems in documents indexing's applications

English word order from Chinese original entries:

(i) (α) documents ✔
 (γ) indexing ✔

(ii) (γ) documents indexing

(iii) (β) computer systems in documents indexing's applications
 (σ) applications ✔

(iv) (γ) computer systems in documents indexing's applications
 (γ) programming ✔

The above Chinese example is discussed as follows:

1. The English original sequence of object (1) - action (2) - agent (3) has been changed to the Chinese sequence of object (α) - agent (β) - action (γ).

2. The whole concepts of entry (i) and (iii) are separated into two concepts as seen below:
(1) Entry (i)

<table>
<thead>
<tr>
<th>A whole concept</th>
<th>電腦程序編寫之應用 (applications of programming of computer systems)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two separated concepts</td>
<td>文件索引編製，電腦之應用，程序編寫</td>
</tr>
<tr>
<td></td>
<td>(Computer programming systems' applications)</td>
</tr>
</tbody>
</table>

(2) Entry (iii)

<table>
<thead>
<tr>
<th>A whole concept</th>
<th>電腦程序編寫 (programming of computer systems)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two separated concepts</td>
<td>電腦 (computer systems)</td>
</tr>
<tr>
<td></td>
<td>之應用，程序編寫</td>
</tr>
<tr>
<td></td>
<td>(programming)</td>
</tr>
</tbody>
</table>

3. The algorithm employed was too complicated, i.e. the algorithm of producing the third entry is:

1. start with the upward-reading connective
2. connect with the next permitted term earlier in the string
3. connect with the downward-reading connective
4. finally connect with the next permitted term latter in the string.

String:

(file索引編製上) (2)
(β) 電腦之應用 (3) (1)
(s) 應用 (4)

Entries:

Example

<table>
<thead>
<tr>
<th>電腦程序編寫之應用，程序編寫</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) (2) (3) (4)</td>
</tr>
</tbody>
</table>
The above discussions have shown the drawbacks of using the set of Greek letters in the Chinese string. In fact, according to the discussion in section 7.4.1, the English original pattern 's)applications$vof$win' that has been changed to the Chinese pattern 's)應用$v$在' (s)applications$v$win), can be used instead of the set of Greek letters. The example below shows the solution taken by the present study.

Example 8.10

**Chinese string:**

*1) 文件
*2) 索引編製
*3) 電腦

**Entries:**

文件
索引編製
電腦

*2) 索引編製

英語

文件
索引編製

*3) 電腦

索引編製，應用

電腦程序編寫

程序編寫，應用

*1) documents
*2) indexing

英語

文件
索引編製

*3) computer systems

程序編寫，應用

*2) programming
In the comparison of Example 8.9 and Example 8.10, we can see that the latter uses the normal operators and yet its entries are more acceptable. In addition, the original sequence of role operators, object - action - agent, is still used by the present study. The comparison of Example 8.9 and Example 8.10 is illustrated below.

Table 8.5 Comparison of the use of the Greek letters and the operator 's' - 'applications'

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Example 8.9 (Chor)</th>
<th>Example 8.10 (The present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of a set of Greek letters (α, β, γ) and a complicated algorithm</td>
<td>the English original pattern 's)applications$vof$win' is changed to the Chinese pattern 's)應用$v$w在' (s)applications$v$win)</td>
<td></td>
</tr>
<tr>
<td>No. of terms in the string</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Coverbs</td>
<td>於 (in, at)</td>
<td>在 (in, at)</td>
</tr>
<tr>
<td>Entries</td>
<td>whole concepts in two entries were separated into two concepts</td>
<td>no separations of whole concepts</td>
</tr>
</tbody>
</table>
8.3 Chapter Summary

The first Chinese PRECIS sample was demonstrated by Austin (1974). He used only one example to show the problem encountered in the application of PRECIS in the Chinese language. The example was related to the role defining term 'role'. The theme interlinks were used in his solution. The present study proved that his solution is right. However, one example is not sufficient to describe all the problems encountered when PRECIS is used in the Chinese language.

The major research which followed Austin was carried out by Chor (1986). This research only used five examples which are insufficient for discovering problems comprehensively. In Chor's research, problems brought about by four role operators were resolved mainly by using theme interlinks. The four role operators are: (1) role definers, (2) agents of transitive actions, (3) two-way interaction, and (4) author-attributed associations.

By comparison, the work in this thesis is more comprehensive. It tested the application of all PRECIS role operators in Chinese, using 498 Chinese strings and fairly sophisticated programs. The results were analysed, then extensive modifications were undertaken to make PRECIS suitable for the Chinese language.

The comparison of Chor's and the present study's solutions proves that the 'theme interlinks' is not the best and the only way to solve the problems. Instead, two Chinese algorithms
were suggested in this research that are according to Chinese syntactic features, i.e., the Chinese upward-reading connective and the Chinese downward-reading connective algorithms. These two algorithms can be used in most of the above four role operators, i.e., agents of transitive actions, two-way interaction, and author-attributed associations.

Furthermore, Chor suggested the use of a set of three Greek letters ($\alpha, \beta, \gamma$) that are different from the normal role operators used in the Manual. However, some drawbacks were found related to the application of this set of Greek letters.

In summary, the comparison of Chor's and the author's solutions to the problems related to four role operators are presented in Table 8.6.

Table 8.6 Comparisons of Chor's and the author's solutions to the problems brought about by the role operators

<table>
<thead>
<tr>
<th>Role operators</th>
<th>Chor's research</th>
<th>The present study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role definers - operator 's applications'</td>
<td>Solution 1: theme interlinks</td>
<td>the English original pattern 's)applications$vo$win' is changed to Chinese pattern 's)應用$m$在' (s)applications$w$win</td>
</tr>
<tr>
<td></td>
<td>No. of terms in the string</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Entries</td>
<td>acceptable</td>
</tr>
<tr>
<td></td>
<td>Solutions 2</td>
<td>$\alpha, \beta, \gamma$</td>
</tr>
<tr>
<td></td>
<td>No. of terms in the string</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Entries</td>
<td>whole concepts in two entries were separated into two concepts</td>
</tr>
</tbody>
</table>

(continued)
(Table 8.6 continued)

<table>
<thead>
<tr>
<th>Agents of transitive actions</th>
<th>Solutions</th>
<th>Theme Interlinks</th>
<th>1. a new algorithm for the Chinese passive construction is developed 2. a blank connective is used in the upward-reading connective</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of terms in the string</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entries</td>
<td>duplication of term in the produced entries due to the use of operator 'q'</td>
<td>no duplication of term in the produced entries</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Two-way interaction</th>
<th>Solutions</th>
<th>Theme Interlinks</th>
<th>1. the algorithm of the Chinese downward-reading connective is used to produce the first entry 2. the original algorithm of operator 'u' is only used for a term prefixed by the operator 'u' and marked as a lead 3. a new algorithm of the Chinese upward-reading connective is developed for the last entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of terms in the string</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entries</td>
<td>duplication of a term in the entries due to the use of operator 'q'</td>
<td>no duplication of a term in the entries</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author-attributed associations - 'A' related 'B'</th>
<th>Solutions</th>
<th>Theme Interlinks</th>
<th>the Chinese operator 'u' (u)related to' (u)is used instead of the original operator 't' - 't)related to', but the term (relate) will not be used as a lead</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of terms in the string</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entries</td>
<td>separation of whole concepts in the displays</td>
<td>no separations of whole concepts in the displays</td>
<td></td>
</tr>
</tbody>
</table>
The above table reveals that three out of five solutions suggested by Chor are inefficient, since more terms are used in Chor's strings than in the present study's strings. Furthermore, all but one solution suggested by Chor's entries have some weaknesses, i.e. a whole concept was separated into two concepts and duplication of terms in the entries. On the other hand, the entries produced by the present study have no these weaknesses. In other words, the present study's solutions are more effective.

The previous two chapters have described problems encountered during the experiment, and suggested solutions to the problems. In the present chapter solutions were compared with those of previous research and found more effective and efficient. The next chapter will demonstrate the whole procedure of manipulating Chinese index entries.
REFERENCES


CHAPTER 9

GENERATION OF CHINESE ENTRIES

This chapter describes the procedure involved in creating a PRECIS Chinese index system. The procedure includes five stages, i.e. (1) input of Chinese strings, (2) production of Chinese entries, (3) production of Chinese sort keys, (4) sorting and merging, and (5) printout of Chinese entries. In total, six programs were included for generating entries. Five programs were written by F. Smith for generation of PRECIS index in English. In addition, the sorting program was modified by F. Smith from the program which was written by D. A. Harrington of Digico. These programs were modified for generation of PRECIS index in Chinese. The flowchart of the whole procedure is seen in Figure 9.1. These stages will be described respectively.

9.1 Input of Chinese Strings

Before inputting the data, a file has to be created using a word processor. The word processor, GALAXY, was used in the experiment because it can process both Chinese and English characters in the Chinese system, ETien Chinese System. Records can be input when the file has been created. The elements of input include: number of string, subject and string. The program allows editing on screen when errors were found. The results of online input and output are illustrated below in Figure 9.2.
<table>
<thead>
<tr>
<th>Stages</th>
<th>Programs</th>
<th>File names</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Input of Chinese strings</td>
<td>Fpflinle</td>
<td>-- Fpflinle</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Filename: Leprec7.tra</td>
</tr>
<tr>
<td>(2) Production of Chinese entries</td>
<td>Fprecpr4</td>
<td>-- Filename: Leprec7.tra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Output file: Lepreent.tra</td>
</tr>
<tr>
<td>(3) Production of Chinese sort keys</td>
<td>Fleekey</td>
<td>-- Entry file: Lepreent.tra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Sortkey file: Leprekey.tra</td>
</tr>
<tr>
<td>(4) Sorting</td>
<td>Fdahsort</td>
<td>-- Input name of input file: Leprekey.tra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Input name of output file: Leprekey.idx</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Enter A,B,C,D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Work file: Srtkey.tra</td>
</tr>
<tr>
<td>(4) Merging</td>
<td>Fmergseq</td>
<td>(sequential file)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Leprekey.idx)</td>
</tr>
<tr>
<td>(5) Printout of Chinese entries</td>
<td>Fpentr</td>
<td>-- Data file: Lepreent.tra</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-- Index file: Leprekey.idx</td>
</tr>
</tbody>
</table>

Figure 9.1 Five stages of generating Chinese PRECIS index entries
---

Number of string

---

0002 三民主義的次序和結構  Subject  String

1 - *1)三民主義  Input string
2 - *p)次序$y和
3 - *g)結構

$z1103三民主義$zp103次序$y和$zg103結構  Manipulation codes

Entries

三民主義
次序和結構
次序，三民主義
結構，三民主義

---

Output entries

Figure 9.2: Online input and output of a record

In addition, more input and output records are shown in Appendix 9.1. After inputting records, those records were sent to a file. The length of each record in this file was fixed, i.e. 128 bytes. The structure of the record is shown in Figure 9.3 and more records are shown in Appendix 9.2.

" 124 bytes  "|CR|LF|

0710002**三民主義的次序和結構**$z1103三民主義...  %0001

---

<table>
<thead>
<tr>
<th>entry number</th>
<th>manipulation code</th>
<th>end of record</th>
</tr>
</thead>
<tbody>
<tr>
<td>length of subject string</td>
<td>string number</td>
<td></td>
</tr>
</tbody>
</table>

CR: carriage return (ASCII 13)
LF: line feed (ASCII 10)

Figure 9.3: The structure of the record
9.2 Production of Chinese Entries

Records produced in the input stage include the whole string of Chinese samples. In this stage, however, these records have to be manipulated to produce separate entries. For example, one entry (see Example 7.3.2) in Figure 9.4 and more entries are seen in Appendix 9.3.

Example 7.3.2

```
030經濟制裁。 南非米被日本十五0004
```

Figure 9.4 The structure of the entry

9.3 Production of Chinese Sort Keys

This stage is mainly for producing sort keys. Six elements were used to determine the sorting order of characters, i.e.

1). nothing
2). '*' : to separate Lead, Qualifier and Display
3). '-' : for operators '4', '5' and '6'
4). ', ' : for date ($d$)
5). '.' : for all other foci

In the program, the numbers 0-9 were represented by "MNOPQRSTUVWXYZ" respectively. The results of a sorted record is shown below (more records are seen in Appendix 9.4):

```
"三民主義，次序和結構，000002"
```

Figure 9.4 The structure of the entry
9.4 Sorting and Merging

9.4.1 Sorting

The procedure of sorting was according to sort keys that have been produced in the previous stage. Four elements have to be chosen, i.e.

A= 2 for alphabetic and Chinese internal codes order
B= start position
C= end position
D= always 2 for ascending order

Every 100 entries were set as a unit for sorting because this number was a convenient size. The results of every 100 sorted entries were stored in the Work File (intermediate file).

The sorting methods of Chinese characters differ from English alphabetic order. Therefore, various Chinese sorting methods have to be discussed first. Chinese characters are normally sorted in three ways: strokes, radicals, and phonetics. Both strokes and radicals are usually mixed, namely sorted by radical components then strokes or arranged by numbers of strokes and followed by radical components. The phonetic system is separated into two kinds: one is the Pinyin system used in China, another is the Mandarin Phonetic Symbol used in Taiwan. These four sorting methods are described as follows:
(1) Sorting by radicals and strokes

The following characters, for example, 人, 木, 伎, 吐, 吹, and 口, are sorted by two radicals: 人 'men', and 口 'mouth', and then arranged by the total number of strokes in ascending order as seen below in Table 9.1.

Table 9.1 Example of sorting by radicals and numbers of total strokes

<table>
<thead>
<tr>
<th>Characters</th>
<th>Radicals</th>
<th>Number of total strokes</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>人</td>
<td>人</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>伎</td>
<td>口</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>吐</td>
<td>口</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>吹</td>
<td>口</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>咖</td>
<td>口</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

(2) Sorting by strokes and radicals

In this method, Chinese characters are sorted first by number of strokes, and then by radicals as seen below in Table 9.2.

