THE EFFECT OF ANAPHORA ON THE COGNITIVE
PROCESSING AND COMPREHENSION OF READERS
OF GERMAN AT VARIOUS LEVELS OF Baseline
GERMAN LANGUAGE ABILITY

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Victoria Clark Berkemeyer, B.A., M.A.T.

* * * * *

The Ohio State University
1991

Dissertation Committee:

E. B. Bernhardt
G. A. Jarvis
J. Belkin

Approved by

[Signature]
Adviser
College of Education
To My Parents, Margaret E. and Egbert B. Clark

And to My Husband, Thomas G. Berkemeyer

With Love and Gratitude
ACKNOWLEDGEMENTS

It is not possible to pursue a successful academic career without the assistance, encouragement, and support of professors, friends, colleagues, and family members. There are many such people in my life to whom I owe much thanks and appreciation. In their unique ways, they have all enabled me to achieve my academic goals, and they continue to encourage me to reach my fullest potential.

First, I wish to express my most sincere gratitude and appreciation to my adviser, Dr. Elizabeth B. Bernhardt. Dr. Bernhardt has heightened my interest in the field of reading and, in particular, eye-tracking research. I greatly appreciate her selecting me to work in the eye-tracking lab, as that experience prepared me well to conduct this research project. Dr. Bernhardt tirelessly read all the drafts of this document, patiently answered all my tedious questions, and graciously accepted my tendency to be too meticulous. For all of her support, I am most grateful.

I am also indebted to Dr. Bernhardt for all of the personal attention and interest she has shown in my academic and professional career. She has taught me to have confidence in my abilities by encouraging me to publish and present my work and to apply for fellowship awards. Without her help and inspiration, I would have never received a Presidential Fellowship and a Graduate Student Alumni Research Award.
I am also very grateful to Dr. Jarvis, who served as my adviser during the earlier phases of my doctoral program. He not only taught me the fundamentals of research design and statistics but instilled in me an ever-growing appreciation for and interest in research endeavors. It has been a privilege to have been his student and to have learned from him.

I extend great thanks and appreciation to Dr. Johanna Belkin for her continuing interest in my academic pursuits and for serving on my general examination and dissertation committees. From her I have gained a much deeper appreciation and understanding of the German language—its origins, its influences, and its historical development. She has not only encouraged my research endeavors and my study of German, she has also enhanced my love of the language itself.

I also wish to acknowledge Cynthia P. McIver, who inspired my study of German in high school, Scott H. Abbott, who enhanced my understanding of German literature and persuaded me to study German at the graduate level, and Charles A. Honeycutt, who encouraged me to enter the field of foreign language education. They each played a pivotal role in my academic development and success. I am very grateful for their guidance and support.

My most heartfelt thanks is offered to my husband, Tom, whose constant support, patience, and love has guided me through every phase of this long and arduous process. During the most trying of times, Tom helped me to put both my achievements and defeats into proper perspective. His calming influence has always offered me comfort and reassurance when I needed it most.

I am also deeply grateful to my parents who had unwavering faith and confidence in my ability to succeed, who taught me always to do my best, and who instilled in me a great love for learning. I am also thankful that I was their
“souvenir of Germany” so many years ago. That initial exposure to a foreign language and culture left a lasting impression on me and inspired my pursuit of this degree. I also wish to thank my brothers and my sister, whose constant support and encouragement I greatly appreciate.

Special thanks also goes to the Berkemeyer family who always showed interest and concern in my academic progress and graciously tolerated my absence from family gatherings during the most time-consuming aspects of the program.

To my good friends, Elizabeth Graham and Patou Brozovic, I offer much thanks and appreciation. I could not have made it through all of the courses, teaching, and the general exam without the support and friendship of those two Chicas. I also want to thank Joe Lehmann, who generously gave of his time and effort in editing this manuscript and who has been a been an extremely helpful and supportive friend, and Kathrine Simon, whose thoughtfulness and concern has meant a lot to me over the past eight years.

My appreciation and thanks are extended to Craig Deville and Christine Cronjaeger for helping me to create the scoring instruments and especially to Pete Heinz who helped me score the recall protocols. Many thanks also go to statistical consultants, Laura Adkins and Fred Ruland, for their help and advice.

I am very grateful to the German Department for allowing me to recruit subjects through their courses and to my research subjects for their time and cooperation. Special appreciation and thanks go to Wynn Kimble and Mary Armentrout for generously taking phone messages from prospective subjects.

Finally, I offer my eternal thankfulness to Him who guides and protects us all and without whom this endeavor would have not been possible.
<table>
<thead>
<tr>
<th>Year(s)</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>April 19, 1961</td>
<td>Born - Frankfurt a/M, Germany</td>
</tr>
<tr>
<td>1983</td>
<td>B.A., Vanderbilt University</td>
</tr>
<tr>
<td></td>
<td>Nashville, Tennessee</td>
</tr>
<tr>
<td>1983-1984</td>
<td>Fulbright Grantee, Universität Stuttgart, Stuttgart, Germany</td>
</tr>
<tr>
<td>1985</td>
<td>Graduate Teaching Associate, Department of Germanic Languages,</td>
</tr>
<tr>
<td></td>
<td>University of North Carolina, Chapel Hill, North Carolina</td>
</tr>
<tr>
<td>1986</td>
<td>M.A.T., University of North Carolina, Chapel Hill, North Carolina</td>
</tr>
<tr>
<td>1986-1987</td>
<td>University Fellowship Awardee, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>1987-1988</td>
<td>Graduate Teaching Associate, Department of English as a Second</td>
</tr>
<tr>
<td></td>
<td>Language, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>1989</td>
<td>Graduate Research Associate, Department of Educational Studies, The</td>
</tr>
<tr>
<td></td>
<td>Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>1989-1990</td>
<td>Presidential Fellowship Awardee, The Ohio State University, Columbus, Ohio</td>
</tr>
</tbody>
</table>
1991 ..........................  Graduate Teaching Associate,
Department of English as a
Second Language, The Ohio
State University, Columbus, Ohio

PUBLICATIONS

Berkemeyer, V. C. (1989). Qualitative analysis of immediate recall protocol
data: Some classroom implications. Die Unterrichtspraxis, 22, 131-137

school German learner. Die Unterrichtspraxis, 21, 6-28.

expectations and behaviors of ESL teachers and students: Some cross-
cultural implications. The Ohio Foreign Language Journal: Special
Research Issue, 12-14.

FIELDS OF STUDY

Major Field: Education

Studies in Second Language Acquisition, Reading in a Second
Professor Elizabeth B. Bernhardt

Studies in Educational Psychology, Educational Statistics, Foreign
Professor Gilbert A. Jarvis

Minor Field: German Philology

Studies in German Linguistics, History of the German Language, Old High
German, and Gothic.
Professor Johanna Belkin
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xx</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>4</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>9</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>12</td>
</tr>
<tr>
<td>Phase I of the Study</td>
<td>13</td>
</tr>
<tr>
<td>Phase II of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Phase III of the Study</td>
<td>14</td>
</tr>
<tr>
<td>Theoretical Bases</td>
<td>14</td>
</tr>
<tr>
<td>A Definition of Anaphora</td>
<td>14</td>
</tr>
<tr>
<td>The Function of Anaphora in Discourse Contexts</td>
<td>27</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>30</td>
</tr>
<tr>
<td>Assumptions of the Study</td>
<td>31</td>
</tr>
<tr>
<td>II. REVIEW OF THE LITERATURE</td>
<td>33</td>
</tr>
<tr>
<td>Introduction</td>
<td>33</td>
</tr>
<tr>
<td>First Language Studies of Anaphora at the</td>
<td></td>
</tr>
<tr>
<td>Micro-Level of Discourse</td>
<td>35</td>
</tr>
<tr>
<td>First Language Studies of Anaphora at the</td>
<td></td>
</tr>
<tr>
<td>Macro-Level of Discourse</td>
<td>51</td>
</tr>
</tbody>
</table>
Eye-tracking Studies of Anaphora ........................................... 62
Second Language Studies of Syntax and Anaphora .................. 72
First Language Studies in the Acquisition of Anaphora in German. 82
Summary of the Literature .................................................. 91

III. DESIGN AND PROCEDURES ........................................... 95

Population and Sample ..................................................... 95
Research Design ............................................................. 96
  Phase I of the Study ..................................................... 96
  Phase II of the Study ................................................... 97
  Phase III of the Study .................................................. 99
Materials ................................................................. 99
  Baseline German Language Ability Test ................................ 99
  Texts ........................................................................ 100
    Phases I and II of the Study ........................................ 100
    Phase III of the Study ................................................. 103
Experimental Procedures .................................................. 104
Data Collection and Instrumentation ................................... 108
Data Analysis ............................................................... 112
  Phase I of the Study ..................................................... 112
  Phase II of the Study ................................................... 114
  Phase III of the Study .................................................. 115
Pilot Study ................................................................. 115

IV. ANALYSIS OF THE DATA ............................................. 117

Introduction ............................................................... 117
Phase I of the Study ..................................................... 118
  ANCOVA Results for Fixation Frequency ......................... 122
  ANCOVA Results for Average Fixation Duration ............. 126
  ANCOVA Results for Total Fixation Duration ............... 130
  ANCOVA Results for Reading Comprehension ................. 134
Phase II of the Study ................................................... 138
  ANCOVA Results for Average Fixation Duration in Blocks.. 139
    Expository Text: "Der Mechanisierte Hof" ................ 139
    Literary Text: "Im Volksgarten" ................................ 161
  ANCOVA Results for Percentage of Total Fixation Time in Blocks 187
    Expository Text: "Der Mechanisierte Hof" ............... 187
    Literary Text: "Im Volksgarten" ............................ 206
Phase III of the Study ...................................... 233
Correlational Analysis ...................................... 233
Qualitative Analysis ........................................ 235
  Personal Pronouns ....................................... 236
  Relative Pronouns ....................................... 249
  Indefinite Pronouns ..................................... 253
  Demonstrative Pronouns ................................ 255
  Demonstrative Noun Phrases .............................. 256
  Possessive Adjectives .................................. 258

V. DISCUSSION, SUMMARY, IMPLICATIONS,
   RECOMMENDATIONS, AND LIMITATIONS ............... 263

  Overview of the Study .................................. 263
  Discussion ................................................ 265
  Phase I of the Study .................................... 266
    Fixation Frequency ..................................... 266
    Average Fixation Duration ............................. 267
    Total Fixation Duration ............................... 269
    Reading Comprehension ................................ 271
  Phase II of the Study ................................... 272
    Average Fixation Duration in Blocks ................. 273
      Expository Text: “Der Mechanisierte Hof” ........ 273
      Literary Text: “Im Volksgarten” .................... 275
    Percentage of Total Fixation Time in Blocks ........ 277
      Expository Text: “Der Mechanisierte Hof” ........ 277
      Literary Text: “Im Volksgarten” .................... 278
  Phase III of the Study ................................... 280
  Correlational Analysis .................................. 280
  Qualitative Analysis .................................... 282

Summary of the Major Findings ......................... 287
  Phase I of the Study ................................... 288
  Phase II of the Study ................................... 290
  Phase III of the Study ................................... 295

  Implications for Pedagogy .............................. 296
  Recommendations for Further Research ................. 298
  Limitations of the Study ............................... 300

APPENDICES .................................................. 304
  A. Information Flier About the Study ................... 304
  B. Letter of Request to Distribute Information Fliers .. 306
C. Authentic and Experimental Versions of the Expository Text: “Der Mechanisierte Hof” ................................. 308
D. Authentic and Experimental Versions of the Literary Text: “Im Volksgarten” ............................................. 310
E. Expository Text: “Europa Heute” ........................................... 312
F. Coreferent Selection Task for the Expository Text: “Europa Heute” .................................................. 315
G. Student Information Form .............................................. 319
H. Screen Display of a Number Pattern ................................. 321
I. Practice Text ................................................................. 323
J. Recall Instructions and Response Form ............................. 325
K. Coreferent Selection Task: Explanation, Examples, and Practice Sentences ........................................... 327
L. Screen Display of the Block Matrix for the Expository Text: “Der Mechanisierte Hof” ................................. 329
   Authentic Text Version ................................................. 330
   Experimental Text Version ............................................ 331
M. Screen Display of the Block Matrix for the Literary Text: “Im Volksgarten” ............................................. 332
   Authentic Text Version ................................................. 333
   Experimental Text Version ............................................ 334
LIST OF REFERENCES .......................................................... 335
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Observed and predicted means and standard deviations of fixation frequency as a function of type of referencing device.</td>
<td>122</td>
</tr>
<tr>
<td>2. Analysis of covariance parameter estimates for fixation frequency.</td>
<td>125</td>
</tr>
<tr>
<td>3. Observed and predicted means and standard deviations of average fixation duration as a function of type of referencing device.</td>
<td>126</td>
</tr>
<tr>
<td>4. Analysis of covariance parameter estimates for average fixation duration</td>
<td>129</td>
</tr>
<tr>
<td>5. Observed and predicted means and standard deviations of total fixation duration as a function of type of referencing device.</td>
<td>130</td>
</tr>
<tr>
<td>6. Analysis of covariance parameter estimates for total fixation duration</td>
<td>133</td>
</tr>
<tr>
<td>7. Observed and predicted means and standard deviations of recall protocol scores as a function of type of referencing device.</td>
<td>134</td>
</tr>
<tr>
<td>8. Analysis of covariance parameter estimates for recall protocol scores.</td>
<td>137</td>
</tr>
<tr>
<td>9. Means and standard deviations of average fixation duration in block 1 of “Der mechanisierte Hof” as a function of type of referencing device.</td>
<td>141</td>
</tr>
<tr>
<td>10. Analysis of covariance of average fixation duration in block 1 of “Der mechanisierte Hof” by type of referencing device.</td>
<td>144</td>
</tr>
</tbody>
</table>
11. Means and standard deviations of average fixation duration in block 2 of "Der mechanisierte Hof" as a function of type of referencing device. .......................... 147

12. Analysis of covariance of average fixation duration in block 2 of "Der mechanisierte Hof" by type of referencing device. .................................................. 148

13. Means and standard deviations of average fixation duration in block 3 of "Der mechanisierte Hof" as a function of type of referencing device. ..................... 149

14. Analysis of covariance of average fixation duration in block 3 of "Der mechanisierte Hof" by type of referencing device. .................................................. 150

15. Means and standard deviations of average fixation duration in block 4 of "Der mechanisierte Hof" as a function of type of referencing device. ..................... 151

16. Analysis of covariance of average fixation duration in block 4 of "Der mechanisierte Hof" by type of referencing device. .................................................. 152

17. Means and standard deviations of average fixation duration in block 5 of "Der mechanisierte Hof" as a function of type of referencing device. ..................... 153

18. Analysis of covariance of average fixation duration in block 5 of "Der mechanisierte Hof" by type of referencing device. .................................................. 156

19. Means and standard deviations of average fixation duration in block 6 of "Der mechanisierte Hof" as a function of type of referencing device. ..................... 157

20. Analysis of covariance of average fixation duration in block 6 of "Der mechanisierte Hof" by type of referencing device. .................................................. 158

21. Means and standard deviations of average fixation duration in block 7 of "Der mechanisierte Hof" as a function of type of referencing device. ..................... 159
22. Analysis of covariance of average fixation duration in block 7 of “Der mechanisierte Hof” by type of referencing device. ................................. 160

23. Summary results of ANCOVAs for average fixation duration in blocks of “Der mechanisierte Hof”. ................................. 161

24. Means and standard deviations of average fixation duration in block 1 of “Im Volksgarten” as a function of type of referencing device .................................. 162

25. Analysis of covariance of average fixation duration in block 1 of “Im Volksgarten” by type of referencing device .... 163

26. Means and standard deviations of average fixation duration in block 2 of “Im Volksgarten” as a function of type of referencing device .................................. 164

27. Analysis of covariance of average fixation duration in block 2 of “Im Volksgarten” by type of referencing device .... 166

28. Means and standard deviations of average fixation duration in block 3 of “Im Volksgarten” as a function of type of referencing device .................................. 168

29. Analysis of covariance of average fixation duration in block 3 of “Im Volksgarten” by type of referencing device .... 169

30. Means and standard deviations of average fixation duration in block 4 of “Im Volksgarten” as a function of type of referencing device .................................. 170

31. Analysis of covariance of average fixation duration in block 4 of “Im Volksgarten” by type of referencing device .... 171

32. Means and standard deviations of average fixation duration in block 5 of “Im Volksgarten” as a function of type of referencing device .................................. 172

33. Analysis of covariance of average fixation duration in block 5 of “Im Volksgarten” by type of referencing device .... 173

34. Means and standard deviations of average fixation duration in block 6 of “Im Volksgarten” as a function of type of referencing device .................................. 174

48. Analysis of covariance of percentage of total fixation time in block 2 of "Der mechanisierte Hof" by type of referencing device .............................................. 193

49. Means and standard deviations of percentage of total fixation time in block 3 of "Der mechanisierte Hof" as a function of type of referencing device .............................................. 194

50. Analysis of covariance of percentage of total fixation time in block 3 of "Der mechanisierte Hof" by type of referencing device .............................................. 195

51. Means and standard deviations of percentage of total fixation time in block 4 of "Der mechanisierte Hof" as a function of type of referencing device .............................................. 196

52. Analysis of covariance of percentage of total fixation time in block 4 of "Der mechanisierte Hof" by type of referencing device .............................................. 197

53. Means and standard deviations of percentage of total fixation time in block 5 of "Der mechanisierte Hof" as a function of type of referencing device .............................................. 199

54. Analysis of covariance of percentage of total fixation time in block 5 of "Der mechanisierte Hof" by type of referencing device .............................................. 200

55. Means and standard deviations of percentage of total fixation time in block 6 of "Der mechanisierte Hof" as a function of type of referencing device .............................................. 202

56. Analysis of covariance of percentage of total fixation time in block 6 of "Der mechanisierte Hof" by type of referencing device .............................................. 203

57. Means and standard deviations of percentage of total fixation time in block 7 of "Der mechanisierte Hof" as a function of type of referencing device .............................................. 204

58. Analysis of covariance of percentage of total fixation time in block 7 of "Der mechanisierte Hof" by type of referencing device .............................................. 205
59. Summary results of ANCOVAs for percentage of total fixation time in blocks of “Der mechanisierte Hof”. 206

60. Means and standard deviations of percentage of total fixation time in block 1 of “Im Volksgarten” as a function of type of referencing device. 207

61. Analysis of covariance of percentage of total fixation time in block 1 of “Im Volksgarten” by type of referencing device. 208

62. Means and standard deviations of percentage of total fixation time in block 2 of “Im Volksgarten” as a function of type of referencing device. 209

63. Analysis of covariance of percentage of total fixation time in block 2 of “Im Volksgarten” by type of referencing device. 210

64. Means and standard deviations of percentage of total fixation time in block 3 of “Im Volksgarten” as a function of type of referencing device. 213

65. Analysis of covariance of percentage of total fixation time in block 3 of “Im Volksgarten” by type of referencing device. 214

66. Means and standard deviations of percentage of total fixation time in block 4 of “Im Volksgarten” as a function of type of referencing device. 215

67. Analysis of covariance of percentage of total fixation time in block 4 of “Im Volksgarten” by type of referencing device. 216

68. Means and standard deviations of percentage of total fixation time in block 5 of “Im Volksgarten” as a function of type of referencing device. 217

69. Analysis of covariance of percentage of total fixation time in block 5 of “Im Volksgarten” by type of referencing device. 218

xvii
70. Means and standard deviations of percentage of total fixation time in block 6 of “Im Volksgarten” as a function of type of referencing device. ........................................ 219

71. Analysis of covariance of percentage of total fixation time in block 6 of “Im Volksgarten” by type of referencing device. ........................................ 220

72. Means and standard deviations of percentage of total fixation time in block 7 of “Im Volksgarten” as a function of type of referencing device. ........................................ 221

73. Analysis of covariance of percentage of total fixation time in block 7 of “Im Volksgarten” by type of referencing device. ........................................ 222

74. Means and standard deviations of percentage of total fixation time in block 8 of “Im Volksgarten” as a function of type of referencing device. ........................................ 223

75. Analysis of covariance of percentage of total fixation time in block 8 of “Im Volksgarten” by type of referencing device. ........................................ 224

76. Means and standard deviations of percentage of total fixation time in block 9 of “Im Volksgarten” as a function of type of referencing device. ........................................ 226

77. Analysis of covariance of percentage of total fixation time in block 9 of “Im Volksgarten” by type of referencing device. ........................................ 229

78. Means and standard deviations of percentage of total fixation time in block 10 of “Im Volksgarten” as a function of type of referencing device. ................................. 230

79. Analysis of covariance of percentage of total fixation time in block 10 of “Im Volksgarten” by type of referencing device. ........................................ 231

80. Summary results of ANCOVAs for percentage of total fixation time in blocks of “Im Volksgarten” ........................................ 232

xviii
81. Pearson product-moment correlations between coreferential tie comprehension, overall comprehension, and baseline German language ability. 234
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Interaction between type of referencing device and baseline German language ability as measured by predicted average fixation duration in block 1 of “Der mechanisierte Hof”</td>
<td>142</td>
</tr>
<tr>
<td>2. Interaction between type of referencing device and baseline German language ability as measured by predicted average fixation duration in block 2 of “Der mechanisierte Hof”</td>
<td>146</td>
</tr>
<tr>
<td>3. Main effect of baseline German language ability as measured by predicted average fixation duration in block 5 of “Der mechanisierte Hof”</td>
<td>155</td>
</tr>
<tr>
<td>4. Main effects of type of referencing device and baseline German language ability as measured by predicted average fixation duration in block 2 of “Im Volksgarten”</td>
<td>167</td>
</tr>
<tr>
<td>5. Main effects of type of referencing device and baseline German language ability as measured by predicted average fixation duration in block 9 of “Im Volksgarten”</td>
<td>183</td>
</tr>
<tr>
<td>6. Interaction between type of referencing device and baseline German language ability as measured by predicted percentage of total fixation time in block 1 of “Der mechanisierte Hof”</td>
<td>190</td>
</tr>
<tr>
<td>7. Main effect of baseline German language ability as measured by predicted percentage of total fixation time in block 4 of “Der mechanisierte Hof”</td>
<td>198</td>
</tr>
<tr>
<td>8. Main effect of type of referencing device as measured by predicted percentage of total fixation time in block 2 of “Im Volksgarten”</td>
<td>211</td>
</tr>
</tbody>
</table>
9. Interaction between type of referencing device and baseline German language ability as measured by predicted percentage of total fixation time in block 9 of “Im Volksgarten” .................................................. 227


11. Percentage of incorrect responses for anaphoric expressions 16-31 of “Europa heute” .............................................. 238
CHAPTER I
THE PROBLEM

Introduction

Much research in recent years has supported the view that reading comprehension is an active and constructive process (Anderson & Pearson, 1984; Bloom & Greene, 1984; Schank, 1984; among others). In general terms, comprehension is defined as the process of relating new or incoming information to that which is already stored in memory. This process requires that readers transform the multiple units of language within a text into a cohesive conceptual representation (Bernhardt & James, 1987; Bransford & Franks, 1971). The reader's conceptual representation is known in the research literature as a discourse model (Bernhardt, 1986). According to Webber (1980), each successive sentence in a text must be determined within and integrated into the reader's discourse model. Garrod and Sanford (1977) further note that integrating information successfully involves "appreciating relationships between various objects, people, or events mentioned in the text" (p.77). In short, readers' overall comprehension will depend largely on how well they are able to reconcile the information in the text with their constructed discourse model.

In order to comprehend information accurately, readers must have precise knowledge of how linguistic features operate within texts. Linguistic features
that have the potential to facilitate or impede readers' abilities to formulate an accurate discourse model are of critical importance to the comprehension process. When managed correctly these features serve to indicate intersentential connections or, as Huggins and Adams (1980) claim, to maximize "the rate of transfer of meaning from a language producer to a language receiver" (p. 141). If, however, readers lack the requisite knowledge or processing capabilities to interpret these features accurately, the connections within the text will become obscured and the readers' comprehension is likely to suffer.

Anaphora is one such linguistic feature capable of influencing the comprehension process (Webber, 1980). Anaphoric forms are coreferential; i.e., they make reference to something else for their interpretation and can not be interpreted semantically in their own right (Halliday and Hasan, 1976). Brown and Yule (1983) describe two general types of coreference: exophora and endophora. Referential relationships that are dependent on the context of the situation and not on a text for their interpretation are said to be exophoric. According to Halliday and Hasan (1976), exophoric relationships play no part in textual cohesion and, therefore, can not impact reading comprehension. Endophoric relationships, in contrast, do form cohesive ties within texts and are dependent on the text for their interpretation. There are two kinds of endophora: anaphora, which depends on preceding discourse for its interpretation, and cataphora, which depends on succeeding discourse for its interpretation. In the following examples from Brown and Yule (1983), the anaphoric expression "it" refers back to "the sun" in sentence (1), whereas the cataphoric expression "it" refers forwards to "the sun" in sentence (2).
1. **Anaphoric Relationship**

Look at the sun. It is going down quickly.

2. **Cataphoric Relationship**

It is going down quickly, the sun. (p. 193).

Endophora is characterized by the coreferential relationships that exist between proforms (e.g., "he," "they," and "my old girlfriend") and coreferents or antecedents (e.g. "the man," "the dirty dishes," and "Maria"). These relationships must be interpreted correctly if readers hope to comprehend texts accurately.

Much research has illustrated that native readers of English have difficulty interpreting anaphoric relationships correctly and that this difficulty impedes their comprehension of texts. First language studies by Bormuth, et al. (1970), Lesgold (1974), Richek (1977), and Walker and Yekovich (1987) indicate that both children and adults sometimes have trouble identifying the correct antecedent of anaphoric forms. Other research reveals that less skilled readers find it difficult to interpret anaphoric expressions within texts (Daneman & Carpenter, 1980; Garnham, et al., 1982). Moreover, Corbett and Chang (1983), Gernsbacher (1989), and Richek (1977) found that comprehension could be improved if anaphoric connections were made more explicit. This research suggests that the misinterpretation of anaphoric structures not only can cause readers to retrieve an incorrect coreferent, but more importantly can prevent them from accurately aligning their evolving discourse models to the writer's intent. As a result, readers' overall comprehension of the text can suffer.
Although research has documented that an inability to interpret anaphora correctly can be detrimental to first language readers, little is known about how this phenomenon impacts the comprehension abilities of foreign language readers.

Statement of the Problem

The implications of anaphoric misunderstanding for foreign and second language readers are perhaps much more profound than for first language readers. Second language readers frequently lack sufficient knowledge of the foreign language to make what Webber (1980) calls sophisticated syntactic, semantic, pragmatic, inferential, and evaluative choices among possible coreferents. Foreign language readers also encounter comprehension difficulties because they must utilize underdeveloped foreign language reading skills or inappropriate first language reading strategies (Alderson, 1984).

In German, the problem of anaphoric comprehension is intensified for the reader because all nouns belong to gendered categories (masculine, feminine, or neuter) and, therefore, are replaced by gendered pronouns when referred to anaphorically. Resolving anaphoric structures accurately in German requires that readers be able to identify the gender of the antecedent. According to Drosdowski (1984), this can be a difficult task because "für Sachbezeichnungen und Abstraka lassen sich nur wenige Hinweise geben, weil sie allen drei Genera angehören" (p. 201). [For representations of things and abstract nouns there are only a few indications [of gender] because these nouns can belong to all three genders.]
In addition, Mills (1986) asserts that noun morphology is potentially confusing even for a native speaker, "since within a gender paradigm the same form can mark different cases" (p. 173). She explains the complexity of the gender and case systems with regard to article use, but the problems apply equally well to pronoun usage. The same form, for instance, can be used for different cases, such as the feminine pronoun "sie" to indicate both nominative and accusative case. In addition, identical forms can cross the gender paradigm to mark the same case, like the dative pronoun "ihm" to represent nouns of masculine or neuter gender. Finally, the same form can be used to mark different cases in different gender paradigms. The relative pronoun "der," for example, can be used to refer to a masculine noun in nominative case or to a feminine noun in dative case.

Mills (1986) and Drosdowski (1984) further note that for the most part the gender system in German honors natural gender distinctions when they apply. Whenever there is a glaring mismatch between the natural and the syntactic genders, however, anaphoric references reflect the natural gender of the noun and not the syntactic gender. The syntactically neuter noun "das Mädchen" [the girl], for example, is referred to with feminine anaphors, like "sie" [she] and not neuter anaphors, such as "es" [it]. In such instances, readers can only rely on semantic clues to determine the intended coreferent because syntactic clues will be misleading. On the whole, however, readers must use syntactic information to interpret anaphoric expressions since, as MacWhinney (1978) maintains, semantic or natural gender clues have limited applicability in German. In other words, the majority of German nouns have no natural gender distinctions, because they represent inanimate objects or abstract concepts.
Because of the complexity of the gender and case system, identifying the correct coreferent in German can be extremely difficult. The following example illustrates this difficulty.

Luise hat eine Bluse. Sie ist sehr groß.

[Luise has a blouse. It/She is very big/tall.]

Even a native German reader must tolerate some ambiguity here, because "sie" could refer to either "Luise" or "eine Bluse" [a blouse]. Both the discourse context and the topic will presumably assist the native reader in locating the correct antecedent. American readers of German, however, are likely to be influenced by their extensive knowledge of English when interpreting German sentences that contain gendered pronouns. It is expected that native readers of English are more inclined to interpret "sie" as "she" (referring exclusively to "Luise"), because in English inanimate objects like "die Bluse" [the blouse] are rarely denoted anaphorically by gendered pronouns.

By inappropriately applying first language reading skills to German text, native readers of English risk creating a discourse model that is radically different from the one intended by the author. In so doing, they greatly jeopardize their capacity to comprehend the text accurately. Studies by Bernhardt (1986) and Brozovic (1988) reveal that both German and French foreign language readers tenaciously adhere to their created discourse models when recalling texts, even if that means misinterpreting or ignoring critical syntactic features. These research findings suggest that when inexperienced foreign language readers encounter a subsequent sentence in a German text,
such as:

Sie ist zeimlich lang.

[It is fairly long.]

they are more inclined to retain their original interpretation of "sie" as "Luise" than to recognize an important semantic constraint. More competent readers of German, on the other hand, will have an easier time identifying "die Bluse" [the blouse] as the only appropriate antecedent for "sie," knowing that it would be extremely unusual to use "lang" [long] to describe a person.

Also disturbing is the fact that sentences employing anaphoric expressions, such as pronouns, do not appear to have complex surface structures. Statistical readability formulas predict that a sentence like "Sie is groß" [She is tall] is inherently easier to comprehend than a sentence such as "Luises Bluse is groß" [Louise's blouse is big] because the words are shorter and there are fewer of them (Vacca, 1981). First language studies indicate, however, that sentences in which key noun phrases are repeated are more readily comprehensible than sentences that substitute noun phrases with less explicit anaphoric references (Bormuth, et al., 1970; Garnham, et al., 1982; Gernsbacher, 1989; Richek, 1977). Second language studies also reveal that discourse containing longer, more complex sentences is easier to understand than the same discourse containing shorter, less complex sentences (Blau, 1982). In addition, Bernhardt (1986) found in an eye-tracking study that "native and highly experienced readers of German tended to allocate considerable attention to functional elements--often proportionately more attention than to
multisyllabic words" (p. 97). Such data clearly illustrate that even though anaphoric expressions are often shorter than their coreferents readers need well-developed processing skills to interpret them correctly.

Webber (1980) concurs with this view in claiming that "antecedents and coreferents are not explicit and that reasoning is involved in constructing them" (p. 152). These reasoning processes are known as inferencing rules and are used by readers to identify the possible set of antecedents for a given anaphoric structure. Webber (1980) further asserts that every anaphoric expression requires that readers employ some inferencing strategies and that if they fail to do this accurately, they "will fail to understand related anaphoric expressions and consequently fail to understand the text" (p. 152). A number of first and second language studies lend support to her contentions (Bormuth, et al., 1970; Blau, 1982; Demel, 1987; Lesgold, 1974; Richek, 1977; Walker & Yekovich; 1987; among others). It is still unclear, however, exactly what processing abilities second language learners need to comprehend texts with anaphoric forms.

This study is an attempt to provide insight into how readers at various levels of baseline German language ability process and comprehend anaphoric expressions in authentic German texts. In order to obtain this information, readers' performance on an authentic German text containing anaphoric forms will be compared to their performance on the same text in which the anaphoric forms have been eliminated and replaced with the same nouns or noun phrases that appeared initially in the text. Specifically, readers' cognitive processing of the text versions will be analyzed using an eye-movement methodology, and their comprehension of the text versions will be assessed via
written recall protocols. These measures will reveal information about the
complex cognitive processes involved in deciphering anaphoric references for
the purpose of comprehension. In addition, subjects' ability to identify
anaphoric forms correctly within a German text and the relationships between
that ability, overall text comprehension, and baseline German language ability
will be investigated.

Significance of the Problem

The majority of previous research in the study of anaphora has employed
English as the language of investigation. Most studies have explored only how
first language learners of English interpret and use anaphoric structures. The
significance of the present study is that it uniquely investigates readers' abilities
to comprehend anaphoric structures in a language other than English. At the
same time this study will make significant contributions to the existing
knowledge base in second language reading.

First language research has already documented that interpreting
anaphoric structures can be difficult for native readers of English. Few studies,
however, have investigated how anaphoric references are understood by native
or nonnative readers of other languages. A significant aspect of this study is
that it will reveal what, if any, types of anaphoric expressions are problematic for
readers of German. First language studies indicate that the acquisition of
anaphora in German is a complex and gradual process, complicated by the the
fact that both gender and case are represented in anaphoric references (Grimm,
1973; Mills, 1977a, 1977b; Park, 1976). Additional investigations suggest that
children learning German often confuse anaphoric forms with other parts of
speech that have similar properties, such as definite articles, and that the word order of a clause or sentence containing anaphoric expressions may influence children's understanding (Grimm, et al., 1975; Park, 1976). Knowing which anaphoric structures are difficult for readers will allow foreign language educators to adjust instruction so that better anaphoric comprehension among learners can be achieved.

This research will also help foreign language educators to determine whether simplified or manipulated texts facilitate reading comprehension. On the basis of research findings generated by first language readers of English, Richel (1977) recommends that reading materials be adjusted, at least in the early stages of instruction, in order to correspond better to learners' linguistic capabilities. In particular, Richel (1977) advocates the manipulation of both the amount and kinds of anaphoric expressions in texts, in order to facilitate the comprehension of first language readers. This conclusion coincides with the findings of Gernsbacher (1989) and Haviland and Clark (1974) who state that the reader's ability to comprehend complex syntactic structures is enhanced when textual connections are made more explicit. In addition, a first language study by Konopak (1988) found that readers gleaned more vocabulary information from a text in which the contextual information was made more "considerate" by increasing the proximity and clarity of relevant connections and the explicitness and completeness of relevant information. She concluded that revised texts have important instructional implications because they produce more accurate and complete word learning for students.

In a global sense, Omaggio (1984) makes the same recommendations for foreign language instruction.
We might obtain the best results [with regard to foreign language reading comprehension] by using simplified versions of authentic materials and gradually move toward incorporating complete, unedited language samples into courses. A second possibility is to provide enough extra linguistic cues to render unedited authentic materials comprehensible. (p. 71-72)

Her preference clearly is to begin foreign language reading instruction with modified texts.

Other researchers, however, argue against the use of altered or unauthentic texts (Bernhardt, 1986; Swaffer, 1985). A rationale frequently cited against their use is that modified texts often contain limited syntactic clues and reduced cohesive elements, which can impair comprehension (Honeyfield, 1977; Grellet, 1981). If, however, manipulated texts could be shown to improve comprehension through the inclusion of more syntactic clues and greater textual cohesion, then this objection would no longer be valid. In addition, Clarke (1980) and Omaggio (1984) among others warn that unaltered authentic texts may contain complexities that impede comprehension and that it is unrealistic to expect beginning and intermediate students to deal with them effectively.

At present there is no consensus among educators as to whether authentic, simplified, or created texts are the most appropriate for foreign language learners. Moreover, few foreign language studies have explored how texts could be manipulated or created to enhance students' reading comprehension abilities. If texts are to be simplified, specific guidelines must be established to determine what textual elements should be manipulated. Those guidelines are best generated via empirical research findings, which this study can provide.
There is also a need to explore whether students' overall reading comprehension abilities in a foreign language affect their capacity to understand anaphoric expressions. Several first language studies have already revealed that a relationship exists between general reading comprehension ability and the ability to interpret complex syntactic features, such as anaphora. Garnham, et al. (1982), for example, discovered that less skilled readers had trouble recognizing coreferential ties in texts, even when referential continuity was enhanced. In addition, a study by Daneman & Carpenter (1980) determined that more proficient readers are able to identify antecedents better than less proficient readers, regardless of the distance between the anaphor and its antecedent. First language research has also shown that both skilled and unskilled readers benefit from simplified anaphoric phrases that resulted in more syntactic clues and cohesive ties (Corbett & Chang, 1983; Gernsbacher, 1989; Richel, 1977). Evidence that the same phenomenon exists in second language reading would tend to support the use of manipulated texts for readers in the beginning and intermediate stages of language instruction.

Purpose of the Study

This study investigates the processing and comprehension of anaphoric expressions in German text by readers at various levels of baseline German ability. In order to carry out this investigation, three specific purposes for the study have been established. The first purpose is to examine the effect of anaphora on the general cognitive processing behaviors and reading comprehension of readers of German at various levels of baseline German
language ability. This purpose is addressed in Phase I of the study. The second purpose of the study is to assess how readers process specific anaphoric references within German texts. Phase II of the study addresses this purpose. The third purpose is to determine if readers of German are able to identify correctly the coreferents of various anaphoric expressions in a German text and if this ability is related to their overall comprehension of the text and their baseline German language ability. This purpose is addressed in Phase III of the study. Precisely stated, this study attempts to answer the following research questions:

Phase I of the Study

Question 1
What effect does the inclusion of anaphoric expressions in German text have on the general cognitive processing and comprehension of that text by readers of German at various levels of baseline German language ability as measured by the indices of fixation frequency, fixation duration, and written recall?

Question 2
How does the baseline German language ability of readers affect their general cognitive processing and comprehension of German text containing anaphoric expressions as measured by the indices of fixation frequency, fixation duration, and written recall?
Phase II of the Study

Question 3
What effect does the inclusion of anaphoric expressions in German text have on the specific cognitive processing of those anaphoric references by readers of German at various levels of baseline German language ability as measured by the indices of fixation duration and percentage of total fixation time in blocks?

Question 4
How does the baseline German language ability of readers affect their specific cognitive processing of anaphoric references in German text as measured by the indices of fixation duration and percentage of total fixation time in blocks?

Phase III of the Study

Question 5
Are readers of German able to identify correctly the coreferents of various anaphoric expressions in a German text and is this ability related to their overall comprehension of the text and to their baseline German language ability?

Theoretical Bases

A Definition of Anaphora

Although there are still disputes about exactly when and how anaphora is acquired, used, and understood, most linguists agree that it is a common
referencing device employed in both spoken and written discourse contexts. Lust (1986) defines anaphora as "the relation between a 'proform' (called an 'anaphor') and another term (called an 'antecedent' or 'coreferent'), wherein the interpretation of the anaphor is in some way determined by the interpretation of the antecedent" (p. 9). In the following sentence, for example,

Alan came home late and he was very tired.

readers' interpretation of the anaphor "he" is determined by their interpretation of the antecedent "Alan." Bloom and Hays (1978) offer a broader definition than Lust (1986) by describing anaphora as the "variation in the form of expression used to designate the same concept on first and subsequent sequences within a discourse" (p. 8). In the above example, this variation is illustrated by the use of "Alan" and "he" to refer to the same concept.

In German, anaphoric structures serve the same reference variation function as they do in English, as Drosdowski (1984) illustrates:

Mit den Formen der 3. Person (er, sie, es; sie) wird von Personen, Dingen und ähnlichem gesprochen. Dabei werden diese Formen vor allem gebraucht, um die unmittelbare Wiederholung von Substantiven in verschiedenen (Teil)sätzen zu vermeiden, um etwas vorher Genanntes identifizierend weiter zu benennen. [People, places, and similar nouns are referred to with the third person forms (he, she, it; they). These forms are used above all to avoid the direct repetition of nouns in different sentences or parts of sentences and to identify further something that was previously mentioned.] (p.318)

The anaphor "he" in the sentence shown above, for example, functions the same way when expressed in German.
Alan kam spät nach Hause und er war sehr müde.

The anaphor "er" [he] is an alternative expression used to refer to the same concept within the discourse, namely "Alan."

Webber (1980) identifies a variety of anaphoric expressions employed in discourse contexts, such as pronouns, definite or substitute noun phrases, and ellipses. According to Sanford (1985), pronouns constitute the most commonly used anaphoric form in English. Their use is also very prevalent in German. Pronouns are used to refer to nouns or noun phrases that have been previously mentioned. Bosch (1983) defines pronouns as "semantically attenuate forms" that are interpretable only through their relationship with other entities in the text (p. 203). Pronouns must agree with the nouns they refer to in number and gender. In addition, their case inflections must reflect the grammatical functions that they serve in their discourse contexts, e.g., subject, direct object, object of a preposition, etc. In English, gender agreement is based strictly on the natural or semantic gender of the noun antecedent, with very few exceptions. In German, the natural or semantic gender of the noun is generally honored, if there is one (Drosdowski, 1984; Mills, 1986). The majority of German nouns, however, represent inanimate or abstract concepts that are simply endowed with a grammatical gender, masculine, feminine or neuter, even if no particular masculine or feminine characteristics are associated with those nouns. The grammatical or syntactic gender of the noun antecedent determines the gender of the pronoun.

There are several different types of pronominal reference, including personal pronouns, possessive pronominal forms, and relative pronouns.
Personal pronouns are one word replacements used to refer to previously named nouns or noun phrases. They operate the same way in both English and German. In the simple sentences above, for instance, the anaphor "he" or "er" is an example of a personal pronoun that refers to the noun "Alan."

Alexander (1988) notes that though they are called personal pronouns, "they do not refer only to people" (p. 73). For example, the personal pronoun "it" or "er" in the following sentences refers to the inanimate object "car" or "Wagen."

Maria has a new car. It is very small.

Maria hat einen neuen Wagen. Er ist sehr klein.

The German sentence above also illustrates the use of the masculine pronoun "er" to refer to the inanimate noun, "der Wagen" [the car] which has masculine syntactic gender. In the English sentence, however, the nongendered pronoun "it" is used to refer to the inanimate object, "car," because syntactic gender distinctions rarely appear in English.

Like personal pronouns, possessive pronominal forms are used to refer to previously named nouns or noun phrases. In addition, they show possession, i. e., that someone or something belongs to somebody or something (Alexander, 1988). The same is true of German possessive pronominal forms, as noted by Drosdowski (1984): "Durch die Pronomen mein und unser, dein und euer, sein und ihr wird ein Besitzverhältnis oder ganz allgemein eine Zugehörigkeit, Zuordnung, Verbundenheit oder Zusammengehörigkeit ausgedrückt" [Through the pronouns my and our, your (singular) and your (plural), his, (its) and her a relationship of ownership or in general an affiliation, an association, a
relationship, or a connection is expressed] (p. 321-322). Determining the intended coreferent for German possessive pronominal forms can be complicated by the fact that the "sein" form is used to refer to both masculine and neuter nouns and, thus, can be interpreted as "his" or "its" depending on the discourse context.

Celce-Muria & Larsen-Freeman (1983) state that possessive pronominal forms fulfill two main functions. They can serve as a possessive adjective before a noun phrase or they can serve as a possessive pronoun by replacing a noun phrase inflected for possession. The same distinction between possessive adjectives and pronouns exists in German as noted by Conant (1974) and Drosdowski (1984). Possessive pronoun forms, however, rarely occur in modern German usage and when they do, typically reflect poetic language. The following sentence pairs illustrate possessive pronominal usage in English and German.

**Possessive Adjective**

Susan is studying chemistry. Her book is on the table.

Susan lernt Chemie. Ihr Buch ist auf dem Tisch.

**Possessive Pronoun**

All computers can calculate. Ours can also draw.


In both examples, the possessive adjective and pronoun refer back to previously mentioned or implied persons and demonstrate their ownership of
the object being described.