Table 9.2 Example of sorting by numbers of total strokes and radicals

<table>
<thead>
<tr>
<th>Number of total strokes</th>
<th>Characters</th>
<th>Radicals</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>付, 付, 付, 付</td>
<td>人</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>本, 本, 本, 本</td>
<td>木</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>忍, 忍, 忍, 忍</td>
<td>心</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>岩, 岩, 岩, 岩</td>
<td>山</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>
(3) Sorting by Pinyin system

The sequence of the Pinyin system is according to Chinese characters' phonetic order which is similar to English, i.e. the alphabetical order as shown in Table 9.3 (Chinese Academic Sinica of Social Sciences, 1987).

Table 9.3 Example of sorting by Pinyin system

<table>
<thead>
<tr>
<th>Characters</th>
<th>Pinyin</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>bao</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>cong</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>ding</td>
<td>7</td>
<td>8</td>
</tr>
</tbody>
</table>

(4) Sorting by Mandarin Phonetic Symbol

Mandarin Phonetic Symbol has its own sequence in which 36 symbols are used, such as ㄅ, ㄆ, ㄇ, ㄈ, ㄉ, and so on. In addition, each syllable is pronounced with one of four tones, i.e. high level '－', high rising '↗', dipping '↙' and high falling '↘' (Li and Thompson, 1981). In some dictionaries, characters with the same pronunciation are arranged according to their respective radicals (Liang, 1984). The arrangement of this system is shown in Table 9.4.
As stated before, the Chinese internal code, BIG-5, was used in the experiment. The arrangement of this code was according to numbers of strokes and then followed by radicals, i.e. the fewer number of strokes was placed first and vice versa. For example, the result of sorted records is shown below (more sorted records are shown in Appendix 9.5):

- "0008" -- record 1 has the lowest value
- "0133" -- record 2 has next lowest value
- "0098" = CARITMTCW
- "0144" = CD-ROM
- "0002" = CD-ROM

The equivalent records to the above positions are shown below:

"0008" = CARITMTCW
"0133" = CD-ROM
"0098" = CARITMTCW
"0144" = CD-ROM
"0002" = CD-ROM

As Table 9.4 shows, the Mandarin Phonetic Symbol sequences are used to sort the records.
The first Chinese characters of the above entries are '人', and '日'. Their number of strokes are two and three respectively. If the first character's numbers of strokes is similar, it is then compared with the second character, and so on. In addition, if a file has mixed English and Chinese entries, the sequence of sorting English index entries will be arranged in front by alphabetical order, and followed by Chinese entries by number of strokes.

9.4.2 Merging

As mentioned before, every 100 entries were sorted as a unit. This stage is merging these separate units to form a whole file and prepare for printout as illustrated below in Figure 9.5.

![Figure 9.5: Procedure of merging](image-url)
9.5 Printout of Chinese Entries

This stage is the last stage which prints the sorted Chinese records. The file of printout is based on the merged results, i.e., index file, matching the file of 'production of entries'. Finally, the sorted and merged Chinese entries are printed (as seen in Appendix 9.6).

In summary, the whole procedure of Chinese index manipulation has been demonstrated in this chapter. The next chapter is the conclusions to the study and the recommendations are suggested.
REFERENCES


This chapter concludes the present study by reviewing the objectives and achievements and also proposing suggestions regarding the future use and research of PRECIS.

10.1 Conclusions

The present study was motivated by the following facts: (1) PRECIS has a linguistic universal feature for computerized subject indexing; (2) the largest Chinese bibliography and index published by the National Central Library of Taiwan still lack subject indexes; (3) both mainland China and Taiwan have created their bibliographic databases based on UNIMARC; and (4) the field 670 of the UNIMARC is reserved for PRECIS.

The present study has aimed to:

(1) experiment with PRECIS for indexing Chinese documents;
(2) generate Chinese subject indexes using PRECIS; and
(3) suggest the use of PRECIS in online retrieval on Chinese bibliographic databases across different Chinese-speaking countries.

The PRECIS system includes two major components, i.e. syntax and semantics. The semantics aspect is related to vocabulary control, namely the creation and maintenance of a thesaurus. Unlike the syntactical component, PRECIS thesaural procedures are based upon rules and relationships that are recognized in international standards (ISO, 1986), therefore they can be applied in other languages without modification. As the
syntactic features are different between English and Chinese, the syntactic factors are mainly considered in this study.

To achieve the first two objectives, pilot and main experiments were carried out mainly according to the following stages, i.e.

(1) analysis of subject statements;
(2) coding;
(3) input, computer processing, and output;
(4) analysis of output based on Chinese syntactic rules;
(5) modifications of role operators and the accompanying programs;
(6) comparison of the findings of the present and previous research;
(7) demonstration of Chinese index manipulation.

In the pilot experiment, a romanization based on the Wade-Giles system was carried out, that is to help the input and output of Chinese characters that otherwise could not be performed due to the unavailability of the Chinese computer system. At this stage, the English version PRECIS programs were tested and a format for input and output was set up. In the main experiment, 492 Chinese titles derived randomly from one issue of a Chinese index, and six strings chosen and translated from the Austin's Manual were analysed, coded, and input using the ETien Chinese Computer system.

Due to the fact that PRECIS was originally devised for indexing in English and the grammatical functions in English and Chinese are expressed in different ways, the attempt to switch this
system from English to Chinese encountered some problems.

Eighteen problem areas were found during the experiment that were grouped into eight categories, i.e. 1. connective codes, 2. coordinate concepts, 3. agents of transitive actions, 4. role definers, 5. two-way interaction, 6. author-attributed associations, 7. following differences, and 8. typography (see chapter 6).

The solutions to those problems are summarized as follows:

The problem brought about by connective codes, i.e. spacing problem, has been resolved by the modification of the related programs.

The problem of coordinate concepts arose because the ampersand (&) is not used in Chinese. Consequently, the ampersand was changed to a Chinese coverb 和 (and).

The passive voice is involved in agents of transitive actions. Since the Chinese passive construction differs from English one and hence the English algorithm cannot be applied in Chinese, a new algorithm of the downward-reading connective for the Chinese passive construction was suggested to resolve the problem.

Eight role defining terms of operator 's' were divided into four groups in Chinese, according to the characteristics of the problems they brought about. Consequently, different solutions were suggested for them, i.e. using the original operator 's' (i.e. for 'applications' and 'use'); Chinese agents of transitive actions (i.e. for 'influence'); and theme
interlinks (i.e. for 'effects', 'attitudes', 'role', 'policies', and 'participation').

Two algorithms used in operator 'u' cannot be applied in Chinese strings since English syntactic structures are different from Chinese ones. Consequently, they were replaced by two new algorithms, i.e. the Chinese downward-reading and the Chinese upward-reading connective algorithms. These Chinese algorithms of operator 'u' were used to solve Problem 16 (two-way interaction), and Problem 4 and 5 (i.e. author-attributed associations - 'A' related to 'B' and 'A' compared with 'B'). Problem related to 't' - 'A' expounded by algorithm of 'B' resolved by the Chinese agents of transitive actions.

In addition, since the English following differences cannot be used for Chinese postpositions, two new codes '$4' and '$5' adopted from the PRECIS program for the Italian language were modified to overcome this problem.

Typographical problems were resolved mainly by the modification of the related programs and the suggestion of using the appropriate hardware and software.

To summarize, there were six types of actions the present study has taken regarding the English version PRECIS role operators, and the accompanying programs, i.e. (1) discarding the English unusable role operators, (2) using alternative role operators, (3) modifying some role operators and the related programs, (4) creating new algorithms and related programs, (5) suggesting the use of appropriate hardware and software, and (6) accepting the rest of the original role operators.
Most of the above solutions have never been suggested before. In order to evaluate the above solutions, the findings of the present study were compared with the solutions suggested by D. Austin and L. Chor (see chapter 8). These two research projects were chosen because they had suggested some methods to resolve problems that resulted from the application of PRECIS in the Chinese language. Comparison of the findings of the present research and those of D. Austin and L. Chor has shown that the findings of the present study can resolve most of the problems encountered and in more effective and efficient ways. Moreover, the comparison also reveals that, so far, the present study is the most comprehensive study of the application of PRECIS in Chinese, as it tested all PRECIS role operators with Chinese strings.

Finally, the whole procedure of manipulating Chinese documents was demonstrated and Chinese subject index entries have been successfully produced (see chapter 9).

Accordingly, the present study has satisfied its first objective, i.e. PRECIS can be used for indexing Chinese documents, though some modifications were needed. Based on this, the second objective has been achieved, i.e. Chinese subject indexes were generated using PRECIS. The third objective is an assumption which is based on the first objective. Since it has been proved that PRECIS can be applied in Chinese, the third objective could also be achieved if libraries and information services use this system in their Chinese bibliographic databases. Consequently, the libraries and information services in Chinese-speaking countries who
subject access to consider designing their bibliographic databases using PRECIS should bear in mind the possibility of using this system not only for printed subject indexes but also online retrieval. In the case of English language, this possibility has already become a reality.

To conclude, the present study has been able to make a theoretical contribution to the development of PRECIS, as it has been able to prove the possibility of using PRECIS in the Chinese language. The comparison of the present study and the previous research have shown that most solutions suggested by the author are better than those by the previous research, especially Chor's research. It has also a practical consequence. From now on, Chinese-speaking countries, such as mainland China, Taiwan and Hong Kong can consider the use of PRECIS to produce Chinese subject indexes for their bibliographies or indexes.

10.2 Recommendations

Suggestions for the future use of PRECIS are given based on: the experience of the present study, the new decision of the British Library, and the development of information technology.

(a) The experience of the present study

Based on the experience resulting from this study, it is suggested that certain considerations should be taken care in order to make the optimum use of PRECIS in Chinese documents.
There are three steps in the implementation of PRECIS that at the moment need to be carried out manually by human indexers, i.e. (1) description of subject statements, (2) analysis of subject statements, and (3) coding. This makes the implementation of PRECIS labour-intensive. In addition, staff is also needed to maintain a thesaurus. Considering these, it is felt that PRECIS is more suitable for large libraries or information services. However, small libraries or information services can also use PRECIS, provided they have enough staff to manipulate this system and maintain a thesaurus.

In addition to that, carefully-planned staff training is strongly recommended for the implementation of PRECIS, since PRECIS is "too complex/ too detailed/ requires too much intellectual effort by staff" (Bakewell, 1978:42) and "is complex to learn and use" (Foskett, 1982:275). Knowledge and skills related to PRECIS, such as the theory of PRECIS, description of subject statements, writing of strings, features of Chinese syntax, and structure of thesaurus are necessary to be acquired by the related staff.

The experiment carried out for the present study (see chapter 9) using a personal computer has shown that the manipulation of index entries needs many programs and is also time-consuming. If a library or information service has a large number of index entries, a higher speed of processing is required, and hence the use of minicomputers or mainframes is suggested instead of personal computers.
In addition, if this system is to be used to manipulate both Chinese and English languages at the same time, the programs will have to be modified.

The future PRECIS users and research should also consider the following trends of the use of PRECIS.

(b) The new decision of the British Library

The British Library has decided that the UKMARC fields 690 (PRECIS string), 691 (SIN) and 692 (RIN) will no longer be used from January 1st 1991 onwards. Instead, the subject descriptors 660 (subject topical descriptors) and 661 (subject geographical descriptors) will be used (British Library, 1990). The use of the new subject system is caused by the following limitations of the old one:

"Many of the features, such as the complex coding and system of role operators are required only to manipulate data and produce the correct typography for hard copy layout. Once data is being searched online, there is no need to identify lead terms or to manipulate the data in the same way. Neither is it necessary to repeat in a subject field data which is already present in other parts of the bibliographic record. For instance the same name can be both author and subject and therefore appear twice. A system which does away with the need for such duplication will be more economic." (British Library, 1990:3)

The shortcomings that negatively affect the efficiency of subject online retrieval arose since PRECIS was originally devised for printed subject indexes, not subject online retrieval.
The new subject system still "takes from PRECIS its basic components, namely an authority file of controlled terms and subject statements for use in bibliographic records and a network of related terms, from broader to narrower and between used and non-used" (British Library, 1990:3).

However, how much this new system will differ from PRECIS is still too early to be discussed, and thus it is open for future research.

(c) The development of Information technology and PRECIS

Some development of information technology, such as artificial intelligence (AI), might help to solve some of the above problems. An investigation has shown that indexing is a prime candidate for using expert systems (Morris, 1990).

A project is being undertaken by M. Dykstra (1990), aiming to design an expert system for automated textual analysis using PRECIS as a model. The experiment is being carried out on English language documents and includes an expanded case grammar capacity.

However, it is doubtful that a similar project can be carried out for Chinese documents, as automated analysis of Chinese sentences or even phrases is a highly complex task due to their syntactic features. The description of subject statements still need to be carried out by human indexers since the quality of PRECIS index entries relies heavily on the quality of the subject description. Expert systems might be applied only in the analysis of Chinese subject description and coding. In the application of expert systems, the knowledge of PRECIS and
Chinese syntactic features may be stored in the knowledge-based system, which in turn may help the human indexers identify a subject statement correctly (for example, the confusion of a term to be a key system or an agent, the distinction between place as location '0' and place as key system '1'). Then, a correct string may be produced by the expert system.

A true knowledge-based expert system has the capability also for machine learning. When the database has stored enough knowledge of PRECIS and Chinese syntactic rules, the expert system should be able to identify a subject statement and produce a string automatically. Subsequently, the drawbacks of 'complex to learn and use', and 'requires too much intellectual efforts by staff' can be resolved and further improvement of PRECIS using the new technology might be achieved.