Relative pronouns, a third type of pronominal reference, are used in clauses "to relate" a descriptive or qualifying phrase to a noun antecedent (Alexander, 1988, p. 16). In German, the most frequently used relative pronouns are "der," "die," and "das," and, according to Drosdowski (1988), "werden als Stellvertreter eines Substantives (+ Artikel) gebraucht" [are used as substitutes in the place of nouns (+ article)] (p. 331). The German relative pronoun forms "der," "die," and "das" correspond to the English forms "who" and "that," used for people, and "which" and "that," used for things. Also common in German is the use of the relative pronouns "welcher," "welche," and "welches," similar to English "which." German and English relative pronoun forms are inflected for case, reflecting the grammatical function that the relative pronoun serves in the clause. Case inflections are obligatory in German relative clause constructions and are important syntactic and semantic markers. In English relative clause constructions, however, case distinctions are less critical and often ignored. This is apparent, for example, in the dwindling use of the relative pronoun "whom" (Celce-Murcia & Larsen-Freeman, 1983).

In addition to marking case, German relative pronouns must reflect the gender and number of their noun antecedents. No such gender and number markers exist in English, although the relative pronoun "who/whom" is reserved specifically for references to human antecedents. The following sentences demonstrate relative pronoun usage in English and German.

The cap which (that) is lying on the table belongs to Frank.

Die Mütze, die (welche) auf dem Tisch liegt, gehört Frank.
In both examples, the relative pronoun "which" ("that") or "die" ("welche") serves as the subject of the relative clause and refers back to the referent noun, "die Mütze." The German relative pronoun also agrees with its feminine, singular antecedent, "die Mütze," in gender and number.

In contrast to pronominal references, which consist of only one-word anaphors, definite or substitute noun phrases are composed of a unified string of words used to refer to a previously introduced concept. In the following English and German sentence pairs, for instance, "the old man" [der alte Mann] is an anaphoric noun phrase substituted for the proper noun "Mr. Harding."

Mr. Harding lives on our street. The old man has the biggest house on the street.

Herr Harding wohnt auf unserer Straße. Der alte Mann hat das größte Haus auf der Straße.

Bloom and Hays (1978) note that definite noun phrases generally serve to provide the reader with additional information about the antecedent involved. In the above example, the reader's knowledge of "Mr. Harding" is expanded by the informative anaphoric description of him as an "old man" [der alte Mann]. This adjunct information will be useful to readers only if they are able to recognize that these two sentences are connected by the fact that "Mr. Harding" and "the old man" [der alte Mann] refer to one and the same person.

Although definite noun phrases supply the reader with more information, ellipses serve just the opposite function. Often denoted in the linguistic literature by the symbol Θ, ellipses indicate that a noun or verb, or a noun or verb phrase, has been deleted. Elliptical constructions have also been called
"the absence of a referring expression where one is predicted by the syntax or the semantics [of the discourse] (Green, 1989, p. 32). Ellipses are used in English and German to avoid repeating nouns or verbs that are believed to be implicitly understood in the context. According to Drosdowski (1984), "wenn gleichwertige Sätze Redeteile gemeinsam haben, braucht das Gemeinsame nur einmal ausgedrückt werden" [when equal sentences have parts of speech in common, the common element only needs to be expressed once] (p. 636). The following are examples of elliptic nouns in English and German.

Alan came home late and Θ was very tired.
Alan kam spät nach Hause und Θ war sehr müde.

The following sentences illustrate elliptic verbs in English and German.

Steven is driving to Italy and Heidi Θ to Sweden.
Steven fährt nach Italien und Heidi Θ nach Schweden.

In both cases, the reader must identify the null noun or verb phrase, i.e., the one that is omitted, as well as the antecedent in order to comprehend the sentence correctly.

Green (1989) further notes that most coreferents for elliptical constructions are unambiguous, but that this is not always the case. In the following English and German sentences, it is not clear whether Mrs. Smith gave Dana a savings bond [einen Pfandbrief] or whether Mr. Smith gave a savings bond [einen Pfandbrief] to Mrs. Smith.
Mr. Smith gave Dana a dictionary and Mrs. Smith Θ (?) a savings bond.

Herr Smith hat Dana ein Wörterbuch gegeben und Frau Smith Θ (?) einen Pfandbrief.

In order to resolve an ambiguous ellipsis, like this one, the reader must look to the content and context of the preceding discourse to provide clues about the author's intended meaning.

Green (1989) expands on Webber's (1980) list of anaphoric expressions, offering a more exhaustive inventory. She begins by noting that, "personal pronouns are the stereotypical anaphoric expressions, but many other expressions pose the same sorts of problems of determining what previously mentioned entities or constructs they are intended to refer to" (p.28). In her list, Green includes the same anaphoric expressions cited by Webber (1980) but also adds several other types, among them: epithets and demonstratives.

Epithets function much like the definite and substitute noun phrases described above, except that the phrase is most often disparaging or abusive. Like definite or substitute noun phrases, epithets also impart additional, if unflattering, information to the reader about their antecedents, as shown in the example sentences below.

Mr. Brown lives in our town. The old miser has the most expensive house in town.

Herr Brown wohnt in unserer Stadt. Der alte Geizhals hat das teuerste Haus in der Stadt.

Demonstratives frequently serve as adjectives when combined with nouns or noun phrases. Demonstrative noun phrases function much the same way
that definite and substitute noun phrases or epithets do. They are merely a combination of a demonstrative and a noun or noun phrase as in “this house” [dieses Haus] or “that woman” [jene Frau]. In the above sentence, for example, the epithet, “the old miser” [der alte Geizhals] could be combined with a demonstrative to form the phrase “this old miser” [dieser alte Geizhals]. In English and German, “this” [diese] and the English plural form “these” are used to refer to someone or something close to the speaker/writer in either space or time. In contrast, “that” [jene] and the English plural form “those” are reserved for persons or things removed from the speaker/writer in space and time (Alexander, 1988; Drosdowski; 1984).

When used alone in the place of a noun or noun phrase, demonstratives can also serve as pronouns. In German, more often than English, demonstrative pronoun usage is quite common, as the following sentences illustrate.

Have you read this book? No, I have not read this (it).


Drosdowski (1984) further notes that with demonstrative pronouns, “weist der Sprecher/Schreiber in besonderer Weise auf eine Person, Sache usw. hin; er deutet sozusagen ‘mit dem Zeigefinger’ auf jemanden oder etwas hin, das bereits bekannt oder näher zu kennzeichnen ist” [(with demonstrative pronouns) the speaker/writer refers to a person, a thing, etc.; he points, so to speak, ‘with his finger’ at someone or something, that is already known or is to be further identified] (p. 324). As with other forms of pronominal reference,
German demonstrative pronouns must be marked for case and match the number and gender of their antecedents. In English, there are no case or gender distinctions; however, English demonstratives do reflect the number of their coreferents, either singular or plural.

Demonstrative pronouns, like “this” [diese] and “that” [jene], can also be used to refer to a series of ideas mentioned in previous discourse. According to Drosdowski (1984), “die neutralen Formen können auf einen ganzen Satz bezogen werden” [the neutral forms can be used to refer to an entire sentence] (p. 324). In the example below, “that” refers to the entire first sentence and not to one specific element in the sentence.

Yesterday we went to the zoo. That was a lot of fun!
Gestern haben wir den Zoo besucht. Das hat viel Spaß gemacht!

Indefinite pronouns in English and German, such as “someone” [jemand], “one” [man], and “nothing” [nichts] have a general and indefinite meaning. Drosdowski (1984) notes that “der Sprecher/Schreiber braucht sie (die Indefinitpronomen), wenn er ein Lebewesen, ein Ding usw. nicht näher bezeichnen will oder kann” [the speaker/writer uses them (the indefinite pronouns) when he does not want to or can not more precisely represent a living thing, a thing, etc.] (p. 335). The following example demonstrates this usage.

He definitely wanted to meet someone.
Er wollte unbedingt jemanden kennenlernen.
(Drosdowski, 1984, p. 334).
He notes, however, that indefinite pronouns can also be used "wenn er (der Sprecher/Schreiber) eine begrenzte Menge, ein begrenztes Maß unbestimmt ausdrücken will" [when he (the speaker/writer) wants to express vaguely a limited quantity, a limited amount] (p. 335). In other words, the author may have a specific group in mind, but chooses to refer to that group in a nonspecific way. It is this usage that is illustrated in the following sentence.

The teachers in the bookstore among others wanted the new book.
Die Lehrer in dem Buchladen unter anderen wollten das neue Buch haben.

Because they are nonspecific, indefinite pronouns generally have antecedents that are not explicitly mentioned in the text but must be inferred from the context.

Going a step beyond inventories that merely describe various types of anaphoric expressions such as those provided by Webber (1980) and Green (1989), Cole (1974) ranked referencing devices according to how much information they conveyed to the reader. The Lakoff-Cole scale, listed below, ranks referencing devices in order from most to least informative.

Proper Name
Margaret Louise is on the phone again.

Definite Description
The insurance salesman is on the phone again.
Epithet

That loudmouth is on the phone again.

Pronoun

He is on the phone again.

Ellipsis

Margaret Louise wants Θ to see you.

The Lakoff-Cole scale is essentially a guideline that authors must follow if they intend to make designated concepts clear to the reader.

According to Bloom and Hays (1978), two general rules regulate the process of concept designation.

1. An expression when uttered [written] designates the most arousable concept it can.

2. The expression that designates a concept on a given occasion is the least informative that would then designated it successfully.” (p. 28)

They claim that arousability is determined by the strength and recency of a previously mentioned concept, as well as its contiguity with other aroused concepts in the discourse. The informativeness of a designator is “simply a measure of how much detail it explicitly supplies” (Bloom & Hayes, 1978, p.28). A third person pronoun, for example, reveals nothing but gender and number, whereas a definite description or a relative clause can be highly informative. Bloom and Hayes (1978) assert that taken together, the arousability of concepts and information value of the words they represent are enough to account for
most simple designations.

It is further noted, however, that "designators are not uttered [written] in a communicative vacuum," but that they require a cognitive and linguistic context (Bloom & Hayes, 1978, p. 33). In short, what has appeared in previous discourse and what is likely to appear in succeeding discourse influences what concept designators an author uses in current discourse. As concept designators, anaphoric expressions, therefore, depend on discourse contexts for their application and interpretation.

The Function of Anaphora in Discourse Contexts

Anaphora is essentially a linguistic device used to refer to previously named or implied concepts within a discourse. The use of anaphoric expressions in texts permits an author to demonstrate semantic variation when referring to previously named concepts. The anaphoric forms discussed above illustrate how this semantic variation can be achieved in texts. Anaphora, therefore, serves to reduce the redundancy of language in a discourse context through the use of alternative, but conceptually congruent, expressions.

According to Wasow (1979), the reduction of language forms via anaphoric references not only shortens sentences but also simplifies them. In order for this simplification to be comprehensible, however, Wasow (1979) cautions that "it is necessary that the speaker of a language be able to identify correctly the elements participating in an anaphoric relation and to determine correctly the meaning of the anaphor on the basis of meaning of the antecedent" (p. 2). In using anaphora the author assumes that readers will be able to recognize this simplification and that the correct coreferent will be accessible to them via the
indirect anaphoric expression.

Webber (1980) warns, however, that the antecedents of anaphoric structures do not refer to specific elements in the text but rather to concepts "evoked in the reader's discourse model" (p. 144). The writer presumes that the coreferent of the anaphoric expression used will correspond to the same concept in the reader's own discourse model. Moreover, the author will generally use "the least informative adequate expression" to designate a previously introduced concept in discourse (Bloom & Hayes, 1978, p. 32). What functions as an "adequate" reference for native readers, however, may not provide sufficient information for foreign language readers. As a result of cultural differences and a lack of proficiency in the language, second language readers may be unable to locate the intended coreferent in their inferior discourse models.

When reading an anaphor, Dell, et al. (1983) claim that the reader must not only identify the intended coreferent, but also "establish the connection between the propositions containing the anaphor and the coreferent" (p. 121). Because of this connecting function, anaphora has the power to affect textual cohesion. In fact, Freebody and Anderson (1983) claim that "a major form of cohesion is referential" (p. 279). Gupta (1982) also asserts that anaphoric usage is "intimately related to and even controlled by the same processes that govern other aspects of discourse cohesion" (p. 6). Halliday and Hasan (1976) define textual cohesion as "the relations of meaning that exist within the text" (p. 4). In their view:
Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a relation of cohesion is set up, and the two elements, the presupposing and the presupposed, are thereby at least potentially integrated into a text. (p. 4)

Lesgold, et al. (1979) contend that a discourse must be coherent to be acceptable; “that is, the full meaning of any sentence in the discourse can be determined only in the context of information provided by other sentences” (p. 291). Cohesion is an important component of texts, therefore, because it can identify for the reader the connections between individual propositions, sentences, and global structures.

Halliday and Hasan (1976) further imply that referential terms, such as anaphoric expressions, must be integrated with other elements in the text to be understood properly.

These items [referential terms] are directives indicating that information is to be retrieved from elsewhere...the information to be retrieved is the referential meaning, the identity of the particular thing or class of things that is being referred to; and the cohesion lies in the continuity of reference, whereby the same thing enters into the discourse a second time. (p. 31)

According to this description, cohesion is an assumption on the part of the reader that there will be continuity of reference in the discourse, so that referential terms can be interpreted accurately. When the continuity of reference in the discourse is apparent and textual cohesion is high, the reader can interpret referential terms easily. However, when the continuity of reference in the discourse is obscure and textual cohesion is low, interpreting referential terms is likely to be more difficult.
It is easy to see, therefore, how anaphora has the power to influence textual cohesion: the *interpretation* of any anaphor in a discourse is dependent on its antecedent and the use of an anaphoric expression *presupposes* the existence of some antecedent. If anaphors are clearly connected to previous discourse, they can serve to enhance textual cohesion and can be readily understood by readers. If anaphoric references are poorly connected to previous discourse, however, textual cohesion may suffer and interpreting the references correctly may prove difficult. What constitutes clear versus poor connections between anaphoric references and antecedent elements in discourse needs further investigation, particularly how such connections are interpreted and understood by second language readers.

Definition of Terms

*Anaphoric Expression*: For the purposes of this study the term "anaphoric expression" will refer to personal, relative, indefinite, and demonstrative pronouns; possessive adjectives; and definite noun phrases.

*Baseline German Language Ability*: The subject’s score on the 1990 reading portion of the National German Examination for High School Students-Level IV. This exam is a discrete-point test designed to measure students’ German language ability on vocabulary, grammar, and reading tasks.

*Average Fixation Duration*: The average time per fixation, measured in milliseconds (msec.), that the eye stops to gather information. A pause of 50 milliseconds or longer with a saccade on either side of the pause constitutes a fixation.
Average Fixation Duration in Blocks: The average time, in milliseconds, spent fixating information within a given block in the text containing either a noun/noun phrase or an anaphoric expression.

Block: In a text passage, one of a number of regular rectangular regions that are delineated by a grid.

Fixation Frequency: The number of times that the eye stops to fixate when reading the text.

Percentage of Total Fixation Time in Blocks: The total amount of time, measured in seconds (sec.), spent fixating within a given block in the text containing either a noun/noun phrase or an anaphoric expression divided by the total fixation time.

Reading Comprehension: Subjects' understanding of the text as measured by their scores on immediate written recall protocols.

Saccade: Eye movements between fixations.

Total Fixation Duration: The total amount of time the eye spends fixating the text, measured in seconds.

Assumptions of the Study

1. It is assumed that because anaphora is a discourse phenomenon (Bloom & Hays, 1978; Webber, 1980), its effect on text processing and comprehension is best investigated in a discourse context.

2. It is expected that subjects will cooperate fully and perform the experimental tasks to the best of their ability.

3. The reading portion of the AATG National German Examination for High School Students--Level IV (1990) is presumed to be an adequate
measure of readers' baseline German language ability.

4. Eye-tracking variables are assumed to be adequate and valid measures of cognitive processing. When subjects are fixating a region of the text, they are presumed to be processing the information contained in that part of the text. In other words, readers are expected to try and interpret the meaning of each word as they encounter it in the text.

5. It is assumed that the readers' written recall protocols accurately reflect their comprehension of the text and thereby, are valid measures of their reading comprehension ability.

6. The coreferent selection task is assumed to be a valid measure of readers' coreferential tie comprehension and will reveal what discourse features and selection strategies influence readers' interpretation of anaphoric relationships.
CHAPTER II
REVIEW OF THE LITERATURE

Introduction

In its most general sense, anaphora is a referencing device used to refer to previously named concepts on subsequent occasions within a discourse. Numerous first language research studies have been conducted to investigate how anaphora is acquired, processed, and understood. The majority of these studies have dealt with the factors that affect the acquisition and comprehension of anaphoric references. Two research perspectives have prevailed in these types of studies. One perspective has treated anaphora as a linguistic phenomenon that operates primarily at the local discourse level; i.e., either within individual sentences or across adjacent sentences. The other perspective has acknowledged that the interpretation of anaphoric expressions is often influenced by a broader discourse context, one that may include several syntactic, semantic, and pragmatic elements. This realization pushed the study of anaphora beyond inquiries focused only on local or micro-levels of discourse to include research that addressed how anaphoric resolution occurs within global or macro-levels of discourse.

Other studies have examined the specific cognitive processes involved in anaphoric resolution and comprehension. These research endeavors have employed an eye-movement methodology to probe the internal mental
processes of readers. This research has generally corroborated the findings of studies conducted at the micro- and macro-levels of discourse, which indicate that readers begin processing anaphors as soon as they are encountered within sentences or texts.

Second language investigations of this issue have been particularly meager. Little research has been conducted to determine how anaphoric expressions are resolved by foreign language learners of English or other languages either intrasententially or intersententially. Some second language research studies in reading have examined how selected lexical, syntactic, or contextual clues influence comprehension, but few have focused explicitly on learners' understanding of anaphoric forms. The few studies that have addressed anaphoric comprehension in second language learners have investigated primarily their acquisition and use of relative pronoun forms and have not considered other types of anaphoric expressions.

Some case study and experimental research has provided minimal information about how anaphora is acquired and used by first language learners of German. These studies have revealed that the acquisition of the anaphoric system in German is a complex and gradual process. Other studies have indicated that the concept of gender is also acquired gradually and may play a significant role in assisting learners to produce and comprehend anaphoric forms.
First Language Studies of Anaphora at the Micro-Level of Discourse

First language studies of anaphora at the micro-level of discourse have provided much insight into the various factors that impact anaphoric comprehension within a single sentence or across two adjoining sentences. Central to this area of research is the work of Haviland and Clark. In 1974 these researchers proposed a comprehension strategy, called the Given-New Contract, that is based on the assumption that the speaker’s/writer’s purpose in communication is to impart new information to a listener/reader, who in turn integrates the new information with old information already stored in memory. Communication, therefore, is defined as a cooperative effort between the speaker/writer and the listener/reader. The speaker’s/writer’s task is to assess the given information, what the audience already knows, as well as the new information, what the audience does not yet know, and to proceed accordingly. The listener’s/reader’s job is to identify the given information and to treat it as an indication of what is already known and stored in memory and then to integrate the new information with the recently activated given information.

In a subsequent work, Clark and Haviland (1977) defined the Given-New Contract as a precept to the speaker/reader:

Try to construct the given and the new information of each utterance in context (a) so that the listener [reader] is able to compute from memory the unique antecedent that was intended from the given information and (b) so that [s]he will not already have the information attached to that antecedent (p. 9).

In addition, they delineated three requirements for the fulfillment of the Given-New Contract: (a) appropriateness—the given part of the sentence must
convey known or knowable information and the new part must convey unknown information; (b) uniqueness—the given information must enable the listener/reader to compute a unique antecedent; and (c) computability—the listener/reader must be assumed to have sufficient knowledge and skill to be able to compute the intended antecedent.

In their 1974 study Haviland and Clark investigated the Given-New Contract, using Direct-Antecedent and Indirect-Antecedent sentence pairs. The sentence pairs below are examples of Direct-Antecedent (1) and Indirect-Antecedent (2) pairs.

1. Direct-Antecedent Pair
   We got some beer out of the trunk. The beer was warm.

2. Indirect-Antecedent Pair
   We checked the picnic supplies. The beer was warm. (p. 514-5).

The reading times of adult readers were measured for the different sentence pairs. As hypothesized, they found that the comprehension time of the target sentences was faster for the Direct-Antecedent pairs than for the Indirect-Antecedent pairs. In the Direct-Antecedent condition the relationship between the given and new information was more “direct,” so less processing time was needed. In the Indirect-Antecedent condition the relationship was less apparent, so the reader had to make some inferences, which required more processing time. Haviland and Clark (1974) concluded that readers do, in fact, employ the Given-New Contract when reading.
One problem with the Haviland and Clark (1974) study was the nature of some of the sentence pairs employed. In order to insure that the Direct-Antecedent pairs were not remembered faster simply because they contained the repetition of a key noun, they tested sentence pairs in which the key noun was repeated but was not necessarily part of the reader’s given information. The following example illustrates this point.

Andrew was especially fond of beer. The beer was warm. (p. 516).

The “beer” mentioned in the first sentence refers to beer in a generic sense and not a specific quantity of beer like the “beer” in the second sentence. The first sentence provides no antecedent for “the beer” referred to in the second sentence because the second sentence fails to include any given information. In using sentence pairs of this type the researchers seemed to violate the very principle they were trying to investigate.

Another flaw in the study was the way in which comprehension was defined. In the study, reading time was equated with comprehension time. Subjects were told to read the sentence pairs and then press a button as soon as they thought they understood what the second sentence meant. The researchers simply had to assume that the sentences were correctly understood because what subjects actually comprehended was not measured. In effect, therefore, the study only indicated how fast readers recognized given-new relationships, but not how well those relationships were understood. Similar problems have plagued many other studies of anaphora at the micro-discourse level. Nonetheless, these studies (particularly, the work of Haviland and Clark,
1974) have provided a useful foundation for subsequent studies that have investigated the role of inference and integration in the understanding of coreferents.

A study by Hupet and Le Boudec (1977) tested the Given-New Contract with a group of adult native readers of French. Four simple sentences were presented to subjects in different orders: (a) a logical order; (b) an orderly mixed order (with one sentence out of logical sequence); and (c) a randomly mixed order (with no two sentences in logical sequence). Subjects were then asked to integrate the four sentences into a single complex idea. The findings fully supported those of Haviland and Clark (1974). The more accessible the intended antecedents were for each simple sentence, the better the subject's reconstruction of the sentences as an integrated whole. The mixed order conditions, in contrast, left students with no way to determine the intended referents for each simple sentence and, therefore, no way to integrate the information well enough to recall it successfully.

Garrod and Sanford (1977) examined the effect of semantic distance on readers' comprehension of related sentence pairs. Semantic distance was defined in terms of Battig and Montag's (1969) concept of "conjoint frequency." Battig and Montag (1969) analyzed the assignment of different exemplars to specific classes of items. Items that were frequently cited and accepted as exemplars of a particular class were labeled high conjoint frequency exemplars and items that were seldom cited and accepted as exemplars were labeled low conjoint frequency exemplars. "Robin," for example, was listed as a high conjoint frequency exemplar of the class "bird," whereas "goose" was listed as a low conjoint frequency exemplar. The reading times for each sentence and
the response times for yes-no questions about the sentences were recorded.