(d) The use of PRECIS in online retrieval

Online retrieval is the logical method of retrieval from computerized bibliographic databases. When PRECIS is used in indexing Chinese documents, consideration should also be given to its use in online retrieval. There are important differences between searching printed subject indexes and online subject searching. Printed indexes are browsed through alphabetically, scanning entries as they appear in the subject statement at a particular point in the index. The searcher is dependent on the indexer's use and coordination of the subject vocabulary according to the rules of the system for which he is indexing. In online retrieval, the choice and coordination of terms is in the hands of the searcher. Searching and retrieval can be improved by the use of a controlled vocabulary such as a list.
of subject headings or a thesaurus. Since PRECIS has a thesaurus, i.e. the RIN file, it would be most useful if this were also available online, enabling searchers to browse through the semantic structure of terms at various stages of the search in order to determine the most effective terms to use for a particular search.

The important developments that are necessary for effective online searching of a PRECIS index are: (1) provision of a browsing facility which can display hierarchical and other relationships for terms in the thesaurus; (2) provision of a capability "to switch from an unused term input by the searcher to the preferred synonym in PRECIS" (Congreve, 1989:6); (3) provision online "to guide the user to build an effective search strategy" (Congreve, 1989:6). These developments should also be borne in mind when considering PRECIS as a system for index generation and vocabulary control in Chinese.

REFERENCES


BIBLIOGRAPHY

O1 TECHNOLOGY COMPANY (ed.) (1988) O1 Chinese system. Taipei:
O1 Technology Company (in Chinese).


international conference on computer processing of Chinese and oriental languages, pp. 531-535.


APPENDIX 2.1: PRECIS operators and codes

SCHEMA OF OPERATORS

<table>
<thead>
<tr>
<th>Primary operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment of core operators</td>
</tr>
<tr>
<td>Core concepts</td>
</tr>
</tbody>
</table>
| 0 | Location
| 1 | Key system
  | Thing when action not present
  | Thing towards which an action is directed, e.g. object of transitive action, performer of intrasitive action.
| 2 | Action; Effect of action
| 3 | Performer of transitive action (agent, instrument); Intake; Factor

<table>
<thead>
<tr>
<th>Extra-core concepts</th>
</tr>
</thead>
</table>
| 4 | Viewpoint-as-form; Aspect
| 5 | Selected instance e.g. study region, sample population
| 6 | Form of document; Target user

<table>
<thead>
<tr>
<th>Secondary operators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coordinate concepts</td>
</tr>
</tbody>
</table>
| f | 'Bound' coordinate concept
| g | Standard coordinate concept

<table>
<thead>
<tr>
<th>Dependent elements</th>
</tr>
</thead>
</table>
| p | Part; Property
| q | Member of a quasi-generic group
| r | Assembly

<table>
<thead>
<tr>
<th>Special classes of action</th>
</tr>
</thead>
</table>
| s | Role definer; Directional property
| t | Author-attributed association
| u | Two-way interaction
### Primary codes

<table>
<thead>
<tr>
<th>Theme interlinks</th>
<th>( \text{x} )</th>
<th>1st concept in coordinate theme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{y} )</td>
<td>2nd/subsequent concept in theme</td>
</tr>
<tr>
<td></td>
<td>( \text{z} )</td>
<td>Common concept</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Term codes</th>
<th>( \text{a} )</th>
<th>Common noun</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \text{c} )</td>
<td>Proper name (class-of-one)</td>
</tr>
<tr>
<td></td>
<td>( \text{d} )</td>
<td>Place name</td>
</tr>
</tbody>
</table>

### Secondary codes

#### Differences

<table>
<thead>
<tr>
<th>Preceding differences 1st and 2nd characters:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3 characters)</td>
</tr>
<tr>
<td>( \text{0} ) Non-lead, space generating</td>
</tr>
<tr>
<td>( \text{1} ) Non-lead, close-up</td>
</tr>
<tr>
<td>( \text{2} ) Lead, space generating</td>
</tr>
<tr>
<td>( \text{3} ) Lead, close-up</td>
</tr>
</tbody>
</table>

3rd character = number in the range 1 to 9 indicating level of difference

| Date as a difference \( \text{d} \) |

<table>
<thead>
<tr>
<th>Parenthetical differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{n} ) Non-lead parenthetical difference</td>
</tr>
<tr>
<td>( \text{o} ) Lead parenthetical difference</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Connectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{v} ) Downward-reading connective</td>
</tr>
<tr>
<td>( \text{w} ) Upward-reading connective</td>
</tr>
</tbody>
</table>

#### Typographic codes

<table>
<thead>
<tr>
<th>( \text{e} )</th>
<th>Non-filing part in italic preceded by comma</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{f} )</td>
<td>Filing part in italic preceded by comma</td>
</tr>
<tr>
<td>( \text{g} )</td>
<td>Filing part in roman, no preceding punctuation</td>
</tr>
<tr>
<td>( \text{h} )</td>
<td>Filing part in italic preceded by full point</td>
</tr>
<tr>
<td>( \text{i} )</td>
<td>Filing part in italic, no preceding punctuation</td>
</tr>
</tbody>
</table>
Appendix 4.1: Two-byte internal code (The BIG-5 code as example)

**The second byte (low byte)**

<table>
<thead>
<tr>
<th>-- 163</th>
<th>...... ...... ...... ...... ...... ...... ...... ......</th>
</tr>
</thead>
<tbody>
<tr>
<td>163</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>-- 164</th>
<th>64 65 66 67 68 69 70 71 72 73 74</th>
</tr>
</thead>
<tbody>
<tr>
<td>164</td>
<td>75 76 77 78 79 80 81 82 83 84 85</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The first byte</th>
<th>176 177 178 179 180 181 182 183 184 185 186</th>
</tr>
</thead>
<tbody>
<tr>
<td>high byte</td>
<td>187 188 189 190 191 192 193 194 195 196 197</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(high byte)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>164</td>
<td>86 87 88 89 90 91 92 93 94 95 96</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>97 98 99 100 101 102 103 104 105 106 107</th>
</tr>
</thead>
<tbody>
<tr>
<td>high byte</td>
<td>164</td>
</tr>
</tbody>
</table>

... ... ... ... ... ... ... ... ... ...
Appendix 4.2: The three-dimensional structure of the three-byte seven-bit code

(Yang, 1985:153)
Appendix 4.3: Structure of the Chinese operating system

Application softwares

<table>
<thead>
<tr>
<th>Application softwares</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS-DOS</td>
</tr>
<tr>
<td>Characters</td>
</tr>
</tbody>
</table>

BIOS (Basic input and output system)

<table>
<thead>
<tr>
<th>BIOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen</td>
</tr>
<tr>
<td>Printout</td>
</tr>
<tr>
<td>Disk</td>
</tr>
<tr>
<td>Communication</td>
</tr>
</tbody>
</table>

Chinese system

Chinese system drivers

Physical devices

Chinese character set

Printer

Disk driver

RS232

(Technology Company, 1988: 1-3)
Appendix 5.1 Sample PRECIS output from the pilot experiment with romanization

0042 Introduction to Chinese system's codes, input systems and characters production

String

*1) chungwen hsit'ung
*p) tzuma
*g) shujufa$v&
*g) tsu te chihtso

Entries
Chungwen hsit'ung
Tzuma, shujufa & tsu te chihtso
Tzuma. Chungwen hsit'ung
Shujufa. Chungwen hsit'ung
Tsu te chihtso. Chungwen hsit'ung

0045 Order and construction of the Three Principles of the People

String

*1) Shan ming chu i
*p) ts'ushu$v&
*g) chiehkou

Entries
Shan ming chu i
Ts'ushu & chiehkou
Ts'ushu. Shan ming chu i
Chiehkou. Shan ming chu i

0049 The Three Principles of the People compared with various principles

String

*1) Shan ming chu i
t)pichiao
1) kechung chui

Entries
Shan ming chu i
pichiao kechung chui
Appendix 5.2 Sample PRECIS output from the main experiment with Chinese

0042 附录中文系统的字词，输入法和字制作

String

1 - *1)中文系统
2 - *p)字词
3 - *g)输入法$v&
4 - *g)字的制作

$z1103中文系统$zp103字词$zg103输入法$v&$zg103字的制作

Entries
中文系统
字词，输入法 & 字的制作
字词，中文系统
输入法，中文系统
字的制作，中文系统

0045 三民主义的次序和结构

String

1 - *1)三民主义
2 - *p)次序$v&
3 - *g)结构

$z1013三民主义$zp103次序$v&$zg103结构

Entries
三民主义
次序 & 结构
次序，三民主义
结构，三民主义

0049 三民主义和各种主义的比较

String

1 - *1)三民主义
2 - *t)比较
3 - 1)各种主义

$z1103三民主义$zt003比较$z1103各种主义

Entries
三民主义
比较，各种主义
Appendix 7.1 Modification to connective codes

rem : Qualifier routine
rem Reads string from right to left - start dimension
rem is Q and end dimension is R
rem 
rem 4000 Q$=""
rem Is Lead within coordinate block?
rem 
rem : Initialise phrase flags
rem
rem 4030 H%=0
K1%=0
rem 
rem Count loop 4050-4120
rem
rem 4050 IF Q%<R% OR Q%<1 THEN 4870
rem IF MIDS(G$,Q%,1)=">" THEN 6400
rem IF MIDS(AS(5),Q%,1)<"2" THEN 4110
rem ON FNP%(Q%) GOTO 4190,4560,4580,4650,4660,4600,4600,4720,4110,4130,4160
rem 4110 Q%=Q%-1
rem GOTO 4050
rem 
rem Upward connective 'sw'
rem
rem 4130 Q2$=RS(E%)
rem IF LEFT$(R$(E%),1)>chr$(161) THEN 4150
rem IF LEN(R$(E%))<1 Then 4150_Li edited for '3rd entry without space'
rem ELSE 4155 'Li edited
rem
rem 4150 Q2$=RS(E%)
rem 'Connective without space' 'Li edited
rem H%=1
rem GOTO 4110
rem 4155 Q2$=RS(E%)
rem 'Li edited
rem
rem 4160 Q2$=RS(E%)
rem '- Connective with space
rem 4170 H%=1
rem GOTO 4110
rem 
rem Adds next focus to string with its
rem differences and any connective
rem 
rem (continued)
IF MID$(A$(2),Q%,1)<"t" Then 4300
IF LEN(R$(E%))<1 THEN 4550
IF LEN(Q$)>0 Then Q$=Q$+" "
    Q$=R$(E%)+Q$+" "
4250 IF H%>0 Then_
    Q$=Q$+Q2$ : Q2$="" : H%=0 : K2%=1
    GOTO 4110
4300 IF len(R$(E%))>0 Then_
    if len(Q5$)<1 then
        Q4$=Q4$+Q1$+Q3$: Q1$="" : Q3$="" : Q2$="": Q5$=""
    if A3%>0 THEN 4540 ' - go add focus in input order
    IF MIDS(A$(5),Q%,1)="2" THEN_ 4500
    Q%=FNO%(Q%): K%<2 ' count back for substitute
    GOTO 4500
    H%=O
    ' reset phrase flags
    K1%=1
    GOTO 4110
4530 Q4$=""
    IF MID$(A$(5),Q%,1)="2" THEN_
    Q%=FNO%(Q%)
    GOTO 4110
rem
rem Alternative concatenation routine
rem processes a predicate phrase
rem in input order starting from top of '2' block - ends up in Display
rem
4540 IF RIGHT$(Q$,1)=" " THEN Q$=Q$+Q4$ 
else Q$=Q4$+Q$ ' - Add focus in input order
    IF Q%=R% THEN 4530
    IF MIDS(A$(2),Q%,1)="q" THEN Q$=Colon$+Q$ 
    ELSE Q$=Pt$+Q$
    GOTO 4530
4550 IF LEN(Q$)>0 Then Q$=Q$+" "
    Q$=Q$+Q2$ : K2%=1
    GOTO 4500
rem
Preceding differences
(continued)
4560 Q6$=Q6$+R$(E%)$" " ' - with space
GOTO 4110
4580 Q6$=FNC$(Q6$,R$(E%))' ' - no space
GOTO 4110
rem
rem Parenthetical difference
rem
4600 Q1$=" ("+R$(E%)+")"
GOTO 4110
rem
rem Following differences
rem
4650 Q5$=R$(E%)+Q5$ 'Li del " "
GOTO 4110
rem
rem Date as difference
rem
4720 Q3$=Comma$+R$(E%)
GOTO 4110
rem
rem Tidying-up bit
rem
4870 IF LEN(Q4$)<1 THEN 4900
Q$=Q$+Q4$
4890 Q4$=""
4900 IF LEN(Q$)<1 THEN RETURN
4910 Q$=FNCLEAN$(Q$)
IF LEFT$(Q$,2)=Pt$ THEN
Q$=RIGHT$(Q$,LEN(Q$)-2):_ 
GOTO 4910
4930 P%=fnpos%(1,q$,".")
IF P%>0 THEN
Q$=LEFT$(Q$,P%-1)+RIGHT$(Q$,LEN(Q$)-(P%+1)): _
GOTO 4930
4970 IF RIGHT$(Q$,2)<>Pt$ THEN RETURN
Q$=LEFT$(Q$,LEN(Q$)-2)
GOTO 4970
4995 RETURN
rem :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
rem:
rem LEAD ROUTINE:
rem Constructs lead focus with any differences:
rem Reads from right to left - thus next counter:
rem position gives start to Qualifier - lastly:
rem it sets counter for Display:
rem:
5000 V9%=9
Q%=I%
ON FNDIFF%(Q%) GOTO 5020,5005,5010
5005 Q5%=FNFOCUS%(Q%) rem Difference 4
LS=FNLEAD4$(Q%,Q5%) : Q%=Q5%-1 : D%=Q5% : RETURN
5010 Q5%=FNFOCUS%(Q%) rem Difference 5
IF F5%>0 THEN 5015
LS=FNFOCUS5$(Q%,Q5%) : LS=FNCLEAN$(LS) : F5%=1 : _
D%=I%+1 : Q%=Q5%-1 : RETURN