The findings revealed that readers required more time to process sentence pairs in which the target item was paired with a low conjoint frequency exemplar than for sentence pairs in which a high conjoint frequency exemplar was used. As a result, sentence pairs like (1) below, involving a high conjoint frequency exemplar, were comprehended more readily than sentence pairs like (2), in which a low conjoint frequency exemplar was presented.

1. High Conjoint Frequency Exemplar
   A bird would sometimes wander into the house.
   The robin was attracted by the larder.

2. Low Conjoint Frequency Exemplar
   A bird would sometimes wander into the house.
   The goose was attracted by the larder. (p. 79).

These results tended to support readers’ use of the Given-New Contract, in that it was more difficult for readers to locate the intended antecedent for a low conjoint frequency item and this difficulty, in turn, forced the readers to make inferences that cost them valuable processing time.

In two similar studies, Garnham (1981; 1984) investigated the encoding of category nouns, e.g., weapon, and instance nouns, e.g., knife, as functions of (a) whether their antecedents were category or instance nouns; and (b) whether the preceding context implied what kind of object was being referenced. When the context revealed no specific information about an instance noun, as in
sentence pair (1) below, the mention of that instance noun in a subsequent sentence presented new information about the antecedent to the reader. The results showed that this new information took additional time to process. When the context had already suggested an instance of a category noun, however, as in sentence pair (2) below, the mention of the suggested instance noun did not require more processing time.

1. The cloth was kept in big rolls.
   The denim would fade in the sun.

2. The cloth was made into jeans.
   The denim would fade in the sun.

The researcher concluded for both studies that interpretations of anaphoric expressions established during reading are the product of both the linguistic input and of the reader’s general knowledge. Garnham (1984) further asserted that the results of the two studies taken together “confirm the importance of an anaphor’s information content in determining how much processing it requires” (p. 1).

Garvey, et al. (1974) conducted a study to examine the role of semantic factors in the assignment of antecedents for potentially ambiguous pronouns. Subjects answered questions about individual sentences and performed a sentence completion task. The results indicated that the semantic property of the verb exerted an important and even “dominant” influence over the selection of antecedents for the ambiguous pronouns (p. 240). The semantic verb
property, labeled *implicit causality* in an earlier study by Garvey and Caramazza (1974), dictates the preferred interpretation of anaphoric expressions in sentences. In the following sentences, for example, the implicit causality of the verbs predicts that the preferred coreferent for “she” in (1) will be Mary and in (2) will be Jane.

1. Jane hit Mary because she had stolen a tennis racket.

2. Jane angered Mary because she had stolen a tennis racket. (p. 601).

According to the findings, verbs vary in their degree of implicit causality, i.e., the degree to which they restrict or promote a particular interpretation. The verb “telephone,” for example, strongly promotes the choice of the subject as the coreferent for an anaphor, whereas the verb “anger” only weakly promotes this choice. The experimenters concluded that implicit semantic features of verbs do influence pronoun-antecedent assignment. They also claimed that this influence appears to operate within and across sentential boundaries.

In a subsequent study, Caramazza, et al. (1977) investigated the implicit causality of different classes of verbs: (a) those that biased the antecedent assignment toward the first noun phrase in the main clause and (b) those that biased the assignment toward the second noun phrase in the main clause. Pairs of sentences containing verbs from both classes were constructed. The subordinate clause in one sentence reflected a reading consistent with the natural bias of the verb, as in sentence (1) below. The second sentence contained a subordinate clause that reflected a reading inconsistent with the natural bias of the verb, as shown in sentence (2).
1. Tom scolded Bill because he was annoying.

2. Tom scolded Bill because he was annoyed. (p. 604)

The reaction time, i.e., the time it took subjects to indicate their choice of an antecedent, was recorded for each sentence. The results indicated that the reaction times were faster for sentences with consistent readings than for sentences with inconsistent readings. This was also the case for control sentences, in which gender differences eliminated the ambiguity.

The researchers found that implicit causality was not the only verb-based feature to influence pronoun assignment. A feature called experciencer constraint also played a role in determining the reader's choice of an antecedent. Experiencer constraint predicts that "it is much more likely that, in discussing a private experience, the experiencer himself (if present) will be the speaker rather than any other participant in the conversation" (p. 607). In other words, a person who has experienced something, e.g., an emotional state or an event, is in a position to make statements about that experience. Certain verbs, by virtue of their semantic properties, mark either their subjects or their objects as the experiencer and, thus, restrict pronoun-antecedent assignment in complement constructions.

This phenomenon had already been reported in an earlier study by Springston (1976; cited in Caramazza, et al., 1977). He found that pronoun assignment times were faster for sentences that exhibited experiencer constraint, such as (1) below, than for sentences like (2) that were unconstrained.
1. Bill told Sue that Mary bored him.

2. Bill told Sue that he bored Mary. (p. 607)

This was the case even for sentences like those above, in which the gender constraint disambiguated the pronoun for readers.

The conclusions of Caramazza, et al. (1977) match those of the Garvey, et al. (1974) study, i.e., “implicit causality is an important determinant of pronoun assignment” (p. 601). The researchers also found that the principle of experiencer constraint significantly influences pronoun resolution. Caramazza, et al. (1977) further argued that their results support the position that pronoun ambiguities are generally resolved at clausal boundaries.

A study by Grober, et al. (1978) also tested subjects’ ability to resolve potentially ambiguous pronouns within sentences. Adult subjects were presented with several sentence fragments and told to complete the sentence. The sentence fragments were similar to the sentence types used in the Caramazza, et al. (1977) and Garvey, et al. (1974) studies; for example, “John must scold Bill because he....” The researchers hypothesized that since the pronoun following the subordinate conjunction is the grammatical subject of that clause, it would be interpreted as being coreferential with the grammatical subject of the main clause. In other words, “he” in the example provided above would be interpreted as referring to “John” in the subjects’ sentence completions. This approach to pronoun resolution was identified as the parallel function strategy in Sheldon’s (1974) study of relative clause acquisition in children.
The results of the Grober, et al. (1978) study indicated that for over 70% of all the sentence fragments, subjects selected the grammatical subject of the main clause as the antecedent for the pronoun in the subordinate clause. In cases where the subject of the main clause was not selected as the antecedent, the researchers concluded that the semantic character of the verb influenced subjects' choices. Apparently, when the bias value of the verb was especially strong it overrode the subject's inclination to use the parallel function strategy. These findings lend further support to the position that the implicit causality of verbs influences pronoun resolution. This study clearly reveals that both syntactic factors, e.g., parallel function, and semantic factors, e.g., implicit causality, determine pronoun-antecedent assignment.

Garnham and Oakhill (1985) examined the effects of both verb semantics and inference making on pronoun resolution. Subjects were presented with sentences containing pronouns and asked to answer yes/no questions about them. Some of the pronouns could be resolved on the basis of gender and number alone and others required the readers to make inferences based on their knowledge of the world. Sentence (1) below illustrates the former type of pronoun and sentence (2) illustrates the latter.

1. Alan lent a pen to Jill, because she wanted to write a letter.
2. Vicky lent a pen to Jill, because she wanted to write a letter. (p.386).

Reading times for the sentences, question-answering times, and responses to the yes/no questions were recorded.
The results indicated that subjects spent more time reading a sentence that required them to make inferences in order to resolve the pronoun. In contrast, subjects needed less time to read the same sentence when it contained a pronoun that could be resolved on the basis of its gender and number alone. This finding suggests that pronouns are resolved “on-line,” i.e., as they are read, and that resolution is not delayed until subjects are asked to answer questions about the sentences. These results are consistent with those of Haviland and Clark (1974) and Garrod and Sanford (1977) in that they all found that “people make the inferences needed to produce coherent representations of texts sentence by sentence as they read them” (p. 392).

The portion of the study investigating the effect of verb semantics on pronoun-antecedent assignment did not corroborate previous research. Garnham and Oakhill (1985) determined that the natural bias of the verb only had a facilitative effect on reading times when the pronoun could also be resolved according to its gender and number. The natural verb bias did not improve processing times in the absence of a gender and number cue. These results differ from the findings of Caramazza, et al. (1977), who determined that pronouns that were consistent with the natural bias of the preceding verb had their antecedents named more quickly than those that were not consistent, regardless of whether a gender clue was given or not. Garnham and Oakhill (1985) concluded that the demand of making an inference slows readers down considerably and possibly masks the implicit causality effect of verbs.

A study by Chang (1980) provided evidence that gender-mismatched names are not retrieved in pronoun-antecedent assignment. He used a probe recognition task to assess readers’ comprehension of sentences, in which the
proper names in the first clause differed in gender, as shown in the following example.

Bill and Mary went to the store, and he bought a quart of milk (p. 292).

The results showed that the presence of a masculine pronoun, e.g., “he,” in the second clause did not lead to retrieval of a feminine name, e.g., “Mary,” from the first clause. According to Chang (1980), these findings suggested that the set of potential antecedents is limited to concepts that do not mismatch the semantic constraints of the pronoun. The phrasing does not mismatch versus matches was deemed important because the pronoun could provide previously unknown information about the antecedent. In the following sentence, for example, “professor” does not mismatch the semantic constraints of “she” and, therefore, can be accepted as a potential referent.

Since the professor had a flat tire, she was late for class.

Chang (1980) further noted that the metaphorical usage of pronouns, such as “she” for ships and “he” or “she” for storms, could be considered a violation of this general rule. He argued, however, that these uses are perhaps best thought of as “learned homonyms of the pronouns,” since “[s]he” can not be employed indiscriminately to refer to inanimate objects. Obviously this principle of pronoun resolution applies well to English language examples. This rule is, however, not applicable to a language such as German in which grammatical, as well as semantic, constraints of the gendered pronoun must be recognized in
order to identify the intended antecedent.

In a later study, Corbett and Chang (1983) investigated the activation of the
correct antecedent versus a second potential antecedent for three types of
anaphoric reference: (a) a proper noun, (b) a pronoun, and (c) an ellipsis.
Responses to yes/no antecedent recognition probes and comprehension
questions were analyzed. The findings indicated that when pronouns or
ellipses were used as the referencing device, the nonantecedent probes were
processed faster than when proper nouns were used. The researcher surmised
that the inclusion of a proper noun in the second clause actually suppressed the
nonantecedent (the other proper name in the first clause) from being activated
and that such suppression did not occur with pronouns and ellipses. In
addition, there were longer response times and lower comprehension scores
for the pronoun and ellipsis conditions.

The study also examined whether semantic information in the predicate of
a clause restricts the search for an antecedent for the subject pronoun of that
clause. In the sentence below the semantic information in the second clause is
sufficient to disambiguate the pronoun.

Scott stole the basketball from Warren and he sank a jumpshot (p. 684).

The results revealed, however, that the predicate information in the clause
was not used to limit the set of possible antecedents; i.e., the nonantecedent
"Warren" was activated in spite of semantic restrictions presented in the second
clause. The researcher concluded that when pronouns or ellipses are
encountered in sentences, both antecedents and nonantecedents (of the
appropriate gender for pronouns) are activated during pronoun-antecedent resolution.

Richek (1977) also varied the types of anaphoric reference presented in experimental sentences. The results showed that manipulating the amount and kinds of anaphoric structures in sentences significantly ($p < 0.01$) affected the comprehension abilities of elementary school students. Specifically, she discovered that repetitions of the original noun form were much easier for students to understand than were pronoun forms, and that ellipses were the most difficult for them to comprehend. This study also indicated that learners' frequent exposure to a syntactic pattern did not influence their ability to comprehend that pattern. In other words, even though a sampling of texts revealed that the readers had encountered elliptic and pronoun structures far more often in their reading than noun forms, this exposure did not enhance their comprehension of those structures. Instead, Richek (1977) determined that readers' comprehension "depends on the amount of information that is immediately available to the reader," i.e., present in the text itself (p. 159).

Research by Gernsbacher (1989) sought to identify ways of facilitating referential access. He proposed two mechanisms by which readers' access to antecedents could be improved: enhancement and suppression. According to Gernsbacher (1989), "enhancement improves the accessibility of previously mentioned concepts by increasing or boosting their activation," whereas "suppression improves concepts' accessibility by decreasing or dampening the activation of other concepts" (p. 99). It was assumed that the informational content of the anaphor would trigger these mechanisms appropriately. A series of six experiments investigated this assumption by varying the referencing
device presented in the experimental sentences. Anaphoric reference was made via very explicit, repeated name anaphors, as in sentence (1) below or via less explicit pronouns, as in sentence (2).

1. Ann predicted that Pam would lose the track race, but Pam came in first very easily.

2. Ann predicted that Pam would lose the track race, but she came in first very easily.

While subjects read each sentence, the activation level of the noun-noun or the noun-pronoun combination was assessed by a probe verification task. Afterwards comprehension was measured via WH questions, such as “Who won the race?”

The results of these studies tended to support the findings of Corbett and Chang (1983) and Richel (1977). The first two experiments revealed that explicit, repeated name anaphors immediately trigger the enhancement of their correct antecedents as they cause the suppression of nonantecedents. The third experiment demonstrated that less explicit pronoun anaphors also trigger the suppression of other nonantecedents, but this occurs less quickly than with explicit anaphors. This was true even for the fourth experiment when the semantic information needed to identify an antecedent appeared in a previous sentence, as in the sentence pair example below.

Bill lost a tennis match to John.

Accepting the defeat, he walked quickly toward the showers.
The results of the fifth experiment indicated that more explicit pronouns, i.e., ones that provide a gender cue, cause the suppression of nonantecedents more powerfully. A final experiment showed that not only rementioned nouns but also newly introduced nouns improve referential access by fostering the suppression of nonantecedents. In the following sentence the newly introduced noun, “Mark,” triggers the suppression of the nonantecedents “Bill” and “John."

Bill handed John some tickets to a concert, but Mark said the tickets were counterfeit.

Gernsbacher (1989) concluded that both suppression and enhancement improve referential access. How rapidly and powerfully these two mechanisms are triggered is a function of the concepts’ explicitness. From the these data and that of other studies, the researcher devised an *explicitness principle*: “The more explicit the concepts, the more likely they are to trigger the suppression of other concepts, and, when used anaphorically, the more likely they are to enhance their antecedents” (p. 135).

These studies have investigated how subjects select antecedents for anaphoric expressions located either intrasententially or across two adjacent sentences. As such, subjects’ ability to identify coreferential ties was measured in the absence of the multiple syntactic, semantic, and cohesive cues inherent in longer pieces of connected discourse. In addition, many of the sentences and sentence pairs used in these studies were designed for explicit experimental purposes. The results, therefore, can not legitimately be generalized to nonmanipulated, authentic discourse examples. In spite of the valuable information these studies have revealed, they offer little insight into
how readers perceive and interpret anaphoric relationships within a macro-level discourse context or within naturally-occurring, authentic text.

First Language Studies of Anaphora at the Macro-Level of Discourse

Several first language studies have revealed the important role that macro-level discourse features play in readers' comprehension of anaphoric structures. An early study by Bormuth, et al. (1970) investigated children's ability to comprehend intrasentential and intersentential syntactic elements, as well as anaphoric forms. The anaphoric expressions tested were selected from such categories as relative pronouns, personal pronouns, noun phrase demonstratives, clause demonstratives, and semantic substitutes. The target syntactic structures were embedded into four- or five-sentence paragraphs created for the study. Comprehension was tested via completion type and multiple-choice type questions.

The findings revealed that a majority of students "were unable to demonstrate a comprehension of the most basic syntactic structures by which information is signaled in language" (p. 355). From the results, the following hierarchy for children's comprehension of syntactic elements was proposed: intrasentential syntactic structures are the easiest for children to understand, intersentential syntactic forms are second, and anaphoric structures are the most difficult for them to comprehend. The researchers indicated, however, that this hierarchy could be an artifact of the measurement techniques; i.e., anaphoric structures were the only ones tested with multiple-choice type questions instead of completion type questions.
Lesgold's (1974) research called the Bormuth, et al. (1970) findings into question. Lesgold (1974) examined children's understanding of the same anaphoric forms investigated by Bormuth, et al., (1970) but constructed answers for the multiple-choice questions that were semantically sensible with the correct choice dependent on the anaphoric syntax. The results revealed a difficulty order that had a significant negative correlation with the hierarchy obtained by Bormuth, et al. (1970). The investigator concluded that children's comprehension of anaphoric forms depends not only on their knowledge of syntactic structures but also on the semantics of the discourse context. Moreover, he asserted that semantics must be accounted for in any model attempting to explain children's understanding of syntax.

Research by Barnitz (1980) also produced results that contradicted the findings of Bormuth, et al. (1970). This study investigated how children in grades two, four, and six comprehend pronoun-referent structures in short constructed passages. In particular, Barnitz (1980) compared: (a) referent type--noun and noun phrase antecedents versus clausal and sentential antecedents, (b) reference order--anaphoric versus cataphoric expressions, and (c) referent distance--intersentential versus intrasentential ties. Children's responses to comprehension questions about the target structures served as the dependent variable. The findings revealed that noun phrase referents were easier to comprehend than clausal or sentential referents and that anaphoric references were easier to understand than cataphoric references. In addition, no significant differences were found between children's comprehension of intrasentential and intersentential ties. The researcher concluded that, with a few exceptions, pronoun-referent structures are generally comprehensible to
children by the time they reach the sixth grade.

A study by Freebody and Anderson (1983) assessed the effects of textual cohesion, topic familiarity, and vocabulary difficulty on children’s comprehension of social studies text. They based their definition of cohesion on the work of Halliday and Hasan (1976), who state that “cohesion lies in the continuity of reference, whereby the same thing enters into the discourse a second time” (p. 31). Freebody and Anderson (1983) hypothesized that when textual cohesion was high, the processing load placed upon readers would not be substantial and, therefore, that they could easily determine the antecedent. When anaphoric references were ambiguous or complicated, however, textual cohesion would be reduced and additional effort would be required for the reader to identify the intended referent. Using this concept of cohesion the researchers created high and low cohesive forms of the passages. High cohesive text versions employed the repetition of key referents with identical lexical items, whereas low cohesive versions included substitutions, pronouns, and ellipses. A third level of cohesiveness, labeled the inconsiderate version (Kantor, 1977), contained propositions irrelevant to the main theme of the text. No significant main effects for cohesion were found, nor was there an interaction between cohesion level and vocabulary difficulty. Echoing Barnitz’s (1980) findings, the researchers concluded that cohesion, “in the specific sense of linguistic ties” may not be as “important in reading” as had been previously believed (p. 293).

Garnham, et al. (1982), however, determined that textual cohesion can impact readers’ understanding of connected discourse. They compared the comprehension of skilled and less skilled readers, who read constructed stories
presented in three versions: (a) original versions; (b) randomized versions, i.e., versions in which the sentences appeared in a random order and, therefore, where the referential continuity had been destroyed; and (c) randomized versions in which sentences appeared in random order, but the referential continuity of the passage had been restored. The findings revealed that restoring referential continuity by replacing pronouns and other terms with fuller and more appropriate noun phrases increased the skilled readers' comprehension of the randomized stories. The investigators concluded that although original story versions were remembered better than random versions, the replacement of nouns with appropriate nouns and noun phrases alleviated the detrimental effects of randomization. In fact, they asserted that restoring referential continuity restored textual cohesion, even though the sequence of events in the randomized story versions may have appeared "slightly odd" (p. 42). This coherence, in turn, improved the comprehensibility of the randomized stories. In addition, the researchers noted that readers who lacked the skills to recognize coreferential ties between sentences to begin with benefited little from the restoration of coherence in the randomized stories.

In contrast to Garnham, et al. (1982), Roen's (1984) study corroborated the findings of Barnitz (1980), Freebody and Anderson (1983), and Lesgold (1974). The researcher investigated adult readers' comprehension of texts at two different levels of reference: low and high. High reference consisted of text versions containing personal pronouns or demonstratives that referred back to previously named nouns, whereas low reference versions included substitute nouns, synonyms, or near synonyms in the place of the previously named nouns. No significant effects for reference were found, indicating that varying
referencing devices within texts may not influence reading comprehension. A significant main effect, however, was found for topic. Roen (1984) concluded that topic may be a strong factor affecting comprehension, whereas micro- and macro-level textual features may exert less influence over comprehension than previously believed.

The controversy over the effects of coreference and cohesion on comprehension was clarified somewhat by Walker and Yekovich (1987). In their experiment, subjects read scripted texts embedded with sentence pairs that were manipulated to reflect different antecedent-anaphor relationships. The referential tie between the two sentences was either: (a) explicit--consisting of a repeated noun phrase; (b) implied--an ellipsis; or (c) absent--the first sentence contained no referent. In addition, the anaphors represented either central or peripheral concepts within the text. Subject's reading times for the anaphoric sentences served as the dependent measure.

The results revealed that the reading times for anaphoric sentences referring to central concepts were the same regardless of the type of referential tie used. In other words, subjects read sentences with explicit referential ties as quickly as they read sentences with implied or absent referential ties, providing the anaphor in the sentence was a concept central to the text. These findings essentially match those of Barnitz (1980), Freebody and Anderson (1983), Lesgold (1974), and Roen (1984). When subjects read anaphoric sentences containing peripheral concepts, sentences with explicit referential ties were read the fastest, followed by implied referential ties, with absent referential ties requiring the most processing time. These results coincide more with the findings of Bormuth, et al. (1970) and Garnham, et al. (1982).
The researchers concluded that the relationship of an anaphor to the topic or script of a text affects the reading time needed to process the anaphor. In particular, Walker and Yekovich (1987) noted that concepts central to the script or topic appear to have an elevated level of activation and "consequently enjoy clear processing advantages" (p. 690). Peripheral concepts, in contrast, are not processed as readily and, in fact, require more explicit textual references to become activated. The nature of the relationship between the anaphor and the discourse topic may account for the disparate results obtained in previous studies.

Lesgold, et al. (1979) also investigated the impact of topic on reading comprehension. Specifically, their study examined Chafe's (1973) notion of foregrounding in discourse comprehension. Foregrounding refers to the status of a concept at various points in a text following the initial introduction of the concept. A concept is considered foregrounded if a speaker/writer can assume that the listener/reader has the concept actively in mind when it is mentioned again in the discourse. If this is not the case than the concept is considered backgrounded. Lesgold, et al. (1979) asserted that "foregrounding and backgrounding determine the acceptability of certain sentence forms in a discourse" (p. 291). The last sentence of (1) below, for example, is more acceptable because the concept "girl" is foregrounded, whereas the last sentence of (2) is less acceptable because the concept "girl" is backgrounded.