(continued)
5660 IF LEFT$(R$(E%),1)>Chr$(161) Then_ 'Li added for $41
D5$=D5$+R$(E%)
Else_
D5$=D5$+""+R$(E%)
GOTO 5640
rem
rem Date as difference
rem
rem
5670 D4$=Comma$+R$(E%)
GOTO 5640
rem
rem Downward connective
rem
rem
5690 K%=1
K1%=1
D2$=R$(E%)
IF LEN(R$(E%))<1 THEN 5640
5695 IF left$(r$(e%),1)>chr$(161) Then 5697_ 'Li edited
ELSE 5770
5697 d2$=r$(e%)
GOTO 5640
rem
rem Concatenation of next focus with display
rem
rem
5770 IF K3%>0 then 5780 'new routine for 1-2-3
Act2%=FNAction%(D%)
IF Act2%>0 THEN K3%=1 : GOSUB 8500 : Goto 5640
5780 IF MID$(A$(2),D%,1)="t" AND D%=D1% THEN 5690
IF D%=D1% THEN 6250 '- adds next focus
5790 IF fnpos%(1,D1$,"$s")>0 THEN_
D$=D$+FNCS(D3$,FNS$(D1$,"s")_}
ELSE_
D$=D$+FNCS(D3$,D1$)
IF LEN(D$)>0 THEN_
D$=D$+D5$
D$=D$+D6$
D1$="" : D3$="" : D6$="" : D5$="" : K2%=0
GLO%=0
IF K%<1 THEN D$=D$+D4$ : D4$="" : GOTO 5960
IF LEN(D$)<1 AND LEFT$(D2$,1)>CHR$(161) THEN_ 'Li edited
D2$=RIGHT$(D2$,LEN(D2$))
IF LEN(D$)<1 AND LEFT$(D2$,1)<CHR$(161) THEN_ 'Li edited
D2$=RIGHT$(D2$,LEN(D2$))
D$=D$+D2$
D2$="" : K2%=1
GOTO 5960
5795 D$=D$+D4$
D4$=""
Goto 5960
-----------------------------------------------
rem
rem Process predicate and put in Display
rem
rem Note - predicate is read upwards in all
rem 's' and 'u' routines so is processed by

(continued)
rem qualifier routine
Q% = Q3%
R% = FNK%(F1%+1)  ' - find end of predicate
7050 GOSUB 4000  ' - process
IF LEN(D$) > 0 THEN D$ = D$ + ""
IF LEFT$(R$(E%), 1) > CHR$(161) THEN_
D$ = D$ + R$(FNASC%(F1%)) + Q$_
   'Li Delet " " for 'u'
ELSE_
   'in Chinese without
D$ = D$ + R$(FNASC%(F1%)) + "" + Q$_
   ' space
IF P1% > 0 THEN_
D$ = D$ + Comma$ + R$(FNASC%(P1%)) : D% = P1% + 3
GOTO 9000
Appendix 7.2 New algorithm for the Chinese downward-reading connective

rem New algorithm for the Chinese downward-reading connective
rem Redisposes Action - Agent as Agent - Action for Downward reading Predicates
rem
rem X1%=U% ' start of block (Act2% holds end of block determined by FNAction)
X3%=FNPOS%(X1%,LEFT$(A$(2),Act2%),"3") 'finds start of 'Agent

IF uflag%>0 then X3% = Act2%
IF X3%<1 THEN RETURN
D3%=D2% : D4%=D% : D5%=Act2% 'store true length of 'Display
DO WHILE X4%>X1%
IF FNPOS%(l,"vw",MID$(A$(1},X4%,l} }<1 THEN EXIT
LOOP
IF MID$(A$(1},X4%,1}="v" THEN X3%=X4% : EXIT
LOOP
x4%=x4%-1
LOOP
IF len(D1$)>0 AND uflag%<1 then d1$=d1$+Pt$
D2%=Act2% : D%=X3% : k%=1 : k1%=1
GOSUB 5520
IF D%>D5% then D5%=D%
D2%=X3%-1 : D%=D4% : k%=1 : k1%=1
GOSUB 5520
D2%=D3% : IF D%<D5% then D%=D5% : K3%=0

rem New algorithm for Chinese downward-reading connective
rem
Def FNAction%(x%)
u%=x% : u1%=0 : uflag%=0 : op2%=0
DO WHILE u%>1
u%=fnup%(u%,1)
u1%=fnpos%(1,"2u",mid$(a$(2},u%,1})
if u1%=1 then op2%=u% : exit loop
if u1%=2 then exit loop
if fnpos%(1,"pqgfr",mid$(a$(2},u%,1})<1 then u1%=0 : exit loop
u%=U%-1
LOOP
if u1%<1 then fnaction%=0 : exit def
ON u1% aOTO nact10, nact20
nact10:
if mid$(a$(5},op2%,1)="1" AND len(r$(fnasc%(op2%}))<1 _ then fnaction%=0 : exit def
u1%=x%+1
u1%=fndwn%(u1%,d2%)
if u1%<1 OR _ fnpos%(1,"3",mid$(a$(2},u1%,1})<1 then fnaction%=0 : exit def

(continued)
p% = fnpos%(x%+1, left$(a$(1),u1%-1),"v")
if p%<1 then u1%=0 : goto nact30
if u1%=d2% then nact30
u1%=fndwn%(u1%+1,d2%)
if u1%>0 then u1%=u1%-1 else u1%=d2%
goto nact30
nact20:
   uflag%=1
   u1%=x%-1
   DO WHILE u1%>0
      u1%=fnup%(u1%,1)
      if fnpos%(1,"gf",mid$(a$(2),u1%,1))<1 then exit loop
      u1%=u1%-1
   LOOP
   u$=mid$(A$(2),u1%,1)
   u1%=x%+1
   DO WHILE u1%<D2%
      u1%=fndwn%(u1%,D2%)
      IF MIDS(A$(2),u1%,1)=U$ then exit loop
   LOOP
   nact30:
   if u1%<1 or u1%>d2% then fnaction%=0 else
   fnaction%=u1%
Appendix 7.3 New algorithm for the Chinese upward-reading connective

rem The Chinese upward-reading connective
rem This routine is added for 1-u-1 3rd entry
7700 if len(d$)>0 then D$=D$+" "
    d$=D$+R$(fnasc%(x3%+1))
    x4%=fnk%(x3%+2)
    r%=x4% : q%=x2%-1
    gosub 4000
    d$=d$+q$ : x4%=q%
    q%=x3%-1 : r%=x2%
    gosub 4000
    d$=d$+q$
    q%=x4%

-------------------------------------------------------------
Def FNAction%(x%)
  u%=x% : u1%=0 : uflag%=0 : op2%=0
  DO WHILE u%>1
    u%=fnup%(u%,1)
    u1%=fnpos%(1,"2u",mid$(a$(2),u%,1))
        if u1%=1 then op2%=u% : exit loop
        if u1%=2 then exit loop
        if fnpos%(1,"pqgfr",mid$(a$(2),u%,1))<1 then u1%=0 : exit loop
    u%=u%-1
  LOOP
  if u1%<1 then fnaction%=0 : exit def
  ON ul% GOTO nact10, nact20
nact10:
  if mid$(a$(5),op2%,1)="1" AND len(r$(fnasc%(op2%)))<1 _
    then fnaction%=0 : exit def
  u1%=x%+1
  u1%=fndwn%(u1%,d2%)
  if u1%<1 OR _
      fnpos%(1,"3",mid$(a$(2),u1%,1))<1 then fnaction%=0 : exit def
  p%=fnpos%(x%+1,left$(a$(1),u1%-1),"v")
  if p%<1 then u1%=0 : goto nact30
  if u1%=d2% then nact30
  u1%=fndwn%(u1%+1,d2%)
  if u1%>0 then u1%=u1%-1 else u1%=d2%
  goto nact30
nact20:
  uflag%=1
  u1%=x%-1
  DO WHILE u1%>0
    u1%=fnup%(u1%,1)
    if fnpos%(1,"gf",mid$(a$(2),u1%,1))<1 then exit loop
    u1%=u1%-1
  LOOP
  u$=mid$(A$(2),u1%,1)
  u1%=x%+1

(continued)
DO WHILE u1%<=D2%
   u1%=fndwn%(u1%,D2%)
   IF MID$(A$(2),u1%,1)=U$ then exit loop
LOOP
nact30:
   if u1%<1 or u1%>d2% then fnaction%=0 else _
   fnaction%=u1%
Appendix 7.4 Modification to Chinese following differences

rem 4650 Q5$=R$(E%)+Q5$ 'Li del space for $41, $51
GOTO 4110

rem ..............................................................
rem ..............................................................
rem LEAD ROUTINE
rem Constructs lead focus with any differences
rem Reads from right to left - thus next counter
rem position gives start to Qualifier - lastly
rem it sets counter for Display
rem ..............................................................
rem 5000 V9%=9
Q%=1%
ON FNDIFF%(Q%) GOTO 5020,5005,5010
5005 Q5%=FNFOCUS%(Q%) rem Difference 4
L$=FNLEAD4$(Q%,Q5%) : Q%=Q5%-1 : D%=Q5% : RETURN
5010 Q5%=FNFOCUS%(Q%) rem Difference 5
' IF F5%>0 THEN return
L$=FNFOCUS5$(Q%,Q5%) : F5%=1 :_
D%=I%+1 : Q%=Q5%-1 : RETURN
'5015 L$=FN LEAD5$(Q%) : L$=FNCLEAN$(L$) : F5%=2 :_
'deleted for $51

rem Following differences
rem 5660 IF LEFT$(R$(E%),1)>Chr$(161) Then_ 'Li added for $41
D5$=D5$+R$(E%)_ 'used in Chinese
Else_
D5$=D5$+" +R$(E%)
GOTO 5640

10620 IF fnpos%(1,"st",MID$(REC$,I%+1,1))>0 THEN 11220
DiffCon$=MID$(REC$,I%+1,1) ' store difference code
A$(1)=A$(1)+DiffCon$ ' add to control string
IF DiffCon$<"v" THEN
A$(2)=A$(2)+RIGHT$(A$(2),1)
ELSE A$(2)=A$(2)+" " ' add focus operator (ex. for connectives)
rem
IF FNPOS%(1,"2350",DiffCon$)>0 THEN_ 'edited for $41
A$(3)=A$(3)+"1" "4" deleted
ELSE A$(3)=A$(3)+"0" ' lead 1 - non-lead 0
IF DiffCon$>"u" THEN_ A$(5)=A$(5)+"3"

(continued)
DEF FNLEAD4$(X%, Y%)
LOCAL X$
LOCAL Z%, Z1%
X$=""
IF LEFTS(R$(E%), 1)>CHRS(161) AND MIDS(A$(4), X%, 1)<"2" THEN_
    Z%=X% : GOTO NLEAD4s60 'Li for $41 Lead term without space
IF MIDS(A$(4), X%, 1)<"2" THEN Z%=X% : GOTO NLEAD4s50
FOR Z%=X% TO Y%+1 STEP -1
IF MIDS(A$(4), Z%, 1)="1" THEN EXIT FOR_
NEXT Z%
NLEAD4s50:
    Z1%=FNASC%(Y%)
    X$=R$(Z1%)
FOR Z%=Z% TO X%
    Z1%=FNASC%(Z%)
    X$=X$+R$(Z1%)
    NEXT Z%
FNLEAD4s60:
    Z1%=FNASC%(Y%)
    X$=R$(Z1%)
FOR Z%=Z% TO X%
    Z1%=FNASC%(Z%)
    X$=X$+R$(Z1%)
    NEXT Z%
FNLEAD4$=X$
END DEF
DEF FNFOCUS5$(X%, Y%)
LOCAL X$
LOCAL Z%
X$=""
FOR Y%=Y% TO X%
    Z%=FNASC%(Y%)
    X$=X$+R$(Z%)
    NEXT Y%
FNFOCUS5$=X$
END DEF
DEF FNLEAD5$(X%)
LOCAL X$
LOCAL Y%, Z%
    Z%=FNASC%(X%)
    Y%=FNPPOS%(1, R$(Z%), "")
    X$=RIGHTS(R$(Z%), LEN(R$(Z%))-Y%)
IF LEFTS(X$, 1)=" " THEN X$=RIGHTS(X$, LEN(X$)-1)
FNLEAD5$=X$
Appendix 7.5 Modification to capitalization

rem for capitalization
def fnalpha$(strg$)
local j$
j$=chr$(161) 'for identifying Chinese
if mid$(strg$,$1,1)$>j$ then goto na 'record
strg$=fncap$(strg$,1)
p$=fnpos%(1,strg$,$Pt$)
if p$>0 AND mid$(strg$,p$+2,1)$<j$ then_
strg$=fncap$(strg$,p$+2)
for j%=l to 5 step 2 'for no capital in Chinese
pun$=mid$(Pt$+Colon$+Hyphen$,$j%,2)
p1%=1 : p%=1
while p$>0
p$=fnpos%(p1%,strg$,$pun$)
if p$>0 AND mid$(strg$,p$+2,1)$<j$ then_
strg$=fncap$(strg$,p$+2)
p1%=p$+2
wend
next j%
p%=1
while p$>0
p$=fnpos%(1,strg$,"%")
if p$>0 then_
strg$=left$(strg$,p%-1)+right$(strg$,len(strg$)-p$)
wend
na:
fnalpha$=strg$
Appendix 7.6 Modification to Chinese punctuation

rem The following for Chinese punctuations
rem Col$ = colon, Hy$ = hyphen, Pt$ = full stop
rem SP$ = space, Comma$ = comma

Col$=chr$((161)+chr$(71) : Hy$=chr$((161)+chr$(88)
Pt$=chr$((161)+chr$(79) : SP$=chr$((161)+chr$(64)
Colon$=Col$+SP$ : Hyphen$=SP$+Hy$+SP$ : Pnt$=PT$+SP$
Comma$=chr$((161)+chr$(77)
Appendix 9.1 Input and output Chinese records

0002 三民主義的次序和結構
  String
  1 - *1)三民主義
  2 - *p)次序$w$和
  3 - *g)結構
    $z1103三民主義$zp103次序$w$和$zg103結構
      Entries
    三民主義
      次序和結構
    次序。 三民主義
    結構。 三民主義