1. I am trying to find a black dog. Yesterday that dog bit a little girl.
   
   She was scared but she wasn't really hurt.
2. Yesterday a black dog bit a little girl. It got away and we are still trying to find it. She was scared but she wasn’t really hurt.

The phenomenon of foregrounding corresponds to Haviland and Clark’s (1974) concept of *given* information and Frederiksen’s (1981) notion of *topicality*.

In a series of five experiments, Lesgold, et al. (1979) measured adult first language readers’ response times and continuation responses for target sentences containing anaphors whose referents were either foregrounded or backgrounded. In all experiments, target sentences took longer to comprehend when the information they referred to was backgrounded. The results also showed that a concept will remain foregrounded as long as it can be integrated with concepts in subsequent sentences. In addition, the researchers found that repeating the context can facilitate the reinstatement of a concept in the reader’s mind, even if the concept itself is not explicitly reintroduced.

Lesgold, et al. (1979) essentially reinforced the earlier findings of Kintsch and Vipond (1978), who also investigated Chafe’s (1973) concept of foregrounding. Kintsch and Vipond (1978) theorized that comprehension occurs in short-term memory, which can only accommodate a small number of propositions from a text at any one time. Which propositions are stored in short-term memory is determined jointly by the location of the propositions in the macro-structure of the text and by how recently they were encountered. In other words, foregrounded propositions are the ones held in short-term memory. If no proposition in short-term memory matches the one being processed, i.e., if the proposition is backgrounded, then a time-consuming search of long-term memory ensues until the desired proposition is found and reinstated. With
regard to anaphoric expressions, the researchers determined that a pronoun is
easier to comprehend if its antecedent is foregrounded, because the
antecedent will still be lodged in short-term memory. They further predicted that
an antecedent will be foregrounded if it is located high in the macro-structure of
the text and is in close proximity to the pronoun.

A study by Clark and Sengul (1979) also provided evidence for the
phenomenon of foregrounding in comprehension. These researchers
examined the reading times of adult first language readers who read short three
sentence paragraphs. Subjects then read a fourth sentence containing the
target information, either a noun or pronoun whose referent was located in
Sentence (1), (2), or (3) of the preceding paragraph. Referents located just one
sentence back, i.e., in Sentence (3), were identified quicker than referents that
were located two or three sentences back, i.e., in Sentence (2) or (1). In
addition, the results showed that subjects identified referents located in the last
clause of Sentence (3) faster than referents located in the first clause of
Sentence (3). Clark and Sengul (1979) concluded that the propositions in the
last clause processed are given a special place in short-term memory and that
they are readily understood when referred to via substitute nouns or pronouns.
According to their research, the last clause boundary appears to be the defining
limit of what constitutes foregrounded information for the short-term memory.

Findings by Daneman and Carpenter (1980) corroborated the work of
Clark and Sengul (1979). The Daneman and Carpenter (1980) study
measured the relationship between reading span (the maximum number of
sentences a subject could read while maintaining perfect recall of the final
words) and readers’ responses to pronoun questions about a text. The distance
between the pronoun and its referent was systematically varied in the text versions. All readers, regardless of reading span abilities, tended to do better on passages with smaller distances between the pronoun and its noun antecedent. The findings also indicated that subjects with larger reading spans performed better on all pronoun-referent distances. The researchers claimed that in order to identify the antecedent of an anaphoric expression correctly, the reader must retrieve information previously given in the text. According to these findings, this is more easily achieved if the information is active in short-term memory, which will be the case if the pronoun and referent appear close together in the text or if the reader has a large reading span.

In two similar studies, McKoon and Ratcliff (1980) and Dell, et al. (1983) examined how antecedent information is activated during the processing of anaphoric expressions. In the McKoon and Ratcliff (1980) study, subjects read paragraphs sentence by sentence and were then presented with a target word from the first sentence of the paragraph for recognition. In the Dell, et al. (1983) study, the target word was presented at unexpected times during the reading of the paragraph. Response times and error rates for the single-word recognition task were recorded.

The results from both studies indicated that both the antecedent, e.g., car, of a superordinate anaphor, e.g., vehicle, and other concepts or nonantecedents in the same proposition as the antecedent became activated very quickly, as early as 250 milliseconds after the anaphor was read. The antecedent remained activated as the sentence was read, but the activation of the nonantecedents deteriorated. The researchers concluded that pertinent antecedent information is initially accessed in the form of propositions, but only
those propositional concepts that are necessary for anaphoric resolution remain activated.

Frederiksen (1981) also investigated how first language learners identify referents for pronouns within texts. Short texts were constructed for the study in which the following variables were manipulated: pronouns versus lexical repetitions, the number of available referents for the pronoun, the position of the pronoun in the sentence, the distance between the pronoun and the referent, and the foregrounding of the referent noun phrase. Reading times, latencies in reporting the referent for the probed pronoun, and error rates were measured.

The results indicated that reading times increased when the referential relationship was pronominal compared to when a lexical item was simply reiterated. There was also a significant increase in reading times for sentences that had two potential referents for the the target pronoun. The investigator determined that the more ambiguous a pronoun is, i.e., the more available potential antecedents it has, the greater the reading time needed to locate its intended antecedent. A significant main effect for foregrounding or topicality was also found. Frederikson (1981) concluded that presenting topical noun phrases in subject position and continuing to refer to them throughout the text increases the likelihood that those items will be selected as referents.

Research by Clifton and Ferreira (1987) also assessed the effects of the discourse topic, the distance between an anaphor and its antecedent (far or near), and the type of reference (a definite noun phrase or a pronoun) on reading times. The findings revealed that propositions containing either definite noun phrases or pronouns with topic antecedents were read more quickly, regardless of the distance factor. In addition, subjects rated how well the
sentences with anaphors "fit" with the preceding context of the story. Subjects rated the stories that had anaphors with topic antecedents significantly higher than stories that had nontopic antecedents, regardless of the distance between the anaphor and its antecedent or the type of reference used.

After further analysis, the researchers determined that some sentences had more than one antecedent that was accessed equally quickly. They hypothesized that it is perhaps discourse centers, not topics, that are the preferred antecedents for anaphoric expressions. Discourse centers are key sentence elements that play important roles in discourse cohesion (Joshi & Weinstein, 1981; cited in Clifton and Ferreira, 1987). Forward-looking centers determine how sentence elements will be linked up to succeeding discourse, whereas backward-looking centers determine how sentence elements will be incorporated into the preceding discourse. Joshi and Weinstein (1981; cited in Clifton and Ferreira, 1987) suggested that backward-looking centers are roughly the same as the topic of the discourse, i.e., what the discourse has presented thus far. Forward-looking centers correspond roughly to the notion of the focus of the discourse, i.e., where the discourse is going. Sentences generally have several forward-looking centers and at most one backward-looking center. Clifton and Ferreira (1987) asserted that both the forward-looking centers (the discourse focus) and the backward-looking centers (the discourse topic) comprise the set of possible antecedents for an anaphoric reference.

Although these studies have revealed much about how anaphora is processed and understood within connected discourse, the conclusions must be viewed cautiously. Many of the studies used texts that were designed and
manipulated for the explicit purpose of testing anaphoric comprehension. The results, therefore, provide little information about how readers interpret anaphoric references in naturally-occurring, authentic texts.

Another major problem with many of the studies was the way in which comprehension of the anaphoric forms was assessed, i.e., via reading times, multiple-choice questions, completion type questions, error rates, and reading spans. No global measure of the subjects' understanding of the text as a piece of connected discourse was obtained. As a result, this research must be carefully interpreted before definitive judgments about the effect of coreference and cohesion on overall reading comprehension can be made.

Eye-tracking Studies of Anaphora

Studies investigating learners' comprehension of anaphoric expressions at the micro- and macro-discourse level have provided valuable information about what textual factors impact anaphoric understanding. They have supplied, however, only limited knowledge about the internal mental processing involved in anaphoric resolution. Eye-tracking research, in contrast, has provided important insight into the cognitive processes that occur during the comprehension of anaphoric expressions. Just and Carpenter (1987) state that "the time a reader spends on various parts of a text and places where he fixates or rereads the text are excellent indices of the ongoing psychological processes [involved in comprehension]" (p. 5). In addition, they note that "the time a readers spends on a word or phrase can indicate when a process occurs and how its duration is influenced by characteristics of the text, the reader, and the task" (p. 5).
Extensive eye-tracking research has furnished important baseline data about the general reading behaviors of native English readers. These data provide a useful framework for interpreting readers’ comprehension of anaphoric forms. First and foremost, eye-tracking research has determined that readers’ eyes do not follow the printed lines of text in a smooth pattern. On the contrary, the eyes make a series of pauses or fixations interspersed with rapid staccato-like movements, known as saccades. It has also been established that the average fixation duration for adult readers is approximately 250 milliseconds and that the average saccade lasts between 10 and 20 milliseconds (Just & Carpenter, 1987). In addition, research has confirmed that the typical reading rate for normal silent reading averages 225 words per minute (Just & Carpenter, 1980).

Research has also verified that decisions about when and where and even how long to fixate are influenced by characteristics of the text and its content. Several eye-tracking studies have examined the role that various textual features play in the resolution of anaphoric expressions. Of particular interest has been the issue of whether the processing of pronouns in texts occurs simultaneously with fixations or whether there is a delay between fixing a word and completing the processing of it. Studies have also addressed the effects of pronoun-antecedent distance, gender cues, implicit causality, episode shift, and spatialisation.

Ehrlich and Rayner (1983) investigated whether pronoun-referent assignment is completed when the pronoun is first encoded or whether there is some lag time between encoding the pronoun and the assignment of a referent. Adult native readers of English read short text passages, in which the distance
between the pronoun and its referent was varied, with near, intermediate, and far conditions. There was no ambiguity over the choice of a referent for the pronouns, i.e., there was only one referent that matched the pronoun in number and gender. The results indicated that the process of pronoun-referent assignment began when the pronoun was encoded, but was not necessarily completed within the same fixation, especially when the antecedent was located far back in the text. The fixation patterns revealed that when an antecedent was placed far from the pronoun in a text, readers began encoding new words in the text following the pronoun while at the same time they continued the process of pronoun-referent assignment. When the referent was located near to or at an intermediate distance from the pronoun, pronoun-referent assignment was sometimes completed during encoding resulting in longer fixations on the pronoun. The researchers concluded that the distance between a pronoun and its antecedent affects the locus and speed of pronoun-referent assignment. In addition, they noted that many cognitive processes associated with the comprehension of a word are executed while the word is being fixated, but that there is a limit to the range of processes that can be carried out immediately. They contended that only the processes directly relevant to lexical access and syntactic parsing are executed before the eyes move on to fixate the next word in the text.

In a separate study by Ehrlich (1983), subjects also read passages in which the distance between the pronoun and the antecedent was manipulated. The antecedent phrases for the pronouns denoted professions in which there was a majority of males. Both the "expected" pronoun "he" and the "unexpected" pronoun "she" were used in text versions. In addition, passages
containing pronouns that blatantly violated the explicit gender of their antecedents were also employed. The data were scored in terms of the fixation nearest the pronoun, the fixation before the pronoun, and the fixation after the pronoun.

The results revealed that when the pronoun and antecedent were not far from each other the fixation duration was longest for the fixation before the encoding of the pronoun and that when the pronoun and antecedent were far apart the fixation duration was longest for the fixation after the encoding of the pronoun. Ehrlich (1983) noted that "there is no single place in the text at which pronoun assignment always occurs" and that pronoun-referent assignment takes place later as the distance between the antecedent and the pronoun increases. These findings corroborated the results of the Ehrlich and Rayner (1983) study. The data also suggested that pronoun-referent assignment may be completed later for the pronoun "she" than for the pronoun "he."

Regressive fixations were also analyzed to determine if readers initiated regressive eye movements to identify the referents of pronouns. The results indicated that readers generally did not look back in the text to find antecedents for pronouns. When the gender of the pronoun contradicted that of the antecedent, however, there was a high percentage of regressive fixations. Ehrlich (1983) concluded that readers are able to resolve pronouns without making regressions back to the referent except in cases of explicit conflict between pronouns and their antecedents. It was determined, therefore, that the function of regressions is not to access information but to check the source of information when there is some inconsistency.
In a similar study, Kerr and Underwood (1985) investigated readers’ eye fixation patterns and reading times for short three sentence paragraphs. The paragraphs contained either “expected” pronouns, i.e., those that match sex stereotypes, such as “she” to refer to “nurse;” “unexpected” pronouns, i.e., those that differ from sex stereotypes, such as “he” to refer to “nurse;” or “neutral” pronouns, i.e., those for which no definitive sex stereotype exists, such as “she” to refer to “student.” The results revealed longer fixation durations and reading times for the sentences containing unexpected pronouns. Because the information provided by the unexpected pronoun was the opposite of what was expected according to the stereotype, the readers needed more time to identify the antecedent and to integrate the information into their discourse model. The researchers concluded that readers use this type of pragmatic information when resolving pronominal references.

Previous eye-tracking research by Ehrlich (1980) determined that pragmatic and semantic information influences readers’ comprehension of anaphoric forms. In particular, the study examined whether the choice of an antecedent is influenced by the verb in the main clause or by the semantics of the sentence. The semantics were altered by changing the conjunctions in the sentence to include “but,” “because,” and “and,” as shown in the example sentences below.

Steve blamed Frank because he spilt the coffee.
Steve blamed Frank and he spilt the coffee.
Steve blamed Frank but he spilt the coffee.
In addition, some sentences contained gender cues and some did not. In the sentence examples above "he" can refer to either "Steve" or "Frank;" therefore, the sentence provides no gender cue to the reader.

The results indicated that when the semantics of a sentence was altered by changing the conjunction in the sentence, the reader's choice of a referent also changed, even though the main verb in the sentence stayed the same. Ehrlich (1980) concluded that this contradicts the implicit causality principle proposed by Garvey and Caramazza (1974) and Caramazza, et al. (1977), whose studies only investigated sentences with the conjunction "because." The researcher claimed that the implicit causality of the verb was probably used in combination with other linguistic elements, such as conjunctions, to select appropriate referents.

A significant main effect was also found for conjunction, but only when no gender cue was present. When gender cues were present subjects' response times were shorter. Ehrlich (1980) argued that readers use general knowledge to choose referents for pronouns when gender does not identify a unique referent. She also concluded on the basis of the response times that readers analyze sentences clause by clause when assigning referents in one clause to a pronoun in another clause.

Vonk (1984) investigated the inferencing processes employed during pronoun resolution by essentially replicating Ehrlich's (1980) study. Subjects read sentences such as the following and had to identify the referent of the pronoun.
Alex lied to Andy because he smelled trouble.

As in the Ehrlich (1980) study, the gender of the pronouns and of the referents was varied to create conditions with and without a gender cue. In addition, the conjunctions in the sentences were varied to include "because," "but," and "and."

The results indicated that the vocalization latencies for sentences with gender cues were smaller than for sentences without gender cues. In addition, Vonk (1984) discovered that the fixation durations for the pronoun were longer for sentences providing a gender cue than for sentences that had no gender cue. In sentences without a gender cue, it was hypothesized that the pronoun could only be assigned by using the semantic information form the conjunction and the verb phrase of the subordinate clause. The results did indicate significantly longer fixation durations for the verb phrase in the no gender cue condition. No significant effect for conjunction, however, was found, although the nonsignificant results were in the predicted direction. There was also no significant interaction between gender cue and conjunction. This finding contradicted the results of Ehrlich (1980), who found differences among the three conjunctions in sentences that had no gender cue.

Vonk (1984) concluded that the reader is able to decide very quickly which aspects of the sentence are crucial to comprehension and to attend to those parts selectively depending on what information is needed. This phenomenon is what Vonk (1984) calls the principle of rational selection of information. The findings also lent support to the claims of Ehrlich and Rayner (1983), who maintained that words are processed as soon as they are fixated at least with
regard to lexical properties and that, although processing is initiated immediately, it may not be completed during the first fixation.

In 1985 Vonk conducted another study that examined the immediacy of inference processing in anaphoric resolution. In the experimental sentences, the main clause contained verbs with strong implicit causality. The verb phrase of the subordinate clause was either congruent or incongruent with the natural bias of the main clause verb. Sentences were also constructed with and without gender cues. Reading times per sentence and per clause and eye fixation durations were measured. The researcher found shorter reading times for sentences that had congruent verb phrases in the subordinate clause, regardless of whether a gender cue was present.

Vonk (1985) concluded that readers start the pronoun-referent assignment process as soon as they encounter relevant cues, but that the process is not completed immediately. As subsequent information is encountered, it is used simultaneously with previously activated cues to advance the assignment process. Several linguistic factors, such as gender cues and verb semantics, are exploited to determine the referents for pronouns. These findings generally corroborated those of Vonk (1984), Ehrlich (1980), and Ehrlich and Rayner (1983).

Kerr and Underwood (1987) measured readers' eye movements to determine the effect of an episode shift on pronoun-referent assignment. Subjects read passages containing an anaphoric sentence that referred to either the main or a scenario-dependent character via an unambiguous pronoun. Scenario-dependent characters are those that are connected or bound to a particular scenario and that should be less available to the reader
when the scenario has shifted to accommodate the next episode. The mention of an usher at a theater, for example, would be less accessible to the reader after the introduction of an episode shift outside the scenario boundary of “at the theater.”

The researchers found that pronominal references to the main character were more quickly recognized than references to scenario-dependent characters. This finding supported the principle of foregrounding observed in previous studies (Frederikson, 1981; Kintsch & Vipond, 1978; Lesgold, et al., 1979; among others). The greater the time changes used in the text to introduce an episode shift, e.g., seven hours versus ten minutes, the more time readers needed to identify the antecedents of scenario-dependent characters. More time was not needed, however, when time changes were kept within scenario boundaries. Kerr and Underwood (1987) concluded that readers create a mental model of the text that contains information about the episodes being described and that they use this information to assist them in the process of pronoun-referent assignment.

Pynte, et al. (1988) investigated the effect of spatialisation on the processing of ambiguous pronouns in French. They noted that French lacks a neutral pronoun, which could be used to distinguish between animate and inanimate nouns. In the absence of grammatical cues, therefore, the reader must rely on the meaning of the sentence to disambiguate the pronoun. The researchers presented sentences on a computer screen one phrase at a time and varied the spatial orientation of the phrase on the screen. In the non-spatialised condition the phrases appeared successively at the same physical location on the screen. In the spatialised condition the phrases also appeared
successively, but in their appropriate physical locations on the screens, as shown below.

<table>
<thead>
<tr>
<th>Non-spatialised Condition</th>
<th>Spatialised Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase 1</td>
<td>Le chat</td>
</tr>
<tr>
<td></td>
<td>Le chat</td>
</tr>
<tr>
<td>Phrase 2</td>
<td>mange</td>
</tr>
<tr>
<td></td>
<td>mange</td>
</tr>
<tr>
<td>Phrase 3</td>
<td>la souris</td>
</tr>
<tr>
<td></td>
<td>la souris</td>
</tr>
</tbody>
</table>

(p.216).

The results indicated that the non-spatialised condition "removes from the reader important positional cues and hence disrupts syntactic operations at the time of segment-by-segment reading" (p. 224). The researchers found that when reading sentences in the spatialised condition, subjects applied both semantic and word order grammatical rules to the assignment of ambiguous pronoun referents. Subjects reading sentences in the non-spatialised condition, however, used only semantic grammatical rules to determine the referents for ambiguous pronouns. These findings illustrated that the reader not only uses syntactic and semantic cues when processing anaphoric expressions, but also positional information about the anticipated referent.

These eye-tracking research studies have provided considerable insight into the internal cognitive processes that occur during the resolution of anaphoric references. In spite of their valuable contributions, the results of these studies must be interpreted cautiously. One major problem is that many of these studies only investigated readers' processing of single or adjacent sentences constructed for experimental purposes. As a result, the findings can
only be generalized to similar types of discourse. In addition, these eye-tracking studies only examined readers’ fixation durations, reading times, and antecedent assignments. No global measure of the readers’ comprehension of the text as an integrated whole was obtained. Although eye fixation indices reveal important information about the processes involved in comprehension, they cannot be construed as direct measures of what or how well the information in the text has been understood. This information must be tested separately via a more global assessment tool, such as the immediate recall protocol.

Second Language Studies of Syntax and Anaphora

Relatively few research studies have investigated second language learners’ comprehension of anaphoric expressions. Most of these studies have examined how L2 learners’ general knowledge of syntax or of discourse structure impacts their reading comprehension. In addition, some studies have addressed how second language learners’ acquire an ability to use and understand specific anaphoric structures, such as relative clauses. In spite of these efforts, the research base in this area is decidedly meager.

In 1986 Barnett investigated the effect of vocabulary and syntactic proficiency on the recall of an authentic text by L2 learners of French. The results indicated that knowledge of vocabulary and syntax interact to influence comprehension. Recall scores increased as vocabulary proficiency increased when students had high or intermediate control of syntax. Similarly, recall scores improved as syntactic proficiency improved when students demonstrated high or intermediate vocabulary skills. The researcher concluded that both sets
of skills are necessary for second language learners' comprehension of French text, and that pedagogy should include attention to both vocabulary building and grammar training.

A major flaw in the Barnett (1986) study is the way in which vocabulary and syntactic proficiency were measured. A multiple-choice rational deletion cloze test for which half of the blanks depended upon a knowledge of vocabulary and the other half on a knowledge of syntax was used to assess vocabulary and syntactic skills. Serious questions about the validity of the cloze procedure as a measure of understanding have been raised (Bernhardt, 1986; Markham; 1984). In addition, the researcher noted that the effect of background knowledge on reader' comprehension was not examined, yet may have been an important factor in their recalls because both texts had "definite cultural connotations" (p. 347). Despite these research flaws, Barnett's (1986) study did reveal that L2 readers' knowledge of syntax does influence their comprehension and, therefore, that explicit syntactic instruction is warranted.

A study by Bernhardt and Berkemeyer (1988) also found that knowledge of syntax played a key role in comprehension for L2 readers of German. Quantitative and qualitative analysis of recall data for four different text types—a newspaper article, a general article, a friendly letter, and a business letter—revealed that misinterpreting syntactic constructions adversely affected comprehension. Contrary to Barnett's (1986) findings that vocabulary and syntactic knowledge interact to influence comprehension, Bernhardt and Berkemeyer (1988) discovered that "a lack of vocabulary rarely interfered with learners' understandings" (p. 27). Insufficient knowledge of syntax, however, crippled L2 readers' comprehension of the German texts. Although students at
all levels were able to construct some meaning from the texts, the researchers concluded that more instructional emphasis on syntax as a "meaning carrier" in texts would improve comprehension.