0003 CD-ROM應用在CARIM TM電腦輔助資訊檢索和專同系統
  String
  1 - *1)資訊   LO
  2 - *2)資訊檢索$31電腦輔助$32CARIM TM$w$和
  3 - *g)專同系統
  4 - *s)應用$w$在
  5 - *3)CD-ROM
    $z1100資訊$z2103資訊檢索$31電腦輔助$32CARIM TM$w$和$zg103專同系統$zs003應用$w$在$z3103CD-ROM
      Entries
    資訊
      CARIM TM電腦輔助資訊檢索和專同系統。應用CD-ROM
    資訊檢索
      CARIM TM電腦輔助資訊檢索。應用CD-ROM
    電腦輔助資訊檢索
      CARIM TM電腦輔助資訊檢索。應用CD-ROM
    CARIM TM電腦輔助資訊檢索
      應用CD-ROM
    專同系統
      應用CD-ROM
    CD-ROM
      應用在CARIM TM電腦輔助資訊檢索和專同系統

0004 窗戶被小偷打破
  String
  1 - *1)窗戶
  2 - 2)打破$w$被$w$
  3 - *3)小偷
    $z1103窗戶$z2003打破$w$被$w$3103小偷
      Entries
    窗戶
      被小偷打破
    小偷
      打破窗戶
0005 南非被日本經濟制裁

String

1. *1)南非
2. *2)經濟制裁$y被Sw
3. *3)日本
   $z1103南非$z2103經濟制裁$y被Sw$z3101日本

Entries

南非
被日本經濟制裁
經濟制裁：南非
被日本
經濟制裁南非

0006 電子文件使用在圖書館

String

1. *1)圖書館
2. *s)使用$ySw在
3. *3)電子文件
   $z1103圖書館$z2103使用$ySw在$z3101電子文件

Entries

圖書館
使用電子文件
電子文件
使用在圖書館

0007 基督教內在教義對西方世界影響

String

1. *1)西方世界
2. 2)影響$y受Sw
3. *3)基督教內在教義NU2
4. *4)基督教
5. *p)內在教義
   $z1103西方世界$z2003影響$y受Sw$z3021基督教內在教義$z3103基督教$z3p103內在教義

Entries

西方世界
受基督教內在教義影響
基督教
內在教義，影響西方世界
影響西方世界

0008 護士對精神病患的態度

String

1. *1)精神病患
2. *3)護士的態度
3. *x2)護士
4. *p)對精神病患的態度
   $x1103精神病患$y3003護士的態度$x1103護士$yp003對精神病患的態度

Entries

精神病患
護士的態度
護士
對精神病患的態度
0009 美國和發展中國家經濟關係
String
1 - *1)美國
2 - *u)經濟關係$V和$W和
3 - *1)發展中國家
$z1103美國$zu103經濟關係$V和$W和$z1103發展中國家
Entries
美國
和發展中國家經濟關係
經濟關係。 美國
和發展中國家
經濟關係。 發展中國家
和美國
發展中國家
和美國經濟關係

0010 三民主義和各種主義比較
String
1 - *1)三民主義
2 - *u)比較$V和$W和
3 - *1)各種主義
$V1103三民主義$zu003比較$V和$W和$z1003各種主義
Entries
三民主義
和各種主義比較

0011 呂氏春秋和名家的關係
String
1 - *1)群經$31中國
2 - 1) ND1
3 - *p)呂氏春秋
4 - *u)關係$V和$W和
5 - *p)名家
$z1103群經$31中國$z1012$zp103呂氏春秋$zu003關係$V和$W和$zp103名家
Entries
群經
中國群經。 呂氏春秋和名家關係
中國群經
呂氏春秋和名家關係
呂氏春秋
名家
和呂氏春秋關係
0012 意義被資訊理論説明

String

1 - 1)意義
2 - 2)説明$V被$W
3 - 3)資訊理論
   $z1103意義$z2003説明$V被$W$z3103資訊理論
   Entries
   意義
   被資訊理論説明
   資訊理論
   說明意義

Entries

日本

建築物

日本

地震後

建築物

日本

地震

建築物

日本

地震後

建築物

日本

0013 日本地震後建築物重建

String

1 - 0)日本
2 - 1)建築物
3 - 2)地震$41後
4 - 2)重建
   $z0103日本$z1103建築物$z2103地震$41後$z2103重建
   Entries
   日本
   建築物
   地震後
   重建

Entries

日本

建築物

日本

地震後

建築物

日本

地震

建築物

日本

地震後

建築物

日本

0014 日本地震後建築物重建

String

1 - 0)日本
2 - 1)建築物
3 - 2)地震$51後
4 - 2)重建
   $z0103日本$z1103建築物$z2103地震$51後$z2103重建
   Entries
   日本
   建築物
   地震後
   重建

Entries

日本

建築物

日本

地震後

建築物

日本

地震

建築物

日本

地震後

建築物

日本

重建

地震後

建築物

日本
0015 日據時期（1895-1945）臺灣公共圖書館研究  
String  
1 - *0)臺灣  
2 - *1)公共圖書館$1895-1945  
3 - 2)研究  
$103臺灣$1103公共圖書館$1895-1945$2003研究  
Entries  
臺灣  
公共圖書館，1895-1945。研究  
公共圖書館，臺灣  
1895-1945。研究  

0016 都市地域計劃社會學家的任務  
String  
1 - x*1)都市地域  
2 - y*2)地域計劃  
3 - y 3)社會學家任務  
4 - x*1)社會學家  
5 - y p)關於都市計劃任務  
$x1103都市地域$y2103地域計劃$y3003社會學家任務$  
$x1103社會學家$yp003關於都市計劃任務  
Entries  
都市地域  
地域計劃，社會學家任務  
地域計劃，都市地域  
社會學家任務  
社會學家  
關於都市計劃任務  

0017 系統理論應用於國際關係研究  
String  
1 - *2)國際關係  
2 - 2)研究  
3 - 2)國際關係研究 ND2  
4 - s)應用$S$w在  
5 - *3)系統理論  
$z2103國際關係$z2003研究$z2022國際關係研究$z003應用$S$w在$  
z3103系統理論  
Entries  
國際關係  
研究，應用系統理論  
系統理論  
應用在國際關係研究
0018 建築物被霜損壞
String

1 - *1)建築物
2 - *2)損壞$V$被$W$
3 - *3)霜
$Z1103$建築物$Z2103$損壞$V$被$W$和$Z3103$霜
Entries
建築物
被霜損壞
損壞，建築物
被霜
霜
損壞建築物

0019 中國和美國外交關係
String

1 - *1)中國
2 - *u)外交關係$V$和$W$和
3 - *1)美國
$Z1103$中國$Z2u103$外交關係$V$和$W$和$Z1103$美國
Entries
中國
和美國外交關係
外交關係，中國
和美國
外交關係，美國
和中國
美國
和中國外交關係

0020 兒童學業成績和身體發育的關係
String

1 - *1)兒童
2 - *2)學業成績
3 - *u)關係$V$和$W$和
4 - *2)身體發育
$Z1103$兒童$Z2103$學業成績$Zu003$關係$V$和$W$和$Z2103$身體發育
Entries
兒童
學業成績和身體發育關係
學業成績，兒童
和身體發育關係
身體發育，兒童
和學業成績關係
0021 電腦程序編寫於文件索引編製上的應用

**String**

1 - *1)文件
2 - *2)索引編製
3 - 2)文件索引編製 ND2
4 - s)應用$vs$w在
5 - 3)電腦程序編寫 NU2
6 - *3)電腦
7 - *2)程序編寫

$z1103$文件$z2103$索引編製$z2022$文件索引編製$zs003$應用$vs$w在$
$z3021$電腦程序編寫$z3103$電腦$zs2103$程序編寫

Entries

文件
索引編製，應用電腦程序編寫
索引編製，文件
應用電腦程序編寫
電腦
程序編寫，應用在文件索引編製
程序編寫，電腦
應用在文件索引編製

0022 精神科護理在職訓練的效果

**String**

1 - x*1)精神疾病
2 - y*2)護理
3 - y p)在職訓練的效果
4 - x*1)護士
5 - y*2)在職訓練
6 - y 3)對精神疾病的效果

$zx1103$精神疾病$y2103$護理$yp003$在職訓練的效果$zx1103$護士$
$y2103$在職訓練$y3003$對精神疾病的效果

Entries

精神疾病
護理，在職訓練的效果
護理 精神疾病
在職訓練的效果
護士
在職訓練，對精神疾病的效果
在職訓練，護士
對精神疾病的效果
0023　教士參與義大利的抵抗運動

String

1 - *0)義大利
2 - x*1)抵抗運動$d1940-1945
3 - y 3)教士的參與
4 - y*1)教士
5 - x p)參與抵抗運動$d1940-1945

Entries

義大利
抵抗運動，1940-1945。教士的參與
抵抗運動。義大利
1940-1945。教士的參與
義大利
教士。參與抵抗運動，1940-1945
教士。義大利
參與抵抗運動，1940-1945

0024　使用Nifedipine，Propranolol和Hydralazine治療人的高血壓

String

1 - *1)人
2 - *2)高血壓
3 - *2)治療
4 - 2)治療人的高血壓ND3
5 - s)使用$V$W在
6 - *3)Nifedipine
7 - *g)Propranolol$V$和
8 - *g)Hydralazine

Entries

人
高血壓。治療。使用Nifedipine，Propranolol和Hydralazine

治療：高血壓。人

使用Nifedipine，Propranolol和Hydralazine

Nifedipine

使用在治療人的高血壓
Propranolol

使用在治療人的高血壓
Hydralazine

使用在治療人的高血壓
0025 臺灣政府對稻米價格的政策

String

1 - *0)臺灣
2 - *1)稻米
3 - *p)價格
4 - *3)政府的政策
5 - *1)政府
6 - *p)對稻米價格的政策

$z0103臺灣$x1103稻米$y103價格$y3103政府的政策$x1103政府$yp003對稻米價格的政策

Entries

臺灣
稻米. 價格. 政府的政策
稻米. 臺灣
價格. 政府的政策
政府. 稻米. 臺灣
政府. 臺灣
對稻米價格的政策

0026 臺灣民間信仰: 傳, 釋和道

String

1 - *0)臺灣
2 - *2)信仰$x31民間
3 - *q)儒
4 - *g)釋$y和
5 - *g)道

$z0103臺灣$z2103信仰$x31民間$zq103儒$yg103釋$yg103到

Entries

臺灣
民間信仰: 儒, 釋和道
信仰. 臺灣
民間信仰: 儒, 釋和道
民間信仰. 臺灣
儒, 釋和道
儒. 民間信仰. 臺灣
釋. 民間信仰. 臺灣
道. 民間信仰. 臺灣
0027 臺灣南部地區局部環流之研究

String

1 - *0)臺灣
2 - *p)南部地區
3 - *p)環流$31局部
4 - 2)研究
   $z0103臺灣$zp103南部地區$zp103環流$31局部$z2003研究

Entries

臺灣
   南部地區。局部環流。研究
南部地區。 臺灣
   局部環流。研究
環流。 南部地區。 臺灣
   局部環流。研究
局部環流。 南部地區。 臺灣
   研究

0028 島群

String

1 - *1)島
2 - *g)群
   $z1103島$zr103群

Entries

島
群。 島

0029 法國集會遊行法及人民團體組織法

String

1 - *1)法國
2 - *p)集會遊行法$g和
3 - *g)人民團體組織法
   $z1103法國$zp103集會遊行法$g和$zg103人民團體組織法

Entries

法國
集會遊行法和人民團體組織法
集會遊行法。 法國
人民團體組織法。 法國
0030 墾丁國家公園的史前文化

Entries

1 - *口)臺灣

2 - *p)墾丁

3 - *1)國家公園

4 - 1) ND3

5 - *g)墾丁國家公園

6 - *2)史前文化

$z0103臺灣$zp103墾丁$z1103國家公園$z1032$zg103墾丁國家公園$z2103史前文化

Entries

臺灣

墾丁．國家公園：墾丁國家公園．史前文化

墾丁． 墾丁

國家公園： 墾丁國家公園．史前文化

國家公園： 墾丁

墾丁國家公園： 史前文化

史前文化： 墾丁國家公園

0031 中國整建閩、浙兩省交通設施調查

Entries

1 - *口)中國

2 - *p)福建$v和

3 - *g)浙江

4 - 1)交通設施$d1986

5 - 2)整建

6 - 6)調查

$z0103中國$zp103福建$zg103浙江$z1103交通設施$z1986$z2003整建$z6003調查

Entries

中國

福建和浙江．交通設施．1986．整建 — 調查

福建． 中國

交通設施．1986．整建 — 調查

浙江． 中國

交通設施．1986．整建 — 調查

交通設施．福建和浙江．中國

1986．整建 — 調查
0032 世界經濟的回顧與展望
String
1 - *0)世界
2 - *2)經濟
3 - 2)回顧
4 - g)展望
Entries
$z0103世界$z2103經濟$z2003回顧$y和$zg003展望

0033 日本經濟的回顧和展望
String
1 - *0)日本
2 - *2)經濟
3 - 2)回顧$y和
4 - g)展望
Entries
$z0103日本$z2103經濟$z2003回顧$y和$zg003展望

0034 美國削減臺灣優惠關稅待遇
String
1 - *1)臺灣
2 - *p)優惠關稅待遇
3 - 1)臺灣優惠關稅待遇 ND2
4 - 2)削減$y被$y
5 - *3)美國
Entries
$z1103臺灣$z2p103優惠關稅待遇$z1022臺灣優惠關稅待遇$z2003削減$y被$y$z3103美國

臺灣
削減臺灣優惠關稅待遇
被美國削減
臺灣
削減
美國
削減臺灣優惠關稅待遇
0035 打破林園的空間佈局與遊園韻律

String

1 - *0) 臺北
2 - *p) 臺北
3 - *p) 打破
4 - *p) 林家花園
5 - *2) 佈局 $31 空間 $v $w 畫
6 - *g) 許律 $31 遊園

$20103 臺北 $zp103 臺北 $zp103 打破 $zp103 林家花園 $z2103 佈局 $31 空間 $v $s $g103 許律 $31 遊園

Entries

臺北。打破。林家花園。空間佈局和遊園韻律
臺北。 臺北
打破。林家花園。空間佈局和遊園韻律
打破。 臺北。 臺灣
林家花園。空間佈局和遊園韻律
林家花園。打破。 臺北。 臺灣
空間佈局和遊園韻律
佈局。 林家花園。打破。 臺北。 臺灣
空間佈局
空間佈局。 林家花園。打破。 臺北。 臺灣
遊園韻律。 林家花園。打破。 臺北。 臺灣
遊園韻律

0036 冷卻對銅－鋁－鋁形狀記憶體合金的影響

String

1 - *1) 合金 $31 形狀記憶體 $32 鋁 $33 鋁－$34 話
2 - 2) 影響 $v 受 $w
3 - *3) 冷卻

$z1103 合金 $31 形狀記憶體 $32 鋁 $33 鋁－$34 鋁－$z2003 影響 $v 受 $w

Entries

合金
銅－鋁－鋁形狀記憶體合金。受冷卻影響
形狀記憶體合金
銅－鋁－鋁形狀記憶體合金。受冷卻影響
鋁形狀記憶體合金
銅－鋁－鋁形狀記憶體合金。受冷卻影響
鋁－鋁形狀記憶體合金
銅－鋁－鋁形狀記憶體合金
受冷卻影響
冷卻
影響銅－鋁－鋁形狀記憶體合金
0037 歐洲的教育實習制度
String
1 - *0)歐洲
2 - *2)教育
3 - *p)實習制度
$z0103歐洲$z2103教育$zp103實習制度
Entries
教育。實習制度
教育。歐洲
實習制度。教育。歐洲

0038 CD-ROM使用在圖書館
String
1 - *1)圖書館
2 - s)使用$v$w在
3 - *3)CD-ROM
$z1103圖書館$zs003使用$v$w在$z3103CD-ROM
Entries
圖書館
使用CD-ROM
CD-ROM
使用在圖書館

0039 臺北市清潔隊員胸腔疾病調查研究
String
1 - *0)臺灣
2 - *p)臺北
3 - *1)清潔隊員
4 - *p)胸腔
5 - *2)疾病
6 - 6)調查研究
$z0103臺灣$zp103臺北$z1103清潔隊員$zp103胸腔$z2103疾病$z6003調查研究
Entries
臺灣
臺北。清潔隊員。胸腔。疾病 — 調查研究
臺北。臺灣
清潔隊員。胸腔。疾病 — 調查研究
清潔隊員。臺北。臺灣
胸腔。疾病 — 調查研究
胸腔。清潔隊員。臺北。臺灣
疾病 — 調查研究
疾病。胸腔。清潔隊員。臺北。臺灣
— 調查研究
未來農業的機械化和自動化

String

1 - *1) 農業
2 - *2) 機械化
3 - *3) 自動化

$21103 農業
$31 未來
$2103 機械化
$2g103 自動化

Entries

未來農業
機械化和自動化
未來農業
機械化
自動化

未來農業
未來農業
機械化
自動化

未來農業
機械化
自動化

未來農業
機械化
自動化

未來農業
機械化
自動化

未來農業
機械化
自動化
Appendix 9.2 File records

"00470040"

"0710002”三民主義的次序和結構“$z1103三民主義zsp103次序sv和$zg103結構$x001"

"1510003”CD-ROM應用在CARIM TM電腦輔助診斷系統“$z11030資訊$z2103診斷系統$z103電腦輔助zsp103電腦系統$zs"

"003應用sv$sw在$z3103CD-ROM%0002"

"0630004”窗戶被小偷打破“$z1103窗戶$z2003打破sv被$sw$z3103小偷%0003"

"0710005”為被日本經貿制度“$z1103為$z2103經貿制度sv被$sw$z3103日本%0004"

"0750006”電子信件使用在圖書館“$z1103圖書館$zsw003使用sv$sw在$z3103電子信件%0005"

"1170007”基督敎在本教義的見解和西方世界影響“$z1103西方世界$z2003影響sv受$sw$z3021基督敎在本教義$z3103基督敎zsp103內在本教義%0006"

"0950008”護士對精神病患的態度“$x1103精神病患$y3003護士的態度sv$z1103護士$yp003對精神病患的態度$x007"

"0850009”美國和發展中國家經貿關係“$z1103美國$z1003經貿關係sv和$sw$z1103發展中國家%0008"

"0810010”三民主義和各種主義比較“$z1103三民主義$zu003比較sv和$sw$z1003各種主義%0009"

"0980011”呂氏春秋和各家的關係“$z1103詳釋$z1003中國$z1012$zsp103呂氏春秋$zu003關係sv和$sw$z1003各家%010"

"0710012”意義被資訊理論說明“$z1103意義$z2003說明sv被$sw$z3103資訊理論%011"

"0500013”日本地震後建築物重建“$z1103日本$z1103建築物$z2103地震%11後$z2103重建%012"

"0800014”日本地震後建築物重建“$z1103日本$z1103建築物$z2103地震%11後$z2103重建%012"

"0970015”日據時期(1895-1945)臺灣公共圖書館研究“$z1103臺灣$z1103公共圖書館$d1895-1945$z2003研究%014"

"1210016”都市計畫社會學家的任務“$z1103都市$z3003城市$sp003社會學家$yp003關於都市計劃任務%1015"

"1110017”系統理論應用於國際關係研究“$z2103國際關係$z2003研究$z2002國際關係研究$sz003應用sv$sw在$z3103系統理論%016"

"0630018”建築物被焚斃損壞“$z1103建築物$z2103損壞sv被$sw$z3103損%017"

"0730019”中國和美國外交關係“$z1103中國$zu003外交關係sv和$sw$z1103美國%018"

"0970020”兒童學術成就和身體發育的關係“$z1103兒童$z2103學術成就$zu003關係sv和$sw$z1103身體發育%019"

"1140021”電腦程序編寫於文件索引編製上的應用“$z1103文件$z2103索引編製$z2003應用sv$sw在$z3021電腦程序編寫%3023"

"1270022”精神科護理在職訓練的成效“$x1103精神病患$y2103護理$yp003在職訓練的成效$x1103護士$y2103在職訓練%3003對精神病患的成效%2021"

"1290023”教師參與義大利的反抗運動“$z1003義大利$z1103反抗運動$zd1940-1945$y3003教師的參與$x1103教師$yp003參與反抗運動$zd1940-1%

"9450022"

"1870024”使用Nifedipine, Propranolol和Hydralazine治療人的高血壓“$z1103人$z2103高血壓$z2103治療$z2003治療人的高血壓$zs003使用%"N$sv$sw在$z3103Nifedipine$zg103Propranolol$sv和$zg103Hydralazine%0023"
1150025**臺灣政府對稻米價格的政策**$z0103臺灣$x1103稻米$yp103 價格$y3103政府的政策$£1103政府$yp003對稻米價格的政策$x0024" 
"0920026**儒，釋，道：臺灣民間信仰**$z0103臺灣$x2103信仰$x31民間 
$zq103儒$ze103釋$yv和$zg103道$x0025 
"0900027**臺灣南部地區局部景流之研究**$z0103臺灣$zp103南部地區$ 
zp103還流$x31局部$x2003研究$x0026 
"0330028**及群**$z1103及$ze103群$x0027 
"0930029**法國集會遊行法及人民團體組織法**$z1103法國$zp103集會 
遊行法$yv和$zg103民間團體組織法$x0028 
"1070030**藝術國家公園的史前文化**$z0103臺灣$zp103藝術$z1103國 
家公園$ze103被$ze103藝術國家公園$x2103史前文化$x0029 " 
"1170031**中國內建館，浙兩省交通設施調查**$z0103中國$zp103福建$yv和$zg103浙江$x1103交通設施$x419868$z2003內建$x6003調查$x0030 
"0770032**世界經濟的回顧與展望**$z0103世界$x2103經濟$x2003回顧$yv和$zg003展望$x0031 
"0770033**日本經濟的回顧與展望**$z0103日本$x2103經濟$x2003回顧$yv和$zg003展望$x0032 
"1130034**美國削減臺灣優惠關稅待遇**$z1103臺灣$zp103優惠關稅待 
遇$x11022臺灣優惠關稅待遇$x2003削減$yv被$swx3103美國$x0033 " 
"1230035**政府對國的空間部局與遊園誌律**$z0103臺灣$zp103臺北$zp 
103政府$x2103林家花園$x2103佈局$x31空間$yv和$zg103遊誌$x31遊園$x003 
4" 
"1170036**冷凝對銅—鋁—錳金屬體合金的影響**$z1103合金$z31形 
狀體合金$x32銅$x33錳—$y34銅—$z2003影響$yv受$swx3103冷凝$x0035 " 
"0650037**歐洲的教育實習制度**$z0103歐洲$x2103教育$x2003實習制 
度$x0036 " 
"0710038**CD-ROM使用在圖書館**$z1103圖書館$x2103使用$xv$xw在$x2003 
CD-ROM$x0037 " 
"1110039**臺北市清潔隊員胸腔疾病調查研究**$z0103臺灣$zp103臺北$ 
z1103清潔隊員$x2103胸腔$x2103疾病$x6003調查研究$x0038 " 
"0820040**未來農業的機械化和自動化**$z1103農業$x31未來$x2103機械 
化$yv和$zg103自動化$x0039 " 

"1150025**臺灣政府對稻米價格的政策**$z0103臺灣$x1103稻米$yp103 價格$y3103政府的政策$£1103政府$yp003對稻米價格的政策$x0024" 
"0920026**儒，釋，道：臺灣民間信仰**$z0103臺灣$x2103信仰$x31民間 
$zq103儒$ze103釋$yv和$zg103道$x0025 
"0900027**臺灣南部地區局部景流之研究**$z0103臺灣$zp103南部地區$ 
zp103還流$x31局部$x2003研究$x0026 
"0330028**及群**$z1103及$ze103群$x0027 
"0930029**法國集會遊行法及人民團體組織法**$z1103法國$zp103集會 
遊行法$yv和$zg103民間團體組織法$x0028 
"1070030**藝術國家公園的史前文化**$z0103臺灣$zp103藝術$z1103國 
家公園$ze103被$ze103藝術國家公園$x2103史前文化$x0029 " 
"1170031**中國內建館，浙兩省交通設施調查**$z0103中國$zp103福建$yv和$zg103浙江$x1103交通設施$x419868$z2003內建$x6003調查$x0030 
"0770032**世界經濟的回顧與展望**$z0103世界$x2103經濟$x2003回顧$yv和$zg003展望$x0031 
"0770033**日本經濟的回顧與展望**$z0103日本$x2103經濟$x2003回顧$yv和$zg003展望$x0032 
"1130034**美國削減臺灣優惠關稅待遇**$z1103臺灣$zp103優惠關稅待 
遇$x11022臺灣優惠關稅待遇$x2003削減$yv被$swx3103美國$x0033 " 
"1230035**政府對國的空間部局與遊園誌律**$z0103臺灣$zp103臺北$zp 
103政府$x2103林家花園$x2103佈局$x31空間$yv和$zg103遊誌$x31遊園$x003 
4" 
"1170036**冷凝對銅—鋁—錳金屬體合金的影響**$z1103合金$z31形 
狀體合金$x32銅$x33錳—$y34銅—$z2003影響$yv受$swx3103冷凝$x0035 " 
"0650037**歐洲的教育實習制度**$z0103歐洲$x2103教育$x2003實習制 
度$x0036 " 
"0710038**CD-ROM使用在圖書館**$z1103圖書館$x2103使用$xv$xw在$x2003 
CD-ROM$x0037 " 
"1110039**臺北市清潔隊員胸腔疾病調查研究**$z0103臺灣$zp103臺北$ 
z1103清潔隊員$x2103胸腔$x2103疾病$x6003調查研究$x0038 " 
"0820040**未來農業的機械化和自動化**$z1103農業$x31未來$x2103機械 
化$yv和$zg103自動化$x0039"
Appendix 9.3 Production of Chinese entries