Blau (1982) researched the effect of syntax on the reading comprehension of ESL students. In particular, the study challenged the sentence length criteria of common L1 readability formulae, which hold that shorter sentences are easier to read. Three versions of short English passages were developed: (a) Version 1 contained short, simple sentences; (b) Version 2 consisted of complex sentences, which included clues to underlying relationships, such as relative pronouns or repeated subjects and verbs in subordinate clauses; and (c) Version 3 contained complex sentences without such syntactic clues. The vocabulary and content of the passages were held constant. Comprehension was measured by multiple-choice questions about the passages.

The results revealed that Version 2, rather than Version 1, as predicted by readability formulae, yielded the highest comprehension scores. In addition, subjects subjectively judged Version 2 as significantly easier to understand than Version 1. Blau (1982) found that short, primarily simple sentences typical of low readability levels did not facilitate reading and may actually have impeded comprehension. Chopy, short sentences were more difficult for ESL learners to read because they destroyed important syntactic relationships between sentences. In contrast, she determined that ESL readers appeared to benefit from the information about syntactic relationships that was conveyed in the complex sentences.

In conclusion, the researcher stated that common readability formulae developed for use with native readers of English are not an appropriate means
of determining what will be easily understood by ESL readers. This study indicates that clearer syntactic relationships in discourse can improve comprehension. As such, it provides strong support for the manipulation of texts to include more syntactic clues and cohesive ties.

A pilot study by Stanley (1984) compared the comprehension of discourse structure by L1 and L2 readers of English. The researcher claimed that most definitions and descriptions of text have been concerned with "local" levels of discourse analysis, such as anaphora, reference and substitution, ellipsis, etc. She further asserted that pedagogy and reading models must be extended beyond the micro-textual level to incorporate macro-textual or global discourse analysis. The study involved the "problem-solution" model of discourse structure, developed by Winter (1977) and extended by Hoey (1983), which is one of the few models that combines surface linguistic features and local levels of discourse analysis with the overall structure of the text. This model of discourse structure was used to analyze four different summary versions. One version was a "model" summary; another was a "no-problem" summary, in which no problem was presented; another was a "no-solution" summary, in which no solution to the problem was presented; and the last version was a "random" summary, constructed by summarizing every third sentence of the original text. Subjects were asked to rank the summaries in order of preference and to give reasons for their choices.

The results showed that both L1 and L2 readers of English could identify the summaries that best reflected the structure of the original text and those that reflected it the least. First language readers most often noted the "readability," "flow," and the informational content of the summary as the reasons for their
rank ordering, whereas second language readers made general comments about the "structure" of the text. The researcher concluded that L2 readers focus more on the structure of the text than on its informational content. Reasons for this finding could be that the problem-solution discourse structure operates differently in the subjects' native language or is understood differently by individual subjects—factors that were not accounted for in this study. Stanley (1984) maintained that more research is needed to determine how and when second language readers develop the ability to recognize text structure and to use it to solve lower level syntactic and lexical problems in reading. Her research has direct relevance to the study of anaphoric acquisition and comprehension, because anaphora is a linguistic feature that operates at both the micro- and macro-levels of discourse.

Studies by Flynn (1984; 1986; 1987; 1989a; 1989b; with Espinal, 1985) are among the few to investigate directly the acquisition of coreferential ties by L2 learners. Flynn’s 1984 study specifically examined whether the principle branching direction (PBD) of the learner’s first language influenced how the learner processed complex sentences in the second language. Principle branching direction is defined as the direction, right or left, in which major recursive devices, such as relative clauses or other types of sentence complementation, can be generated in relation to a main clause. Languages such as English and Spanish are characterized as right-branching because subordinate clauses tend to be to the right of the main clause, as illustrated in the following sentence.
This is the book that the man I met on the train going to Tokyo wrote.

Languages such as Japanese, in contrast, are characterized as left-branching because subordinate clauses tend to be to the left of the main clause. The example sentence shown above would be expressed in the following way in Japanese.

This Tokyo to go train met man wrote book is.

Flynn (1984) hypothesized that if the PBD of the adult learner's first language matched that of the second language, the processing and acquisition of complex sentences in the L2 would be facilitated. If the PBD of the learner's first language was different, however, from that of the target language, then processing would be disrupted and acquisition would be delayed. The researcher had native speakers of Spanish and Japanese repeat aloud complex English sentences presented to them orally. The experimental sentences contained subordinate clauses both with and without pronominal anaphors that varied in branching direction, either right or left.

The results indicated that the PBD of the learner's first language was a significant predictor of the learner's processing and acquisition of complex sentence structures in a second language. Speakers of Spanish, where the PBD of the first language matched English, imitated the experimental sentences significantly better than the Japanese speakers. In addition, Japanese speakers made significantly more anaphora errors than the Spanish speakers on structures which included anaphora. Flynn (1984) concluded that, "second
language learners are sensitive to differences in the PBD of the first and second language and use this sensitivity to formulate hypotheses about the second language, specifically hypotheses about anaphora" (p. 81).

In subsequent studies, Flynn investigated the role of the head-initial/head-final parameter in second language learners’ acquisition of English relative clauses (Flynn, 1986; Flynn, 1987; Flynn, 1989a; Flynn, 1989b; Flynn & Espinal, 1985). Head-initial languages such as English and Spanish place the head—a noun in a noun phrase construction or a verb in a verb phrase construction—before the complement and other structures in the sentence. In the English example sentence below, the head “child” precedes the complement “who is eating rice.”

The child who is eating rice is crying.

In Japanese and Chinese, which are head-final languages, the complement precedes the head. Shown below is the same example sentence translated from the Japanese, where the head “child” follows the complement “rice-object eating.”

Rice-object eating child-subject crying is.

Flynn asserted in these experiments that when first and second language features differ, L2 acquisition is delayed because learners must learn new structures that conform to the L2 grammar. When first and second languages features match, however, L2 acquisition is facilitated because learners do not
need to learn new structures but can rely on their L1 knowledge in forming English constructions.

In these studies, groups of Chinese, Japanese, and Spanish learners were tested on their elicited imitation of English restrictive relative clauses. The groups differed according to whether the head direction of their native language did or did not correspond to English. The results of these studies revealed that adult L2 learners were constrained in their early hypotheses about grammatical anaphora by the head-direction of their first language, regardless of whether the L1 was a head-initial or a head-final language.

The findings also disclosed significant differences in acquisition patterns between learners who had an L1 that matched English in head direction versus learners whose L2 did not match English in head direction. An analysis of the learners' errors in Flynn's (1989a) study further revealed that Spanish speakers made primarily lexical errors while Japanese speakers made more structural errors. The researcher found that acquisition was significantly disrupted for the Japanese and Chinese speakers because they had to revise their knowledge of head direction to accommodate English constructions. Acquisition of these sentence structures for the Spanish speakers, however, was greatly facilitated as they did not have to adjust their understanding of head direction but could rely on their L1 knowledge. These findings generally corroborated the results of Flynn's (1984) earlier investigation of principle branching direction in learners' first and second languages. On the basis of these studies, the researcher concluded that an adequate theory of second language acquisition must take into account the structural aspects of the learner's first language, such as principle branching direction and head
direction.

The results of the Flynn studies must be interpreted cautiously because of the method of assessment employed. The elicited sentence imitation task measures only subjects’ ability to mimic sentences aloud. It is questionable whether this ability can be construed as comprehension of pronominal reference. In a 1986 study Flynn even described the elicited imitation task as a “production” task not a “comprehension” task. The sentence imitation task, therefore, may not be a valid assessment of the subjects’ actual comprehension ability with respect to cohesive ties.

In 1983 Tarallo and Myhill investigated the acquisition of relative clauses for L2 learners of left-branching languages (Chinese and Japanese) and right-branching languages (German and Portuguese). In particular, they measured the learners’ ability to judge the grammaticality of relative clause constructions in individual sentences in the second language. The grammaticality of the sentences was manipulated by placing incorrect resumptive pronouns at various grammatical positions in the relative clauses. The following German sentence illustrates an ungrammatical example from the study, in which a resumptive pronoun, “er” [ne] has been incorrectly included in the subject position of the relative clause.

\[ \text{Der Mann, der er aus Deutschland kommt, ist reich.} \]

[The man who he from Germany comes is rich.]

\[ \text{[The man who comes from Germany is rich.]} \]
The results revealed a high percentage of acceptance for ungrammatical resumptive pronouns, but the pattern of acceptance varied according to whether the learner was studying a left- or right-branching language. Resumptive pronouns were tolerated less often in the subject position by students of right-branching languages (German and Portuguese), whereas students of left-branching languages (Chinese and Japanese) were less tolerant of resumptive pronouns in the direct object position. One reason for this finding offered by the researchers was that learners’ knowledge of basic phrase structure rules of the L2 prompted them to reject L2 structures which did not conform to those rules. Students of Portuguese and German learn from the very beginning that the basic word order for these languages is Noun Phrase (Subject)-Verb Phrase-Complements. Thus, these students would recognize that resumptive pronouns are less likely to occur in subject position. The experimenters concluded that many of the errors made could best be explained as natural processes of language acquisition and not as first language interference.

A study by Demel (1987) investigated the relationship between adult ESL learners’ understanding of coreferential ties in a text and their overall comprehension of the text. Specifically, she examined whether the errors learners made when selecting coreferents for certain linguistic elements within a text distorted their recall of the text. Subjects were characterized according to English language proficiency (high and low) and L1 anaphoric system (languages in which overt pronouns are not required when the coreferent is available from elsewhere in the text and languages in which overt pronouns are permitted even when the coreferent is available from the context).
The results showed no differences between English language proficiency levels or between L1 anaphoric systems. The correlation between the dependent measures of coreferential tie comprehension and overall text comprehension was 0.68. An analysis of the errors made by the L2 readers revealed that comprehension problems with respect to coreferential ties stemmed from either (a) an unfamiliarity with a descriptive expression used as an antecedent, e.g., “the old man,” or (b) a lack of knowledge of the cultural concept conveyed by a particular descriptor, e.g., the negative feelings imparted by calling one’s father an “old man.” Demel (1987) concluded that instruction about how cohesive elements within texts operate would improve the comprehension of L2 readers.

Because the research base is so small, relatively little is known about L2 learners’ acquisition and comprehension of anaphoric expressions. Second language studies of general syntax and discourse structure imply that such an important linguistic feature as an anaphora could strongly influence learners’ understanding of connected discourse. The studies by Flynn suggest that the learners’ first language plays a key role in how anaphoric references are comprehended in a second language. More research is clearly needed, however, to determine how anaphora is acquired, used, and understood by second language learners.

First Language Studies in the Acquisition of Anaphora in German

Few research studies have investigated how first language learners of German acquire and comprehend anaphoric forms. The majority of studies that
have been conducted have examined only how German relative clause constructions are acquired and understood by children. The earliest evidence of children's comprehension and use of relative clauses comes from the diary studies of Stern and Stern (1928). They found that children inserted a meaningless syllable in the place of a relative pronoun until about age three. It seemed that the semantic and pragmatic information in the clause was sufficient to make the children's meaning clear. Even though the children could not produce correctly inflected relative pronoun forms, this did not mean that they could not form intelligible relative clause constructions.

This finding was supported by the work of Grimm (1973). She reported that children aged three to four years old used primarily "wo" in the relative clauses they produced. In German, "wo" [where] is the locative interrogative pronoun, that can also be combined with prepositions to form relative pronouns. Mills (1986) notes, however, that in some dialects "wo" is used as a relative pronoun in the subject and object positions. It was predominantly this usage of "wo" that was observed in the children's speech, as shown by the example below.

Das ist ein Mädchen, wo in die Schule geht.
That is a girl who in the school goes
[That is a girl who goes to school.] (p.205).

Grimm (1973) also found that at about age four the children produced an intermediate form by combining "wo" with a standard relative pronoun, e.g., "der," "die," "das," etc. The researcher concluded that both forms are used until
the child feels confident that the standard inflected form alone is correct and conveys the semantic information clearly.

A study conducted by Park (1976) considered the effects of branching direction and the position of the relative clause in relation to the main clause on children's ability to imitate sentences. Three types of relative clause constructions were tested: (a) final right-branching, (b) initial left-branching, and (c) initial right-branching. The following example sentences illustrate these relative clause types.

1. Final Right-Branching Relative Clause
   Ich brauche eine Sekretärin, die gut tippt.
   I need a secretary who well types!
   [I need a secretary who types well.]

2. Initial Left-Branching Relative Clause
   Wer das schreibt, muß vollkommen besoffen sein.
   whoever that writes must completely drunk be
   [Whoever writes such things must be completely drunk.]

3. Initial Right-Branching Relative Clause
   Der Mann, der Würste verkauft, ist mein Vater.
   the man who sausages sells is my father
   [The man who is selling sausages is my father.] (p.206-7).
The results revealed that the children found it easier to imitate relative clauses in final position rather than in initial positions. Of the relative clauses in initial position, left-branching clauses were easier to imitate than right-branching clauses. Because the main clauses were interrupted, sentences with right-branching relative clauses were thought to be more difficult for the children to interpret.

Grimm, et al. (1975) investigated children's comprehension of right-branching relative clauses using a toy-moving task with four and five year olds and an interview procedure with six and seven year olds. The researchers manipulated (a) the position of the relative clause (initial or final) and (b) the function of the relative pronoun (subject or object). Their findings indicated that initial subject clauses were easier to understand than final subject clauses, but final object clauses were easier than initial object clauses. In general, sentences where the relative pronoun was the subject of the clause were easier to comprehend than sentences where the relative pronoun was the object of the clause. Mills (1986) surmised that the frequency with which relative pronouns occur in nominative case and the association of the subject with initial positions in the sentence makes interpreting subject relative clauses easier.

The investigators also suggested that children may have difficulty recognizing relative clause structures because relative pronouns closely resemble definite articles. To test this hypothesis, they measured children's understanding of the same sentences using a deviant linguistic form, the relative pronoun followed by "wo." In general, the compound form improved comprehension. Grimm, et al. (1975) concluded that the insertion of "wo" highlighted the relative pronoun and eliminated any confusion with the definite
articles. As a result, the children found these clauses easier to recognize and process.

Mills (1977a) tested the comprehension of slightly older children, aged five to eight, using the same types of clauses and a picture selection task. In addition to manipulating the position of the relative clause and the case of the relative pronoun, a masculine form clearly marked as nominative or accusative was used as either the relative pronoun or as the definite article for the second noun in the relative clause. The following sentences illustrate these structures.

1. Masculine Relative Pronoun in Nominative Case

   Der Mann, der das Mädchen ruft, folgt dem Jungen.
   the man who the girl calls follows the boy
   [The man who calls the girl follows the boy.]

2. Masculine Definite Article in Accusative Case

   Die Frau, die den Mann erschießt, schlägt den Polizisten.
   the woman who the man shoots hits the policeman
   [The woman who shoots the man hits the policeman.]

The results mirrored those of Park (1976) in that relative clauses in final position were easier to understand than those in initial position. Moreover, clauses with relative pronouns as the subject were comprehended better than clauses with object relative pronouns, which matched the Grimm, et al. (1975) findings. Mills (1977a) also noted that the older children still had problems interpreting the case information within the relative clause. She concluded that
the relative pronoun can only be identified as such when it is clear that it is not a definite article and this clarification comes relatively late in the clause. The definite article, on the other hand, is easier to recognize as such because it is immediately followed by a noun.

In yet another study, Mills (1977b) varied the location of the relative clause within the sentence and examined the tendency of children to associate the subject with the first position in the clause. Using deliberately ambiguous clauses, the researcher measured the comprehension of adults and children aged 6 to 13 years old. Any semantic or pragmatic information in the clause that might help to disambiguate the meaning was purposely limited. In addition, words were chosen that would leave the relative pronoun and the definite article of the second noun in the clause unmarked for case. The following example sentences demonstrate this ambiguity.

1. Ambiguous Relative Clause in Initial Position
   Die Katze, die das Mädchen sieht, beißt den Hund.
   the cat that the girl sees bites the dog
   [The cat that sees the girl bites the dog.] OR
   [The cat that the girl sees bites the dog.]

2. Ambiguous Relative Clause in Final Position
   Der Hund sieht das Mädchen, das die Krankenschwester ruft.
   the dog sees the girl who the nurse calls
   [The dog sees the girl who calls the nurse.] OR
   [The dog sees the girl whom the nurse calls.]
The findings showed that the relative pronoun was chosen as the subject of the clause significantly more often than as the object, regardless of the age group. Moreover, the preference to interpret the relative pronoun as the subject of the clause was not affected by the location of the relative clause within the sentence. The investigator concluded that these results support previous research, which suggests that “the subject is strongly related to the first position in the clause.” (Mills, 1986, p. 212). The researcher further noted that this preference may be due to an association with the standard German word order SVO (Subject-Verb-Object) of main clauses.

In addition to relative clause acquisition, some research has investigated children’s acquisition of natural versus syntactic gender in German. This issue has important implications for learners’ acquisition and comprehension of anaphora, because German anaphoric expressions reflect the syntactic gender of their coreferents. Mills (1978) investigated the ability of children aged 5 to 10 to assign the correct definite article to 10 familiar nouns, all of which were the names of toys. The subjects were presented with all three gender possibilities, “der” for masculine, “die” for feminine, and “das” for neuter, and asked to choose the correct one. Mistakes were made only by 5- and 6-year-olds, and even they made very few mistakes.

The researcher found that the selection of the gender seemed to be dependent on the type of noun presented. The greatest number of errors were made on the masculine and neuter animate nouns. In contrast, few mistakes were made on masculine and neuter inanimate nouns. The feminine gender nouns, both animate and inanimate, had the highest percentage of correct responses. Mills (1978) reported that in some instances the noun was clearly
animate but the sex of the noun was not readily apparent. In these cases, the children tended to assign the feminine form, "die" to the noun.

The investigator concluded that two main factors influence children's assignment of gender: (a) the animacy of the noun, and (b) a tendency to overgeneralize their use of "die." Mills claimed (1986) that the overgeneralization of the feminine form could be attributed to the frequency with which it appears in the language. "Die," as a feminine and plural marker, accounts for 50% of all the nominative and accusative forms of the definite article in German.

Research by MacWhinney (1978) also examined children's ability to assign gender to German nouns. Real and nonce words with similar structures were chosen to determine the effects of inherent semantic gender and phonological ending on the selection processes of children aged 3 to 12 years. The words were presented under three different conditions: (a) with no cues, (b) with overt cues through the previous mention of the indefinite article in accusative case, or (c) with overt cues through the previous mention of the pronoun in accusative case.

The results indicated that age improved the children's ability to perform the assignment task. In particular, older children were better able to make use of overt cues and phonological information, especially when dealing with the nonce words. There was also a tendency for the older children to overgeneralize the neuter form "das," particularly with the nonce words. This finding contradicted the work reported by Mills (1978), which found that "die" was most often overgeneralized. Mills (1986) claimed that the children's preference for "das" could be explained by the fact that many of the words used
were monosyllabic. Monosyllabicity is strongly associated with masculine and neuter genders, but not with the feminine gender. In addition, MacWhinney (1978) found that German children make little use of inherent semantic gender when determining syntactic gender, because this information has limited applicability.

A study by Böhme and Levelt (1979; cited in Mills, 1986) tested children's comprehension of the masculine and neuter possessive pronoun "sein" [his/its] versus the feminine possessive pronoun "ihr" [her/its] for toys that reflected natural gender and objects distinguished only by syntactic gender. Children three to five years old performed worse on the toys which provided a natural gender cue than on the objects that were differentiated only by syntactic gender. The concept of natural gender did not appear to aid comprehension, perhaps because this concept is not yet fully acquired by children of this age group. The researchers also suggested that morphophonological regularities are used before semantic cues in determining gender assignment. In addition, comprehension was measured via a production task and Mills (1986) maintained there may be a difference in the children's ability to comprehend the distinction between natural and syntactic gender and their ability to produce forms that reflect this distinction.

Some studies have also claimed that masculine anaphoric forms are acquired before feminine forms. Deutsch and Pechmann (1978) found that children produced the masculine pronoun "ihm" [him] before the feminine form "ihr" [her]. In addition, Scupin and Scupin (1910) and Stern and Stern (1928) reported the preferred use of "sein" [his/its] over "ihr" [her]. Mills (1986) asserted that in these studies the use of the masculine form dominated only when the
natural gender was the concept behind the linguistic form. She concluded, therefore, that there is no evidence that masculine forms are acquired first, although she conceded that they may be overgeneralized once they are acquired.

Obviously little is known about anaphoric comprehension in German. The few research studies that have been done have examined primarily children's understanding of anaphoric references. This research indicates that first language acquisition of anaphora in German is a complex process, complicated by the fact that both case and gender must be represented in anaphoric forms. Studies in the acquisition of gender indicate that children acquire an understanding of syntactic gender very early, perhaps via rote learning. The concept of natural gender, however, requires more time to be thoroughly acquired. Investigations in the acquisition of German relative clause constructions suggest that relative pronoun forms are acquired very gradually and that the word order of the clause may influence children's understanding more than either the case or gender of the relative pronoun. Clearly more first language research in the acquisition and comprehension of German anaphoric references is necessary to establish an adequate research base in this area. In particular, studies that investigate anaphoric forms other than relative pronouns and studies that include adult subjects are needed.

Summary of the Literature

Webber (1979) asserts that understanding anaphoric language requires two complementary tasks: "(1) identifying what a text potentially makes available for anaphoric reference and (2) constraining the candidate set of a
given anaphoric expression down to one possible choice” (p. vi). This chapter has reviewed several research studies that have investigated the various factors associated with these two tasks. In 1987, Just and Carpenter outlined a list of heuristic rules that readers employ when resolving anaphoric expressions. The list shown below is an updated and expanded version of the one provided by Just and Carpenter (1987). It essentially summarizes what is known about the discourse features that readers use to identify anaphoric forms and to constrain the set of candidate antecedents for a given anaphoric expression.

1. Look for a recently mentioned referent with the same gender and number as the anaphor (Chang, 1980; Corbett & Chang, 1983; Garnham & Oakhill, 1985; Gernsbacher, 1989).

2. If there are two possible candidate antecedents in a previous clause, favor the one that fulfills the same grammatical role in its clause as the anaphor does in its clause. This is known as the parallel function strategy (Grober, et al., 1978; Sheldon, 1974).

3. If there are two possible candidate antecedents in a previous clause, favor the one that is more thematically prominent, i.e., the one that is the topic of the discourse or is foregrounded (Clark & Sengul, 1979; Clifton & Ferreira, 1987; Frederikson, 1981; Kerr & Underwood, 1987; Kintsch & Vipond, 1978; Lesgold, et al, 1979; Roen, 1984; Walker & Yekovich, 1987).

4. If there are two possible candidate antecedents in some previous clauses, favor the antecedent in the most recent clause (Daneman &
Carpenter, 1980; Dell, et al., 1984; McKoon & Ratcliff, 1980).

5. Use the semantic properties of the verb to help determine which candidate is the more likely referent. This is known as implicit causality (Caramazza, et al., 1977; Garvey, et al., 1974; Garvey & Caramazza, 1974; Grober, et al., 1978).

6. Use general knowledge about the candidate antecedents in the referential situation to help determine which one is the more likely referent (Garnham, 1981; 1984; Garnham & Oakhill, 1985; Garrod & Sanford, 1977; Gernsbacher, 1989; Kerr & Underwood, 1985; Lesgold, 1974).

The studies presented above have focused primarily on the discourse features that readers use to interpret anaphoric structures. Precise mental processing of these discourse features is also necessary for anaphoric resolution to occur. Just and Carpenter (1987) note that only a few research studies have examined the cognitive processes involved in anaphoric comprehension. This research has revealed “some interesting performance characteristics associated with the processes themselves” (Just & Carpenter, 1987, 207). The following list summarizes what little is known about these cognitive processes:

1. Resolution begins when a reader first encounters an anaphor and continues while the anaphor is being fixated (Ehrlich, 1983; Ehrlich & Rayner, 1983; Vonk, 1984; 1985).
2. The duration of the resolution process is longer if the antecedent occurred much earlier in the discourse than if occurred only recently (Daneman & Carpenter, 1980; Ehrlich, 1983; Ehrlich & Rayner, 1983).

3. The duration of the resolution process is longer if no gender cue is present (Ehrlich, 1980; 1983; Vonk, 1984).

4. During the search for the antecedent of an anaphor, nonantecedents of the appropriate gender are also activated to some degree (Corbett & Chang, 1983; Dell, et al., 1984; McKoon & Ratcliff, 1980).

5. Spacing and word order features can influence antecedent assignment (Pynte, et al., 1988).

These studies have revealed much about the discourse features and processes utilized in anaphoric resolution and comprehension. There is, however, much that is still not known about how anaphora is acquired, used, and understood by both first and second language learners of English and other languages. It is quite apparent from this review of literature that additional research in this area is warranted. Of particular benefit would be studies that examine: (a) readers' cognitive processing of anaphoric structures, (b) their overall comprehension of macro-level discourse that contains anaphoric expressions, and (c) their comprehension of specific anaphoric forms within authentic macro-level discourse contexts.
CHAPTER III
DESIGN AND PROCEDURES

Population and Sample

The population from which the subjects were drawn consisted of native speakers of German and nonnative (American) readers of German at various levels of German language ability enrolled at The Ohio State University (OSU). An adult population of college students was chosen to avoid some of the problems encountered by first language researchers working with children. These studies revealed that school-aged children had not yet attained mastery in the comprehension and use of anaphoric structures in their native language (Bormuth et al, 1970; Garnham, et al., 1982; Lesgold, 1974; Richel, 1977). In addition, these students sometimes had difficulty understanding and completing the experimental task. Adult college students, on the other hand, have achieved an adequate understanding of anaphora in their first language to insure that this confounding variable would not affect the results. Moreover, these students were presumably mature enough to understand and perform the experimental task without difficulty.

All participants in the study were volunteers who were compensated ten dollars ($10.00) for their participation. The subjects were recruited via information fliers distributed in German classes and posted in the German Department (Appendix A). German instructors were requested in writing to
distribute the fliers in their classes (See Appendix B).

A total of 80 subjects participated in the study, but valid data were obtained from only 54 of them. The high rate of subject loss (n = 26) was due to several factors that affected the ability of the eye-tracking equipment to track the pupil reliably and, therefore, reduced the validity of the data for those subjects. These factors included: excessive head movement; heavy eyelids; long, dark eyelashes; squinting; and excessive blinking.

Research Design

Phase I of the Study

In Phase I of the study, Analysis of Covariance (ANCOVA) was used for statistical analysis to assess the effect of anaphora on the processing and comprehension of readers of German with various amounts of language ability. There were three independent variables examined in Phase I of this study. The measure of baseline German language ability was treated as a continuous, quantitative independent variable and was included as the covariate in the ANCOVA tests. The second independent variable, type of referencing device, was treated as a fixed, categorical variable with two levels: (a) authentic, representing an authentic text with anaphoric expressions; and (b) experimental, representing the same text except with the anaphoric expressions replaced in the text by their antecedent nouns and noun phrases. Text, the third independent variable, was treated as a random or generalization, categorical variable as described in Coleman (1964; 1973) and Clark (1973) and had two levels: (a) expository text and (b) literary text. Four quantitative dependent variables were evaluated in Phase I of this study: (a) fixation frequency, (b)
average fixation duration, (c) total fixation duration, and (d) comprehension of the text.

A 2 X 2 (text X type of referencing device) mixed ANCOVA was run separately for each of the four dependent variables in Phase I of the study. Because of the complexity of the research design, a BMDP statistical program that could accommodate mixed (fixed and random) effects, repeated measures, and one covariate was used for data analysis. This statistical program subjected the data to an ANCOVA procedure that tested the following model. This model attempted to account for all possible sources of variability in the subjects' dependent measures:

\[ Y_{ijk} = S_i + \mu + \alpha_j + \beta_k + \alpha \beta_{jk} + \Delta_1 X_i + \Delta_2 X_i(1) + \Sigma_{ijk} \]

where \( Y_{ijk} \) = the dependent measure, \( S_i \) = the subject effect, \( \mu \) = the constant or intercept, \( \alpha_j \) = the text effect, \( \beta_k \) = the type of referencing device effect, \( \alpha \beta_{jk} \) = the text X type of referencing device interaction, \( \Delta_1 X_i \) = the baseline German language ability effect, \( \Delta_2 X_i(1) \) = the baseline German language ability effect when no anaphora is present (a means of including the baseline X type of referencing device interaction where 1 is an indicator variable that is 0 when anaphora is present and 1 when anaphora is not present), and \( \Sigma_{ijk} \) = the error variance.

**Phase II of the Study**

In Phase II of the study, ANCOVA was also used for statistical analysis, to assess how readers of German process specific anaphoric expressions within
texts. In contrast to Phase I of the study, in which repeated measures and random variable effects had to be included in the ANCOVA tests, Phase II allowed for a less complex ANCOVA procedure to be used. There were two independent variables investigated in Phase II of this study. The covariate, baseline German language ability, was treated as a continuous, quantitative independent variable. Type of referencing device, the second independent variable, was treated as a fixed categorical variable with two levels: (a) authentic and (b) experimental.

Two quantitative dependent variables were evaluated in Phase II of the study: (a) average fixation duration in blocks and (b) percentage of total fixation time in blocks. The dependent variables were measured in specific blocks of a matrix mapped onto both versions of each text. Block size was determined by the size (length) of the anaphoric expressions in the authentic text. The repeated noun phrases in the experimental text were usually longer than the anaphoric expressions in the authentic text, and, therefore, took up more block space. To make the comparisons more accurate, fixation durations and times in blocks were averaged for the noun phrases in the authentic text. In other words, the fixation duration for a noun phrase that took up two blocks, e.g., “der Traktor” [the tractor], was averaged over the two blocks and compared to the fixation duration for the corresponding anaphoric expression, e.g., “er” [he/it], that took up only one block. Only those blocks containing anaphoric references in the authentic text version were compared with the corresponding blocks in the experimental text version containing repeated nouns or noun phrases.

Separate one-way ANCOVAs were run for each block of interest in each text for each dependent variable. A SAS statistical program that included fixed
effect variables and one covariate was used for data analysis.

Phase III of the Study

In Phase III of the study, the Pearson product-moment correlation procedure was used for data analysis. Three quantitative dependent variables were evaluated: (a) overall comprehension of the text, (b) coreferential tie comprehension, and (c) baseline German language ability. Subjects' coreferential tie errors were also analyzed qualitatively to determine how discourse features within the text, selection strategies, and baseline language ability impact antecedent selection.

Materials

Baseline German Language Ability Test

In order to assess each subject's baseline German language ability, all participants in the study took the 1990 reading portion of the American Association of Teachers of German (AATG) National German Examination for High School Students--Level IV. This exam is a discrete-point test designed to measure students' German language ability on vocabulary, grammar, and reading tasks. The Level IV version of the national exam was chosen because it tests all of the major grammatical features of the language and, therefore, provided a comprehensive assessment of the subjects' language ability with regard to grammatical features. The exam is a nationally recognized measure of German language competence. The 1990 Level IV version of the examination has a Kuder-Richardson 20 reliability coefficient of 0.95 for nonnative speakers of German and a coefficient of 0.94 for native speakers of
German. The subjects' scores on the exam are a measure of their baseline German language ability and were entered as quantitative variables in the data analysis.

**Texts**

**Phases I and II of the Study**

To enhance the generalizability of the results, two different German texts were used in Phases 1 and II of this study. Authentic texts containing anaphoric references that appeared in German readers were selected. The first text was a short expository paragraph entitled "Der mechanisierte Hof" [The Mechanized Farmyard] (Appendix C). The text was a slightly adapted excerpt from the book *Deutschland heute* [Germany Today] by Schulz and was presented as an authentic reading passage in the German textbook *How to read German: A short cut for non-linguists* (Law, 1964). This book has not been used for German instruction at OSU.

The paragraph was slightly abbreviated in order to fit the limited space available on the computer monitor. The modified text used in the study contained nine lines, including the title. In the excerpt, a farmer describes the mechanization of farms today and the changes that this brings to the farm. The text had a total of 80 words, seven of which were anaphoric references chosen for investigation. Three of the anaphoric expressions were gendered personal pronouns that referred to things or abstract nouns. One of these pronouns was plural and referred to "die wenigen Tage" [the few days]; the other two pronouns were singular and referred to "der Traktor" [the tractor] and "die Mechanisierung" [the mechanization] respectively. Four of the anaphoric
references were plural personal pronouns that referred to people and all had the same antecedent, "die Bauern" [farmers].

The second text, entitled "Im Volksgarten" [In the Park], was an excerpt from a short literary work by Altenberg (Appendix D). This text appeared in the German reader Texte und Übungen: Intermediate readings and exercises (Kunst, 1977). This book has not been included in the German curriculum at OSU.

This passage was also shortened and adapted slightly to conform to the restrictions of the computer monitor. The modified text used in the study contained ten lines, including the title. The excerpt is essentially a dialogue between a mother and a daughter. They are in a park and the daughter is flying a balloon. The text had a total of 82 words, ten of which were anaphoric expressions chosen for investigation. Six of the anaphoric references were gendered personal pronouns that referred to things. All of them were singular and referred to "der Ballon" [the balloon]. Three of the anaphoric expressions were singular, gendered personal pronouns that referred to people; two of them referred to the daughter, "Anna," and one of them referred to "dem armen Mädchen" [the poor girl] who was also in the park. There was also one singular demonstrative pronoun that referred to "diesen Ballon" [this balloon].

The texts "Der mechanisierte Hof" and "Im Volksgarten" were chosen because: (a) they are authentic, naturally-occurring passages; (b) they are texts used for reading instruction; (c) they contain a high density of anaphoric references despite their short length; and (d) their length could be accommodated by the limited space on the eye-tracking computer monitor.
For each text there were two text versions. The first version (the authentic version) contained nouns and noun phrases initially that were subsequently referred to in the text by appropriate anaphoric forms. The second version (the experimental version) also contained nouns and noun phrases initially, but they were subsequently referred to in the text by the same nouns and noun phrases. In all other respects the two versions of each text were identical. The text versions differed only in the type of referencing device used: anaphoric expressions for the authentic version and repeated nouns or noun phrases for the experimental version.

The experimental versions of both texts were created by replacing the naturally occurring pronoun references in the authentic text versions with the same nouns or noun phrases that appeared initially in those versions. All manipulations and adaptations were performed by the researcher. The slightly adapted and abbreviated authentic text versions and the experimental text versions of both texts were then reviewed for grammatical accuracy and authenticity by two native speakers of German.

The following excerpt from a study by Freebody and Anderson (1983) illustrates two contrasting versions of an English text created by this procedure.

Version of the text containing anaphoric expressions:

All countries have laws about how trade and business can be carried on with other countries. One of the oldest ways that governments control exchange is through a "tariff." This is most often a tax on goods coming into a country. It is added to their price and so makes them cost more.
Version of the text containing the exact same nouns and noun phrases:

All countries have laws about how trade and business can be carried on with other countries. One of the oldest ways that governments control trade with these laws is through a "tariff" law. The tariff is most often a tax on goods coming into a country. The tax is added to the price of the goods and so it makes the goods cost more. (p. 282).

Phase III of the Study

In Phase III of the experiment, a longer German text was used. The inclusion of a longer German text in the study was deemed especially important because of the limitations encountered in text selection for Phase I of the experiment. The size and space limitations of the computer display monitor attached to the eye-tracking equipment restricted the length of the texts in Phase I to a maximum of 11 lines or approximately 100 words. There was some concern that a text of this restricted length might not allow anaphora to operate completely and naturally as a discourse phenomenon. The analysis of data obtained with a longer German text was therefore included.

The longer text was an expository excerpt entitled "Europa heute" [Europe Today] from the book, Deutsche Geschichte des 19. und 20. Jahrhunderten [German History of the 19th and 20th Centuries] by Mann (1958). It appeared as a supplemental reading in the German textbook, How to read German: A short cut for non-linguists (Law, 1964). The text was 397 words long and described Europe's position in the world in 1958 from a historical perspective (Appendix E). This text was selected because: (a) it is an authentic, naturally-occurring text and (b) it contains a high density of anaphoric expressions.

There were 31 anaphoric expressions in the passage examined in the coreferent selection task (Appendix F). Among these anaphoric references, 12
were singular, gendered personal pronouns that referred to things. Two of the anaphors were plural, gendered personal pronouns that referred to things, and two were plural personal pronouns that referred to people. There was also one demonstrative noun phrase among the 31 anaphoric expressions. In addition, seven singular possessive adjectives and one plural possessive adjective were included in the task. Four relative pronouns, one demonstrative pronoun, and one indefinite pronoun were also investigated.

Two additional singular, gendered personal pronouns and two additional demonstrative pronouns were included in the coreferent task, but were not analyzed. These four anaphoric expressions were eliminated from the data analysis, because no consensus could be reached among the researcher, a highly proficient nonnative speaker, and a native speaker of German as to precisely what were the intended antecedents for these anaphors.

A high percentage of the anaphoric expressions tested (35%) referred to Europe, the text's central theme. The passage contained three gendered personal pronouns, one demonstrative noun phrase, and seven possessive adjectives, whose antecedent was Europe. The remaining anaphoric expressions referred to a variety of both macro-level and micro-level discourse concepts in the text.

Experimental Procedures

Data were collected individually from each subject using the eye-tracking equipment in the Foreign Language Research Laboratory located in Arps Hall at The Ohio State University. The data collection session included taking the National German Exam, reading the short texts while the eyes were tracked,
writing the recalls for the short texts, reading and recalling the longer text, and performing the coreferent selection task. This session lasted approximately two hours per subject.

When subjects arrived at the laboratory, they filled out a student information form, which included their name, address, telephone number, etc. (Appendix G). Then the researcher explained the agenda for the data collection session to them. They were told that first they would have 40 minutes to take an exam, which tests their German language ability. After they had finished the exam, there would be a brief demonstration on how the eye-tracking equipment functions. The subjects were then informed that they would read three short German texts on the computer monitor. The subjects were told that they could take as much time as they needed to read the texts and to read for comprehension. The subjects were also notified that their reading times would be recorded but were not of interest to the researcher and were noted only for computation purposes. Prior to reading the texts, the subjects were also informed that they would be asked to write down everything that they could remember about each text as soon as they finished reading it. They were also told that their recalls should be written in their first language. In addition, the subjects were notified that following the eye-tracking session they would be asked to read and recall a longer German text as well as to perform a brief “grammar-like” task with the text.

After these procedures and instructions were explained to the subjects, the reading comprehension section of the AATG National German Examination for High School Students-Level IV was administered to them. They had a total of 40 minutes to complete the test and this time limit was strictly observed.
When they had finished the exam, subjects were randomly assigned to read one version of each text. Subjects read the authentic version of one text and the experimental version of the other text. Two sets of data, therefore, were collected from each subject during this phase of the experiment, resulting in a repeated measures design. The order of the texts and the text versions were varied, to counterbalance order effects. There were four possible groups in which subjects can be placed: (a) Text 1-authentic version, Text 2-experimental version; (b) Text 1-experimental version, Text 2-authentic version; (c) Text 2-experimental version, Text 1-authentic version; (d) Text 2-authentic version, Text 1-experimental version.

Subjects were first asked to position themselves in front of the computer monitor, which was approximately 21 inches from their head. The subject's head was then stabilized in front of the monitor using a chin rest and head bar designed to minimize head movement. A low-intensity, infrared light source was positioned, so that it was reflected off of the cornea of the left eyeball. An infrared sensitive video camera was then adjusted to track the left eye, i.e. register the light that was reflected off of the left eye. The location of each reflection revealed where on the computer screen the subject was fixating. At this point subjects were reminded not to move their heads or close their eyes while reading, because the movement would produce inaccurate data.

The subjects were then asked to look at a mark (0) in the center of the computer screen. The subject's eye position at that moment was defined as the zero point. The subjects were also asked to look at nine numbers placed at regular intervals on the screen for a total of 25 seconds (Appendix H). The researcher said the numbers aloud and asked the subject to stare deliberately
at each number as it was called out. This procedure allowed eye fixations to be recorded for each number on the screen. The number patterns were used to calibrate the eye-tracking data for each individual subject.

The subjects then read a short practice text on the screen to familiarize them with the equipment and the experimental procedures (Appendix I). After reading the practice text, the subjects read the first text version corresponding to the group to which they were assigned. They were told to take as much time as they needed to read the text. Upon finishing the text, the subjects received unlined paper and written instructions telling them to write down, in their native language, everything they could remember about the text (Appendix J). They were given as much time as they needed to write their recalls. When the subjects completed their recalls, they were repositioned in front of the computer screen and asked to read the second text version for their group. When they finished reading, they again received paper and instructions to write down, in their native language, everything they could recall about the second text. While they were reading the eye-tracking computer monitored the subjects' eye movements and measured those movements according to the fixation position and the chronometric variables of fixation frequency and fixation duration.

After the data had been collected on the eye-tracking computer and the subjects had recalled the text, they were given a longer text to read. They were told to take as much time as they needed to read the text. When they were finished reading, they received paper and written instructions telling them to write down, in their native language, everything they could remember about the text. They were given as much time as they needed to complete their recalls. Once the subjects had completed their recall protocols, they were asked to
perform a coreferential location task using the same longer text that they had just read. The subjects received a copy of the longer text in which the anaphoric expressions of interest had been underlined. They were given written instructions telling them to write the correct referent for each underlined anaphoric expression in the space provided above the anaphoric expression. The subjects were first provided with a few examples and practice sentences to ensure that they understood the task correctly (Appendix K).

Data Collection and Instrumentation

A Micromeasurements System 1200 Eye Monitor interfaced with an IBM PCXT personal computer was used for data collection. The System 1200 is a video-based eye-monitoring system controlled by a microcomputer. The system is designed to measure the pupil area as well as horizontal and vertical eye movements. Sixty times a second, a digitized signal sent information about the position of the subject's eye to the IBM computer where the data was stored. Eye positions were recorded as ordered pairs, with the abscissa indicating the horizontal position of the eye and the ordinate indicating the vertical position. The subjects' eye-movement parameters were quantified by a specialized computer program. The data collected for each subject was stored on computer diskettes and used for further analyses.

The System 1200 was used to measure the following dependent variables:

1. Fixation frequency--the number of times that the eye stops to fixate when reading the text;

2. Average fixation duration, or the average time per fixation, measured in milliseconds (msec.), that the eye stops to gather information. A pause
of 50 milliseconds or longer with a saccade on either side of the pause constitutes a fixation;

3. Total fixation duration, or the total amount of time the eye spends fixating the text, measured in seconds (sec.);

4. Average fixation duration in blocks, or the average time, in milliseconds, spent fixating information within a given region of the text containing either a noun/noun phrase or an anaphoric expression;

5. Percentage of total fixation time in blocks or the total amount of time spent fixating within a given region of the text containing either a noun/noun phrase or an anaphoric expression divided by the total fixation time.

In order to determine what information subjects had understood about the text, the dependent variable of reading comprehension was measured via immediate written recall protocols. This measure of reading comprehension was chosen because it is considered a direct reflection of what the reader has actually processed and understood about the text (Bernhardt & James, 1987). Johnston (1983) calls the recall protocol "the most straightforward assessment of the result of the text-reader interaction" (p. 54). According to Bernhardt and James (1987), the value of this measurement device lies in its ability "to determine the extent to which adequate and accurate comprehension is occurring" (p. 71). Recalls were written in the subject's native language (either English or German) to avoid the confounding effects of a second language (Lee, 1986). Bernhardt (1986) also asserts that if the target language is used: (a) those who score the recalls may become distracted by the grammatical
errors and focus less on the student's actual comprehension; and (b) students will attend to grammar, vocabulary, and spelling in the target language and may not recall as much information.

A weighted propositional scoring instrument was developed for each text according to the procedure described by Bernhardt (1988). The texts were first divided into pausal units by two native readers and two highly proficient nonnative readers of German. This was accomplished by having the raters read each text to themselves and to mark all those places in the text where they paused. Pausal units generally came at the end of syntactically related sentence units, as in the following example.

"Ich möchte / einen blauen Ballon / haben!" / sagte Anna.
[I want / a blue balloon / (to have)! / says Anna.]

The raters were then asked to divide the pausal units or text propositions equally among four levels: the lowest level being the least important 25% of the propositions in the text, the next level being the next least important 25% of the propositions, and so forth, ending with the most important 25% of the text. The lowest 25% of the propositions received a value of 1, the next 25% a value of 2, the next highest 25% a value of 3, and the highest or most important 25% of the propositions a value of 4. Bernhardt (1988) maintains that the level 4 information represents the thread of the discourse in a telegraphic style. The raters' individual pausal units and the point values for those units were averaged to produce the final scoring instrument for each text. In conflicting instances, the shorter pausal units were selected for the scoring instruments, as
recommended by Bernhardt (1988).