"0000000143
"026民主主義次序和結構＋0001
"025次序。 民主主義＋0001
"024結構。 民主主義＋0001
"058資訊米CARIM TM電腦輔助資訊檢索和導向系統。應用CD-ROM＋0002
"052資訊檢索米CARIM TM電腦輔助資訊檢索。應用CD-ROM＋0002
"060電腦輔助資訊檢索米CARIM TM電腦輔助資訊檢索。應用CD-ROM＋0002
"042CARIM TM電腦輔助資訊檢索米應用CD-ROM＋0002
"026導向系統米應用CD-ROM＋0002
"054CD-ROM米應用在CARIM TM電腦輔助資訊檢索和導向系統＋0002
"022窗戶米被小偷打破＋0003
"020小偷米 打破窗戶＋0003
"026南非米被日本經濟制裁＋0004
"030經濟制裁。 南非米被日本＋0004
"024日本米經濟制裁南非＋0004
"026圖書館米使用電子信件＋0005
"028電子信件米使用在圖書館＋0005
"036西方世界米受基督教內在教義影響＋0006
"036基督教米內在教義。影響西方世界＋0006
"038內在教義。 基督教米影響西方世界＋0006
"026精神病患米護士的態度＋0007
"028護士米對精神病患的態度＋0007
"032美國米和發展中國家經濟關係＋0008
"036經濟關係。 美國米和發展中國家＋0008
"036經濟關係。 發展中國家米和美國＋0008
"032發展中國家米和美國經濟關係＋0008
"030三民主義米和各種主義比較＋0009
| 040群經米中國群經 | 呂氏春秋和名家關係 + 0010 |
| 034中國群經米呂氏春秋和名家關係 | 呂氏春秋和名家關係 + 0010 |
| 026呂氏春秋米和名家關係 | 呂氏春秋和名家關係 + 0010 |
| 026名家米和呂氏春秋關係 | 呂氏春秋和名家關係 + 0010 |
| 026意義米被資訊理論說明 | 資訊理論米說明意義 + 0011 |
| 024資訊理論米說明意義 | 資訊理論米說明意義 + 0011 |
| 032日本米建築物 | 日本米建築物. 地震後. 重建 + 0012 |
| 034建築物 | 日本米建築物. 地震後. 重建 + 0012 |
| 040地震 | 建築物. 日本米建築物. 地震後. 重建 + 0012 |
| 036重建 | 地震後. 建築物. 日本米 + 0012 |
| 032日本米建築物 | 地震後. 重建 + 0013 |
| 034建築物 | 日本米建築物. 地震後. 重建 + 0013 |
| 040地震 | 建築物. 日本米建築物. 地震後. 重建 + 0013 |
| 034地震後 | 建築物. 日本米建築物. 地震後. 重建 + 0013 |
| 036重建 | 地震後. 建築物. 日本米 + 0013 |
| 039臺灣米公共圖書館 | 臺灣米公共圖書館, 1895-1945. 研究 + 0014 |
| 043公共圖書館 | 臺灣米, 1895-1945. 研究 + 0014 |
| 038都市地域米地域計 | 都市地域米地域計. 社會學家任務 + 0015 |
| 040地域計 | 都市地域米社會學家任務 + 0015 |
| 032社會學家米關於都市計 | 都市計. 社會學家任務 + 0015 |
| 034國際關係米研究 | 國際關係米研究. 社會學家任務 + 0016 |
| 034社會學家米國際關係研究 | 國際關係米研究. 社會學家任務 + 0016 |
| 022建築物米被破損壞 | 建築物米被破損壞 + 0017 |
| 026損壞 | 建築物米被破+ 0017 |
| 020破米損壞建築物 | 建築物米被破+ 0017 |
| 026中國米和美國外交關係 | 中國米和美國外交關係 + 0018 |
| 030外交關係 | 中國米和美國外交關係 + 0018 |
| 030外交關係 | 美國米和中國外交關係 + 0018 |
| 026中國米和中國外交關係 | 美國米和中國外交關係 + 0018 |
034 兒童米學業成績和身體發育關係 + 0019
038 學業成績。 兒童米和身體發育關係 + 0019
037 身體發育。 兒童米和學業成績關係 + 0019
038 文件索引編製，應用電腦程序編寫 + 0020
040 索引編製。 文件米應用電腦程序編寫 + 0020
040 電腦米程序編寫。 應用在文件索引編製 + 0020
042 程序編寫。 電腦米應用在文件索引編製 + 0020
036 精神病患米護理。 在職訓練的效果 + 0021
038 護理。 精神病患米在職訓練的效果 + 0021
038 護士米在職訓練。 對精神病患的效果 + 0021
040 在職訓練。 護士米對精神病患的效果 + 0021
045 義大利米抵抗運動，1940-1945。 護士的參與 + 0022
049 抵抗運動。 義大利米，1940-1945。 護士的參與 + 0022
043 義大利米護士。 參與抵抗運動，1940-1945 + 0022
045 護士。 義大利米參與抵抗運動，1940-1945 + 0022
064 人米血壓，治療。 使用 Nifedipine，Propranolol 和 Hydralazine + 0023
066 血壓。 人米治療。 使用 Nifedipine，Propranolol 和 Hydralazine + 0023
066 治療。 高血壓。 人米使用 Nifedipine，Propranolol 和 Hydralazine + 0023
038 Nifedipine 米使用在治療人的高血壓 + 0023
039 Propranolol 米使用在治療人的高血壓 + 0023
039 Hydralazine 米使用在治療人的高血壓 + 0023
034 臺灣米米。 價格。 政府的政策 + 0024
036 米。 臺灣米價格。 政府的政策 + 0024
036 價格。 米。 臺灣米政府的政策 + 0024
038 政府的政策。 價格。 米。 臺灣米 + 0024
034 臺灣米政府。 對米價格的政策 + 0024
036 政府。 臺灣米對米價格的政策 + 0024
034 臺灣米民間信仰： 像，釋和道 + 0025
042 信仰。 臺灣米民間信仰： 像，释和道 + 0025
034民間信仰。台灣民間信仰，釋道儒 + 0025
028儒。民間信仰。台灣儒 + 0025
028釋。民間信仰。台灣皈 + 0025
028道。民間信仰。台灣道 + 0025
036台灣米南部地區。部落景流。研究 + 0026
038南部地區。台灣米局部流。研究 + 0026
044景流。南部地區。台灣米局部流。研究 + 0026
038局部流。南部地區。台灣米研究 + 0026
012鳥米群 + 0027
016群。鳥米 + 0027
038法國米集會遊行法和人民團體組織法 + 0028
026集會遊行法。法國米 + 0028
030人民團體組織法。法國米 + 0028
052台灣米墾丁。國家公園：墾丁國家公園。史前文化 + 0029
054墾丁。台灣米國家公園：墾丁國家公園。史前文化 + 0029
052國家公園。墾丁。台灣米墾丁國家公園。史前文化 + 0029
028墾丁國家公園米史前文化 + 0029
032史前文化。墾丁國家公園米 + 0029
054中國米福建和浙江。交通設施，1986。整建 + 調查 + 0030
050福建。中國米交通設施，1986。整建 + 調查 + 0030
050浙江。中國米交通設施，1986。整建 + 調查 + 0030
058交通設施。福建和浙江。中國米，1986。整建 + 調查 + 0030
028世界米經濟。回顧和展望 + 0031
030經濟。世界米回顧和展望 + 0031
028日本米經濟。回顧和展望 + 0032
030經濟。日本米回顧和展望 + 0032
036台灣米優惠關稅待遇。被美國削減 + 0033
038優惠關稅待遇。台灣米被美國削減 + 0033
032美國米削減台灣優惠關稅待遇 + 0033
"052臺灣米台，板橋，林家花園，空間佈局和遊園韻律＋0034
"054台， 臺灣米板橋，林家花園，空間佈局和遊園韻律＋0034
"054板橋， 臺北， 臺灣米林家花園，空間佈局和遊園韻律＋0034
"054林家花園， 板橋， 臺北， 臺灣米空間佈局和遊園韻律＋0034
"050佈局， 林家花園，板橋， 臺北， 臺灣米空間佈局＋0034
"046空間佈局， 林家花園，板橋， 臺北， 臺灣米＋0034
"050韻律， 林家花園，板橋， 臺北， 臺灣米遊園韻律＋0034
"046遊園韻律， 林家花園，板橋， 臺北， 臺灣米＋0034
"048合金米銅－鋅－鋁形狀記憶體合金。受冷飛影響＋0035
"058形狀記憶體合金米銅－鋅－鋁形狀記憶體合金。受冷飛影響＋0035
"060鋁形狀記憶體合金米銅－鋅－鋁形狀記憶體合金。受冷飛影響＋0035
"064鋁－鋁形狀記憶體合金米銅－鋅－鋁形狀記憶體合金。受冷飛影響＋0035
"042銅－鋅－鋁形狀記憶體合金米受冷飛影響＋0035
"040冷飛米影響銅－鋅－鋁形狀記憶體合金＋0035
"026歐洲米教育，實習制度＋0036
"028教育， 歐洲米實習制度＋0036
"030實習制度， 教育， 歐洲米＋0036
"024圖書館米使用CD-ROM＋0037
"026CD-ROM米使用在圖書館＋0037
"052臺灣米台， 清潔隊員，胸腔，疾病－調查研究＋0038
"054台， 臺灣米清潔隊員，胸腔，疾病－調查研究＋0038
"054清潔隊員， 臺北， 臺灣米胸腔，疾病－調查研究＋0038
"054胸腔， 清潔隊員， 臺北， 臺灣米疾病－調查研究＋0038
"052疾病， 胸腔， 清潔隊員， 臺北， 臺灣米－調查研究＋0038
"036農業米未來農業，機械化和自動化＋0039
"030未來農業米機械化和自動化＋0039
"026機械化， 未來農業米＋0039
"026自動化， 未來農業米＋0039
"026機械化， 未來農業米＋0039
"026自動化， 未來農業米＋0039
Appendix 9.4 Production of Chinese sort keys

"000143

"三民主義，次序和結構，
"次序。 三民主義，
"結構。 三民主義，
"資訊，CARIMTM電腦輔助資訊检索和導向系統。應用CD-ROM，
"資訊檢索，CARIMTM電腦輔助資訊檢索。應用CD-ROM，
"電腦輔助資訊檢索，CARIMTM電腦輔助資訊檢索。應用CD-ROM，
"CARIMTM電腦輔助資訊檢索，應用CD-ROM，
"導向系統，應用CD-ROM，
"CD-ROM，應用在CARIMTM電腦輔助資訊檢索和導向系統，
"窗戶，被小偷打破，
"小偷，打破窗戶，
"南非，被日本經濟能制裁，
"經濟制裁。 南非，被日本，
"日本，經濟制裁南非，
"圖書館。使用電子信件，
"電子信件，使用在圖書館，
"西方世界，受基督教內在教義影響，
"基督教，內在教義。影響西方世界，
"內在教義。 基督教，影響西方世界，
"精神病患，護士的態度，
"護士，對精神病患的態度，
"美國，和發展中國家經濟能關係，
"經濟關係。 美國，和發展中國家，
"經濟關係。 發展中國家，和美國，
"發展中國家，和美國經濟能關係，
"三民主義，和各種主義比較，
"群經，中國群經。呂氏春秋名家關係，
"中國群經，呂氏春秋和名家關係",
"呂氏春秋，和名家關係",
"名家，和呂氏春秋關係",
"意義，被資訊理論說明",
"資訊理論，說明意義",
"日本，建築物。地震後。重建",
"建築物。 日本，地震後。重建",
"地震。 建築物。日本，地震後。重建",
"重建。 地震後。建築物。日本。",
"日本，建築物。地震後。重建",
"建築物。 日本，地震後。重建",
"地震。 建築物。日本，地震後。重建",
"地震後。 建築物。日本，重建",
"重建。 地震後。建築物。日本。",
"臺灣，公共圖書館，NUVR-NVQR。研究，"
"公共圖書館。 臺灣，NUVR-NVQR。研究，"
"都市地域，地域計劃。社會學家任務，"
"地域計劃。 都市地域，社會學家任務，"
"社會學家，關於都市計劃任務，"
"國際關係，研究。應用系統理論，"
"系統理論，應用在國際關係研究，"
"建築物，被霜損壞，"
"損壞。 建築物，被霜，"
"霜，損壞建築物，"
"中國，和美國外交關係，"
"外交關係。 中國，和美國，"
"外交關係。 美國，和中國，"
"美國，和中國外交關係，"
"兒童，學業成績和身體發育關係，"
學業成績。 兒童，和身體發育關係， 000058"
身體發育。 兒童，和學業成績關係， 000059"
文件，索引編製。 應用電腦程序編寫， 000060"
索引編製。 文件，應用電腦程序編寫， 000061"
電腦，程序編寫。 應用在文件索引編製， 000062"
程序編寫。 電腦，應用在文件索引編製， 000063"
精神疾病，護理。 在職訓練的效果， 000064"
護理。 精神疾病，在職訓練的效果， 000065"
護士，在職訓練。 對精神疾病的效果， 000066"
在職訓練。 護士，對精神疾病的效果， 000067"
義大利，抵抗運動，NVQM-NVQR。 護士的參與， 000068"
抵抗運動。 義大利，NVQM-NVQR。 護士的參與， 000069"
義大利，護士。 參與抵抗運動，NVQM-NVQR， 000070"
護士。 義大利，參與抵抗運動，NVQM-NVQR， 000071"
人，高血壓。 治療。 使用Nifedipine，Propranolol和Hydralazine， 000072"
高血壓。 人，治療。 使用Nifedipine，Propranolol和Hydralazine， 000073"
治療。 高血壓。 人，使用Nifedipine，Propranolol和Hydralazine， 000074"
Nifedipine，使用在治療人的高血壓， 000075"
Propranolol，使用在治療人的高血壓， 000076"
Hydralazine，使用在治療人的高血壓， 000077"
臺灣，稻米。 價格。 政府的政策， 000078"
稻米。 臺灣，價格。 政府的政策， 000079"
價格。 稻米。 臺灣，政府的政策， 000080"
政府的政策。 價格。 稻米。 臺灣， 000081"
臺灣，政府。 對稻米價格的政策， 000082"
政府。 臺灣，對稻米價格的政策， 000083"
臺灣，民間信仰。 僑，釋和道， 000084"
信仰。 臺灣，民間信仰。 僑，釋和道， 000085"
民間信仰。 臺灣，僑，釋和道， 000086"
"需。民間信仰。臺灣，"
"釋。民間信仰。臺灣，"
"道。民間信仰。臺灣，"
"臺灣，南部地區。局部環流。研究，"
"南部地區。 臺灣，局部環流。研究，"
"環流。南部地區。 臺灣，局部環流。研究，"
"局部環流。南部地區。 臺灣，研究，"
"鳥，群，"
"群。鳥，"
"法國，集會遊行法和人民團體組織法，"
"集會遊行法。 法國，"
"人民團體組織法。 法國，"
"臺灣，墾丁。國家公園： 墾丁國家公園，史前文化，"
"墾丁。 臺灣，國家公園： 墾丁國家公園，史前文化，"
"國家公園。 墾丁。 臺灣，墾丁國家公園，史前文化，"
"墾丁國家公園，史前文化，"
"史前文化。 墾丁國家公園，"
"中國，福建和浙江。交通設施，NVUS。整建 — 調查，"
"福建。 中國，交通設施，NVUS。整建 — 調查，"
"浙江。 中國，交通設施，NVUS。整建 — 調查，"
"交通設施。 福建和浙江。中國， NVUS。整建 — 調查，"
"世界，經濟。回顧和展望，"
"經濟。 世界，回顧和展望，"
"日本，經濟。回顧和展望，"
"經濟。 日本，回顧和展望，"
"臺灣，優惠關稅待遇。被美國削減，"
"優惠關稅待遇。 臺灣，被美國削減，"
"美國，削減臺灣優惠關稅待遇，"
“臺灣．臺北．板橋．林家花園．空間佈局和遊園設計，
000115”
“桃園． 臺灣．板橋．林家花園．空間佈局和遊園設計，
000116”
“板橋． 臺北．臺灣．林家花園．空間佈局和遊園設計，
000117”
“林家花園． 板橋． 臺北． 臺灣．空間佈局和遊園設計，
000118”
“佈局． 林家花園． 板橋． 臺北． 臺灣．空間佈局，
000119”
“空間佈局． 林家花園． 板橋． 臺北． 臺灣．
000120”
“遊園設計． 林家花園． 板橋． 臺北． 臺灣．
000121”
“合金．銅—鋅—鋁形狀記憶體合金．受冷浸影響，
000122”
“形狀記憶體合金．銅—鋅—鋁形狀記憶體合金．受冷浸影響，
000123”
“鋁形狀記憶體合金．銅—鋅—鋁形狀記憶體合金．受冷浸影響，
000124”
“銅—鋁形狀記憶體合金．銅—鋁—鋁形狀記憶體合金．受冷浸影響，
000125”
“銅—鋁—鋁形狀記憶體合金．受冷浸影響，
000126”
“冷浸，影響銅—鋁—鋁形狀記憶體合金，
000127”
“歐洲．教育．實習制度，
000128”
“教育． 臺灣．教育．實習制度，
000129”
“歐洲．教育．實習制度，
000130”
“教育． 臺灣．實習制度．教育．
000131”
“圖書館．使用CD-ROM，
000132”
“CD-ROM．使用在圖書館，
000133”
“臺灣． 臺北．清潔隊員．胸腔．疾病 — 調查研究，
000134”
“ 臺北． 臺灣．清潔隊員．胸腔．疾病 — 調查研究，
000135”
“清潔隊員． 臺北． 臺灣．胸腔．疾病 — 調查研究，
000136”
“胸腔． 清潔隊員． 臺北． 臺灣．疾病 — 調查研究，
000137”
“疾病． 胸腔．清潔隊員． 臺北． 臺灣． — 調查研究，
000138”
“農業．未來農業．機械化和自動化，
000139”
“未來農業．機械化和自動化，
000140”
“自動化． 未來農業，
000141”
“機械化． 未來農業，
000142”
“自動化． 未來農業，
000143”
“機械化． 未來農業，
000144”
### Appendix 9.5 Sorted Chinese records

<table>
<thead>
<tr>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;0008&quot;</td>
</tr>
<tr>
<td>&quot;0133&quot;</td>
</tr>
<tr>
<td>&quot;0010&quot;</td>
</tr>
<tr>
<td>&quot;0077&quot;</td>
</tr>
<tr>
<td>&quot;0075&quot;</td>
</tr>
<tr>
<td>&quot;0076&quot;</td>
</tr>
<tr>
<td>&quot;0072&quot;</td>
</tr>
<tr>
<td>&quot;0098&quot;</td>
</tr>
<tr>
<td>&quot;0002&quot;</td>
</tr>
<tr>
<td>&quot;0027&quot;</td>
</tr>
<tr>
<td>&quot;0012&quot;</td>
</tr>
<tr>
<td>&quot;0053&quot;</td>
</tr>
<tr>
<td>&quot;0104&quot;</td>
</tr>
<tr>
<td>&quot;0029&quot;</td>
</tr>
<tr>
<td>&quot;0020&quot;</td>
</tr>
<tr>
<td>&quot;0044&quot;</td>
</tr>
<tr>
<td>&quot;0060&quot;</td>
</tr>
<tr>
<td>&quot;0034&quot;</td>
</tr>
<tr>
<td>&quot;0038&quot;</td>
</tr>
<tr>
<td>&quot;0110&quot;</td>
</tr>
<tr>
<td>&quot;0015&quot;</td>
</tr>
<tr>
<td>&quot;0108&quot;</td>
</tr>
<tr>
<td>&quot;0103&quot;</td>
</tr>
<tr>
<td>&quot;0054&quot;</td>
</tr>
<tr>
<td>&quot;0055&quot;</td>
</tr>
<tr>
<td>&quot;0140&quot;</td>
</tr>
<tr>
<td>&quot;0086&quot;</td>
</tr>
<tr>
<td>&quot;0107&quot;</td>
</tr>
<tr>
<td>&quot;0031&quot;</td>
</tr>
<tr>
<td>&quot;0123&quot;</td>
</tr>
<tr>
<td>&quot;0046&quot;</td>
</tr>
<tr>
<td>&quot;0036&quot;</td>
</tr>
<tr>
<td>&quot;0040&quot;</td>
</tr>
<tr>
<td>&quot;0041&quot;</td>
</tr>
<tr>
<td>&quot;0067&quot;</td>
</tr>
<tr>
<td>&quot;0003&quot;</td>
</tr>
<tr>
<td>&quot;0142&quot;</td>
</tr>
<tr>
<td>&quot;0018&quot;</td>
</tr>
<tr>
<td>&quot;0119&quot;</td>
</tr>
<tr>
<td>&quot;0128&quot;</td>
</tr>
<tr>
<td>&quot;0030&quot;</td>
</tr>
<tr>
<td>&quot;0093&quot;</td>
</tr>
<tr>
<td>&quot;0124&quot;</td>
</tr>
<tr>
<td>&quot;0049&quot;</td>
</tr>
<tr>
<td>&quot;0059&quot;</td>
</tr>
<tr>
<td>&quot;0057&quot;</td>
</tr>
<tr>
<td>&quot;0069&quot;</td>
</tr>
<tr>
<td>&quot;0118&quot;</td>
</tr>
<tr>
<td>&quot;0117&quot;</td>
</tr>
<tr>
<td>&quot;0096&quot;</td>
</tr>
<tr>
<td>&quot;0074&quot;</td>
</tr>
<tr>
<td>&quot;0047&quot;</td>
</tr>
<tr>
<td>&quot;0120&quot;</td>
</tr>
<tr>
<td>&quot;0085&quot;</td>
</tr>
<tr>
<td>&quot;0013&quot;</td>
</tr>
<tr>
<td>&quot;0091&quot;</td>
</tr>
</tbody>
</table>
Appendix 9.6 Printout of Chinese entries

CARIM TM 聲幕輔助資訊检索
應用 CD-ROM == 0002
CD-ROM
使用在圖書館 == 0037
應用在 CARIM TM 聲幕輔助資訊检索和專同系統 == 0002
Hydralazine
使用在治療人的高血壓 == 0023
Nifedipine
使用在治療人的高血壓 == 0023
Prapranolol
使用在治療人的高血壓 == 0023

人民團體組織法 " 法國 == 0028

三民主義
次序和結構 == 0001
和各種主義比較 == 0009

小偷
打破窗戶 == 0003

中國
和美國外交關係 == 0018
福建和浙江，交通設施，1986。整建 - 調查 == 0030

台灣
呂氏春秋和名家關係 == 0010

內在政體。 基督教
影響西方世界 == 0006

公共圖書館。 臺灣
1895-1945。研究 == 0014

文件
索引編製。應用電腦程序編寫 == 0020

日本
建築物。 震後。重建 == 0012
建築物。震後。重建 == 0013
經濟。回顧和展望 == 0032
經濟制裁南非 == 0004

世界
經濟。回顧和展望 == 0031

史前文化。 遺址國家公園 == 0029

外交關係。 中國
和美國 == 0018
外交關係。 美國
和中國 == 0018

未來教育
擴展化和自動化 == 0039

民間信仰。 臺灣
農，神和道 == 0025

交通設施。 福建和浙江。中國
1986。整建 - 調查 == 0030

名家
和呂氏春秋關係 == 0010

合計
鋼 - 鎳 - 鋁 - 形狀記憶體合金。受冷環境影響 == 0035

地區計劃。 都市地域
社會學家任務 == 0015

地圖。 建築物。 日本
地圖。閩南。重建 == 0012
地圖。重建 == 0013
地震後，建築物。日本
重建 == 0013
在職訓練。護士
對精神病患的效果 == 0021
次序。三民主義” == 0001
自動化。未來農業 == 0039
西方世界
受基督教內在教義影響 == 0006
佈局。林家花園。板橋。臺北。臺灣
空間佈局 == 0034
冷鍍
影響銅－鋁－鋁形狀記憶體合金 == 0035
呂氏春秋
和名家關係 == 0010
局部環境。南部地區。臺灣
研究 == 0026
形狀記憶體合金
銅－鋁－鋁形狀記憶體合金。受冷鍍影響 == 0035
系統理論
應用在國際關係研究 == 0016
身體發育。兒童
和學業成績關係 == 0019
兒童
學業成績和身體發育關係 == 0019
抵抗運動。義大利
1940-1945。教士的參與 == 0022
林家花園。板橋。臺北。臺灣
空間佈局和遊園設施 == 0034
板橋。臺北。臺灣
林家花園。空間佈局和遊園設施 == 0034
法國
集會遊行法和人民團體組織法 == 0028
治療。高血壓。人
使用Nifedipine, Propranolol和Hydralazine == 0023
社會學家
關於都市計劃任務 == 0015
空間佈局。林家花園。板橋。臺北。臺灣 == 0034
信仰。臺灣
民間信仰：儒，釋和道 == 0025
南非
被日本統治狀態 == 0004
南部地區。臺灣
局部環境。研究 == 0026
建築物
被棄損壞 == 0017
建築物。日本
地震後。重建 == 0012
地震後。重建 == 0013
政府。臺灣
對稻米價格的政策 == 0024
政府的政策。價格。稻米。臺灣 == 0024
美國
和中國外交關係 == 0018
和發展中國家經濟關係 == 0008
削減臺灣經貿實例待遇 == 0033
重建。地震後。建築物。日本 == 0012
浙江。中國
交通設施，1986。整建 — 調查 == 0030
疾病：胸腔、清潔隊員、臺北、臺灣
- 調查研究 == 0038

索引編製：文件
應用電腦程序編寫 == 0020
胸腔、清潔隊員、臺北、臺灣
疾病 - 調查研究 == 0038

高血壓：人
治療、使用Nifedipine、Propranolol和Hydralazine == 0023
國家公園、墾丁、臺灣
墾丁國家公園、史前文化 == 0029

國庫關係
研究、應用系統理論 == 0016

基督徒
內在教義、影響西方世界 == 0006

教士、義大利
參與抵抗運動，1940-1945 == 0022
教育、歐洲
實用制度 == 0036

清潔隊員、臺北、臺灣

胸腔、疾病 - 調查研究 == 0038

都市區域
區域計劃、社會學家任務 == 0015

鳥
群 == 0027

發展中國家
和美國經濟關係 == 0008

程序編寫、電腦
應用在文件索引編製 == 0020

窗戶
被小偷打碎 == 0003

結構、三民主義 == 0001
集會遊行法、法國 == 0028

意義
被資訊理論說明 == 0011

損壞、建築物
被損 == 0017

經濟、日本
回顧和展望 == 0032

經濟、世界
回顧和展望 == 0031

經濟制度、南非
被日本、南非

經濟關係、美國
和發展中國家、發展中國家
和美國 == 0008

義大利
抵抗運動、1940-1945、義大利的參與 == 0022

教士、參與抵抗運動、1940-1945 == 0022
群、鳥 == 0027

詳情
中國詳情、國民党和國民關係 == 0010

資訊
CARIM TM電算輔助資郵檢索和語音系統，應用CD-ROM == 0002

資訊理論
說明意義 == 0011

資郵檢索
CARIM TM電算輔助資郵檢索，應用CD-ROM == 0002
農業
未來農業，機械化和自動化 == 0039
遊園設計，林家花園，按樣，台北，臺灣 == 0034
道，民間信仰，臺灣 == 0025
電子信件
使用在圖書館 == 0005
電腦
程序編寫，應用在文件索引編製 == 0020
電機輔助資訊檢索
CARIM TM電機輔助資訊檢索。應用CD-ROM == 0002
圖書館
使用CD-ROM == 0037
使用電子信件 == 0005
實習制度，教育，歐洲 == 0036
福建，中國
交通設施，1986，設施 == 調查 == 0030
精神病患
護士的態度 == 0007
護理，在職訓練的效果 == 0021
臺灣
按樣，林家花園，空間佈局和遊園講義 == 0034
清潔隊員，胸腔，疾病，調查研究 == 0038
臺灣
公共圖書館，1895-1945，研究 == 0014
民間信仰，原，釋和道 == 0025
南部地區，局部污染，研究 == 0026
政府，對玉米價格的政策 == 0024
臺北，按樣，林家花園，空間佈局和遊園講義 == 0034
臺北，清潔隊員，胸腔，疾病，調查研究 == 0038
稻米，價格，政府的政策 == 0024
墾丁，國家公園，墾丁國家公園，史前文化 == 0029
廟，神社和學校，美國削減 == 0033
鎚，鋁，鋁形狀記憶體合金
受冷水影響 == 0035
價格，稻米，臺灣，政府的政策 == 0024
歐洲
教育，實習制度 == 0036
稻米，臺灣，價格，政府的政策 == 0024
鎚，鋁，鋁形狀記憶體合金
鎚，鋁，鋁形狀記憶體合金，受冷水影響 == 0035
鋼，鋼，鋼形狀記憶體合金
鋼，鋼，鋼形狀記憶體合金，受冷水影響 == 0035
俄，民間信仰，臺灣 == 0025
墾丁，臺灣
國家公園，墾丁國家公園，史前文化 == 0029
墾丁國家公園，史前文化 == 0029
學業成就，學童
和身體發展關係 == 0019
導向系統
應用CD-ROM == 0002
機械化，未來農業 == 0039
優惠關稅待遇，臺灣，美國削減 == 0033
環流，南部地區，臺灣
部環流，研究 == 0026
覆
捐建建築物 == 0017
雕像、林家花園、建築。台北、臺灣
遊園雕像 == 0034
鏡、民間信仰。臺灣 == 0025
教士
在職訓練。對精神病患的效果 == 0021
對精神病患的態度 == 0007
經理。精神病患
在職訓練的效果 == 0021