The created scoring instruments were then compared to the subjects' written protocols and points were awarded for each pausal unit recalled correctly. The total number of points obtained for a text constituted the subject's comprehension score for that text. Subjects' scores on the recalls depended on the importance of the pausal units they remembered. A subject recalling many units with a value of 4, therefore, earned a higher recall score than a subject recalling primarily units with a value of 2.

The researcher scored all of the recalls twice. Intra-rater reliability was established via Pearson product-moment correlations. For both the expository text, "Der mechanisierte Hof," and the literary text, "Im Volksgarten," a correlation of 0.9988 was obtained. A correlation of 0.9984 between the first and second sets of scores was obtained for the longer German text, "Europa heute."

To insure inter-rater reliability a highly proficient reader of German was trained to use the scoring instruments. This independent rater scored a randomly selected sample of 25% of the recalls collected for each text. In cases of great disparity between the researcher's score and that of the independent rater, the recalls in question were reevaluated and areas of conflict were mutually resolved. Pearson product-moment correlations between the researcher's scores and the scores of the independent rater were 0.9890 for "Der mechanisierte Hof," 0.9942 for "Im Volksgarten," and 0.9977 for "Europa heute." Since all of these correlations were very strong, one scorer was considered sufficient to score the majority of the recalls for each text.

The researcher consulted a native speaker of German to create the scoring instrument for the coreferent selection task. This precaution insured that all
syntactically and semantically appropriate coreferents were considered when scoring the subjects' responses. The total number of correctly identified referents for the 31 underlined anaphoric expressions comprised the subject's score for the coreferent selection task.

Data Analysis

All of the dependent variables were analyzed separately using the statistical procedures outlined in the Research Design section. The following null hypotheses were tested.

Phase I of the Study

\( H_01: \) (A) There is no significant difference due to type of referencing device as measured by fixation frequency.
(B) There is no significant difference due to baseline German language ability as measured by fixation frequency.
(C) There is no significant difference due to text as measured by fixation frequency.
(D) There is no significant interaction between type of referencing device and baseline German language ability as measured by fixation frequency.
(E) There is no significant interaction between type of referencing device and text as measured by fixation frequency.

\( H_02: \) (A) There is no significant difference due to type of referencing device as measured by average fixation duration.
(B) There is no significant difference due to baseline German language ability as measured by average fixation duration.

(C) There is no significant difference due to text as measured by average fixation duration.

(D) There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration.

(E) There is no significant interaction between type of referencing device and text as measured by average fixation duration.

$H_0^3$: (A) There is no significant difference due to type of referencing device as measured by total fixation duration.

(B) There is no significant difference due to baseline German language ability as measured by total fixation duration.

(C) There is no significant difference due to text as measured by total fixation duration.

(D) There is no significant interaction between type of referencing device and baseline German language ability as measured by total fixation duration.

(E) There is no significant interaction between type of referencing device and text as measured by total fixation duration.

$H_0^4$: (A) There is no significant difference due to type of referencing device as measured by recall protocol score.
(B) There is no significant difference due to baseline German language ability as measured by recall protocol score.
(C) There is no significant difference due to text as measured by recall protocol score.
(D) There is no significant interaction between type of referencing device and baseline German language ability as measured by recall protocol score.
(E) There is no significant interaction between type of referencing device and text as measured by recall protocol score.

**Phase II of the Study**

\(H_5\): (A) There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in blocks.
(B) There is no significant difference due to type of referencing device as measured by average fixation duration in blocks.
(C) There is no significant difference due to baseline German language ability as measured by average fixation duration in blocks.

\(H_6\): (A) There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in blocks.
(B) There is no significant difference due to type of referencing device as measured by percentage of total fixation time in blocks.
(C) There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in blocks.

**Phase III of the Study**

H₀ 7:  (A) There is no significant relationship between coreferential tie comprehension and overall comprehension.
(B) There is no significant relationship between coreferential tie comprehension and baseline German language ability.
(C) There is not significant relationship between overall comprehension and baseline German language ability.

**Pilot Study**

In order to evaluate the efficacy of the experimental procedures and materials, a pilot study was conducted. The subjects were drawn from the same population that was used for the actual experiment, i.e., students at The Ohio State University. One of the major purposes of the pilot study was to determine the suitability of the materials, in particular the selection of the texts and the lay-out of the shorter texts on the computer screen. Also of concern to the researcher was that the instructions (both written and verbal) were clear to the students and that they were capable of performing the various experimental procedures without difficulty.

The results of the pilot study indicated that all of the materials were suitable for the experiment. In addition, the shorter texts were found to be positioned on the computer monitor in such a way that they could be clearly and easily seen.
The pilot study also revealed that the subjects clearly understood the written and verbal instructions given to them and that they were able to perform all of the experimental tasks adequately. No adjustments in the materials or the experimental procedures were therefore deemed necessary.
CHAPTER IV
ANALYSIS OF THE DATA

Introduction

Research evidence suggests that native readers of English sometimes have difficulty interpreting anaphoric relationships correctly and that this difficulty can impair their comprehension of texts (Bormuth, et al., 1970; Lesgold, 1974; Richek, 1977; Walker & Yekovich, 1987). Additional studies indicate that comprehension improves when anaphoric connections are made more explicit (Corbett & Chang, 1983; Gemsbacher, 1989; Richek, 1977). The present study investigated the processing and comprehension of anaphoric expressions in German text by readers at various levels of baseline German ability. This research had three specific purposes: (a) to examine the effect of anaphora on the general cognitive processing behaviors and reading comprehension of readers of German at various levels of baseline German language ability, (b) to assess how readers process specific anaphoric references within German texts, and (c) to determine if readers of German are able to identify correctly the coreferents of various anaphoric expressions in a German text and if this ability is related to their overall comprehension of the text and their baseline German language ability. This research also sought to provide general data about the text processing behaviors of American readers of German.
Phase I of the Study

To assess the effect of anaphora on the processing and comprehension of readers of German with varying amounts of language ability, Analysis of Covariance (ANCOVA) was used for statistical analysis in Phase I of the study. There were three independent variables investigated in this phase of the study. The first independent variable, baseline German language ability, was treated as a continuous, quantitative variable and was included as the covariate in the ANCOVA tests. Type of referencing device, the second independent variable, was treated as a fixed, categorical variable with two levels: (a) authentic, representing an authentic text with anaphoric expressions; and (b) experimental, representing the same text except with the anaphoric expressions replaced by their antecedent nouns and noun phrases. The third independent variable, text, was treated as a random or generalization (Coleman, 1964; 1973; Clark, 1973), categorical variable with two levels: (a) expository text and (b) literary text.

The four quantitative, dependent variables measured in Phase I of this study were: (a) fixation frequency, (b) average fixation duration, (c) total fixation duration, and (d) reading comprehension. The first three dependent variables were measured with a Micromeasurements System 1200 Eye Monitor interfaced with an IBM PCXT personal computer. The fourth variable, reading comprehension, was assessed through immediate written recall protocols performed by subjects after reading the texts.

Separate 2 X 2 (Text X Type of Referencing Device) mixed ANCOVA tests were run for each of the four dependent variables in Phase I of the study. Due to the complexity of the research design, a BMDP statistical program that could
accommodate mixed (fixed and random) effects, repeated measures, and one covariate was used for data analysis. This statistical program subjected the data to an ANCOVA procedure that tested the following model. This model attempted to account for all possible sources of variability in the subjects' dependent measures:

\[ Y_{ijk} = S_i + \mu + \alpha_j + \beta_k + \alpha \beta_{jk} + \Delta_1 X_i + \Delta_2 X_i(l) + \Sigma_{ijk} \]

where \( Y_{ijk} \) = the dependent measure, \( S_i \) = the subject effect, \( \mu \) = the constant or intercept, \( \alpha_j \) = the text effect, \( \beta_k \) = the type of referencing device effect, \( \alpha \beta_{jk} \) = the text X type of referencing device interaction, \( \Delta_1 X_i \) = the baseline German language ability effect, \( \Delta_2 X_i(l) \) = the baseline German language ability effect when no anaphora is present (a means of including the baseline X type of referencing device interaction where \( l \) is an indicator variable that is 0 when anaphora is present and 1 when anaphora is not present), and \( \Sigma_{ijk} \) = the error variance.

The BMDP statistical program generates parameter estimates for all the terms in the model and computes significance levels for the covariate and the fixed effect variables. The parameter estimates for the model are interpreted differently depending on whether the term is a covariate, a random or a fixed effect. The parameter estimate for the covariate is a regression coefficient that represents the slope of the regression line for the covariate when all other effects are held constant. For random effects, the parameter estimate is an estimate of the population variance due to that particular random effect.
parameter estimates for fixed effects are estimates of the coefficients of the dummy variables the statistical program has created to test those effects. These estimates are regression coefficients and can be used to predict the expected difference in the means between the levels of the fixed effect variables.

In order to assess the interaction effect between the covariate, baseline German language ability, and the type of referencing device an indicator variable (I) was added to the model. When anaphora was present in the text (the authentic level of the type of referencing device variable) the indicator variable was zero (I = 0), producing the following model.

\[ Y_{ijk} = S_i + \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + \Delta_1 X_i + \xi_{ijk} \]

When anaphora was not present in the text (the experimental level of the type of referencing device variable), the indicator variable was one (I = 1) and the following model was obtained.

\[ Y_{ijk} = S_i + \mu + \alpha_j + \beta_k + \alpha\beta_{jk} + (\Delta_1 + \Delta_2) X_i + \xi_{ijk} \]

If \( \Delta_2 = 0 \) or if \((\Delta_1 + \Delta_2)\) was not significantly different from \(\Delta_1\), the effect due to baseline German language ability was the same or not significantly different for both levels of type of referencing device and no significant interaction was present. If \( \Delta_2 \neq 0 \) and \((\Delta_1 + \Delta_2)\) was significantly different from \(\Delta_1\), the effect due to baseline German language ability was significantly different for the two levels of type of referencing device and a significant interaction was present.
The full model was tested using restricted maximum likelihood to obtain parameter estimates for each of the terms in the model. Restricted maximum likelihood provides the best estimates for the parameters, but only tests the significance of fixed effect variables. For this reason, maximum likelihood was used to test the significance of random effects. The only random effects of interest were text and the type of referencing device X text interaction. To test the significance of the type of referencing device X text interaction the full model was compared with the following model in which the type of referencing device X text interaction ($\alpha \beta_{jk}$) had been removed.

$$Y_{ijk} = S_i + \mu + \alpha_j + \beta_k + \Delta_1X_i + \Delta_2X_i(l) + \Sigma_{ijk}$$

In order to assess the effect due solely to text, the text X type of referencing device interaction ($\alpha \beta_{jk}$) must be eliminated from the full model as shown in the model directly above. This modified model was then compared with the following model in which the text effect ($\beta_k$) had been removed.

$$Y_{ijk} = S_i + \mu + \alpha_j + \Delta_1X_i + \Delta_2X_i(l) + \Sigma_{ijk}$$

This analysis is done using a likelihood ratio test. The results are reported as a test statistic with a Chi-square distribution. The significance test for this statistic reveals whether the random effect due to text is significant or not.

The results of the ANCOVA tests for each of the four dependent variables examined in Phase I of the study are reported separately in the following
sections.

**ANCOVA Results for Fixation Frequency**

The ANCOVA test for the dependent variable of fixation frequency produced the observed and predicted means shown in Table 1 and the parameter estimates listed in Table 2.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
<th>Predicted Mean</th>
<th>Predicted SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic</td>
<td>186.2037</td>
<td>71.1476</td>
<td>190.1246</td>
<td>25.4648</td>
</tr>
<tr>
<td>Experimental</td>
<td>191.7592</td>
<td>80.4952</td>
<td>187.7855</td>
<td>25.4648</td>
</tr>
</tbody>
</table>

N = 108, cell size = 54

Hypothesis 1A: **There is no significant difference due to type of referencing device as measured by fixation frequency.** An examination of Table 1 reveals that the average observed fixation frequency was 186.2037 fixations for the text with the authentic type of referencing device and was 191.7592 fixations for the text with the experimental type of referencing device. The ANCOVA failed to detect a significant difference between these two groups, p > 0.926 (Table 2). The null hypothesis of no difference between the two types of referencing
devices was, therefore, retained.

A review of the predicted means for type of referencing device in Table 1 reveals a fixation frequency pattern different from the one obtained for the observed means. The predicted means are generated by the model being tested and account for all sources of variability in the model. They are, therefore, better estimates of the actual means for the population than are the observed sample means. The means predicted by the model indicate that on the average subjects needed slightly more fixations to process a text with anaphoric expressions (190.1246) than a text containing repeated nouns and noun phrases (187.7855). Although this finding is not statistically significant, it implies that greater processing demands are being placed on the reader when interacting with texts that contain anaphoric references.

Hypothesis 1B: There is no significant difference due to baseline German language ability as measured by fixation frequency. Table 2 reveals significant differences among subjects’ fixation frequencies due to their baseline German language ability and, thus, led to a rejection of the null hypothesis, \( p < 0.001 \). These findings demonstrate that subjects’ baseline German language ability does impact the frequency with which they fixate the text. The estimated coefficient, -1.828, produced by the ANCOVA test indicates that, holding all other factors in the model constant, an increase of one point in baseline German language ability score would yield an average decrease of 1.828 fixations in fixation frequency. In practical terms, this result means that a subject with a very high baseline German language ability score of 60 was predicted to need 73.12 fewer fixations to process these texts than a subject with a relatively
low baseline German language ability score of 20.

Hypothesis 1C: There is no significant difference due to text as measured by fixation frequency. The results of the likelihood ratio test to determine the effect of text on fixation frequency revealed that the two texts were significantly different from each other, \( p < 0.000 \). The null hypothesis of no effect due to text was, therefore, rejected. It can be concluded that subjects’ fixation frequencies vary significantly from text to text and that the estimate of that variance for the population of texts is 847.781 (Table 2).

Hypothesis 1D: There is no significant interaction between type of referencing device and baseline German language ability as measured by fixation frequency. The null hypothesis of no significant interaction between type of referencing device and subjects’ baseline German language ability was retained for fixation frequency, \( p > 0.832 \) (Table 2).

Hypothesis 1E: There is no significant interaction between type of referencing device and text as measured by fixation frequency. The likelihood ratio test revealed no significant interaction between type of referencing device and text as measured by fixation frequency, \( p > 0.983 \). This result led to a retention of the null hypothesis of no interaction.
Table 2
Analysis of Covariance Parameter Estimates
for Fixation Frequency

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject*</td>
<td>2327.220</td>
<td>718.830</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>261.845</td>
<td>29.997</td>
<td>0.000</td>
</tr>
<tr>
<td>Text*</td>
<td>847.781</td>
<td>1281.087</td>
<td></td>
</tr>
<tr>
<td>Type of Referencing Device</td>
<td>1.170</td>
<td>12.558</td>
<td>0.926</td>
</tr>
<tr>
<td>Text X Type of Referencing Device*</td>
<td>11.717</td>
<td>384.318</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>-1.828</td>
<td>0.565</td>
<td>0.001</td>
</tr>
<tr>
<td>Baseline X Type of Referencing Device</td>
<td>0.119</td>
<td>0.561</td>
<td>0.832</td>
</tr>
<tr>
<td>Error*</td>
<td>2248.433</td>
<td>445.256</td>
<td></td>
</tr>
</tbody>
</table>

* = random effects
ANCova Results for Average Fixation Duration

The ANCOVA test for the dependent variable of average fixation duration produced the observed and predicted means shown in Table 3 and the parameter estimates listed in Table 4.

Table 3
Observed and Predicted Means and Standard Deviations of Average Fixation Duration as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
<th>Predicted Mean</th>
<th>Predicted SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic</td>
<td>0.2732</td>
<td>0.0732</td>
<td>0.2616</td>
<td>0.0132</td>
</tr>
<tr>
<td>Experimental</td>
<td>0.2678</td>
<td>0.0729</td>
<td>0.2793</td>
<td>0.0132</td>
</tr>
</tbody>
</table>

N = 108, cell size = 54

Hypothesis 2A: **There is no significant difference due to type of referencing device as measured by average fixation duration.** Table 3 shows that the observed average fixation duration was 0.2732 msec. for the text with the authentic type of referencing device and was 0.2677 msec. for the text with the experimental type of referencing device. This difference was not found to be significant by the ANCOVA, p > .0206 (Table 4), thus leading to a retention of the null hypothesis.

A review of the predicted means for type of referencing device in Table 3 reveals a pattern for average fixation duration different from the one obtained for the observed means. The means predicted by the model indicate that, on the
average, subjects’ predicted average fixation durations were slightly lower (0.2616 msec.) for text versions that contained anaphoric expressions than for text versions that contained repeated nouns and noun phrases (0.2793 msec.). Although this finding is not statistically significant, it suggests that fewer processing demands are being placed on the reader when interacting with texts that contain anaphoric references.

Hypothesis 2B: **There is no significant difference due to baseline German language ability as measured by average fixation duration.** An examination of Table 4 reveals that subjects’ average fixation durations differed significantly according to their baseline German language ability. This finding led to a rejection of the null hypothesis, \( p < 0.048 \). These results indicate that subjects’ baseline German language ability does influence their average fixation times. The ANCOVA produced an estimated coefficient of -0.001, which indicates that, holding all other factors in the model constant, an increase of one point in baseline score would yield an average decrease of 0.001 msec. in average fixation durations. In practical terms, this result means that a subject with a very high baseline German language ability score of 60 was predicted to have an average fixation duration that was 0.04 msec. shorter than the average fixation duration of a subject with a relatively low baseline German language ability score of 20.

Hypothesis 2C: **There is no significant difference due to text as measured by average fixation duration.** The likelihood ratio test to determine the effect of text on average fixation duration indicated that the the two texts were significantly
different from each other, $p < 0.044$. The null hypothesis of no effect due to text was, therefore, rejected. These results show that subjects’ average fixation durations differ significantly from text to text and that the estimate of the variance for the population of texts is between 0.0005 and 0.000 (Table 4).

Hypothesis 2D: **There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration.** The ANCOVA revealed no significant interaction between type of referencing device and subjects’ baseline German language ability as measured by average fixation duration, $p > 0.064$ (Table 4). This result led to a retention of the null hypothesis of no interaction.

Hypothesis 2E: **There is no significant interaction between type of referencing device and text as measured by average fixation duration.** The results of the likelihood ratio test revealed no significant interaction between type of referencing device and text for average fixation duration, $p > 1.000$. Thus, the null hypothesis of no interaction was retained.
Table 4  
Analysis of Covariance Parameter Estimates  
for Average Fixation Duration

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject*</td>
<td>0.004</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.330</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>Text*</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Type of Referencing Device*</td>
<td>-0.009</td>
<td>0.007</td>
<td>0.206</td>
</tr>
<tr>
<td>Text X Type of Referencing Device*</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.048</td>
</tr>
<tr>
<td>Baseline X Type of Referencing Device</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.064</td>
</tr>
<tr>
<td>Error*</td>
<td>0.001</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

* = random effects
**ANCova Results for Total Fixation Duration**

The ANCOVA test for the dependent variable of total fixation duration produced the observed and predicted means shown in Table 5 and the parameter estimates listed in Table 6.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
<th>Predicted Mean</th>
<th>Predicted SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic</td>
<td>50.3367</td>
<td>21.8067</td>
<td>49.9943</td>
<td>8.5996</td>
</tr>
<tr>
<td>Experimental</td>
<td>51.6546</td>
<td>25.2856</td>
<td>51.5577</td>
<td>8.5996</td>
</tr>
</tbody>
</table>

N = 108, cell size = 54

Hypothesis 3A: **There is no significant difference due to type of referencing device as measured by total fixation duration.** The means for total fixation duration according to type of referencing device are reported in Table 5. The observed mean total fixation duration was 50.3367 sec. for the text with the authentic type of referencing device and was 51.6546 sec. for the text with the experimental type of referencing device. The ANCOVA failed to detect a significant difference between these two types of referencing devices, \( p > 0.870 \) (Table 6). The null hypothesis of no difference between the two groups was, therefore, retained.
A review of the predicted means for type of referencing device in Table 5 reveals a pattern for total fixation duration similar to the one obtained for the observed means. The means predicted by the model reveal that on the average subjects had slightly shorter total fixation durations (49.9943 sec.) for texts that contain anaphoric expressions than for texts containing repeated nouns and noun phrases (51.5577 sec.). Although this finding is not statistically significant, it implies that fewer processing demands are being placed on the reader when interacting with texts that contain anaphoric references.

Hypothesis 3B: There is no significant difference due to baseline German language ability as measured by total fixation duration. Table 6 reveals significant differences among subjects’ total fixation durations as a function of their baseline German language ability. This finding led to a rejection of the null hypothesis, \( p < 0.000 \). These results demonstrate that subjects’ baseline German language ability does impact the the total amount of time they spend fixating the text. The estimated coefficient, -0.682, produced by the ANCOVA test indicates that, holding all other factors in the model constant, an increase of one point in baseline score would yield an average decrease of 0.682 sec. in total fixation duration for the text. In practical terms, this result means that a subject with a very high baseline German language ability score of 60 was predicted to have a total fixation duration that was 27.28 sec. shorter than the total fixation duration of a subject with a relatively low baseline German language ability score of 20.
Hypothesis 3C: There is no significant difference due to text as measured by total fixation duration. The results of the likelihood ratio test to determine the effect of text on fixation frequency revealed that the two texts were significantly different from each other, $p < 0.000$. The null hypothesis of no effect due to text was, therefore, rejected. It can be concluded that subjects’ total fixation durations vary significantly from text to text and that the estimate of that variance for the population of texts is 65.865 (Table 6).

Hypothesis 3D: There is no significant interaction between type of referencing device and baseline German language ability as measured by total fixation duration. The null hypothesis of no significant interaction between type of referencing device and subjects’ baseline German language ability was retained for total fixation duration, $p > 0.836$ (Table 6).

Hypothesis 3E: There is no significant interaction between type of referencing device and text as measured by total fixation duration. The likelihood ratio test revealed no significant interaction between type of referencing device and text as measured by total fixation duration, $p > 0.992$. This result led to a retention of the null hypothesis of no interaction.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject*</td>
<td>222.280</td>
<td>60.531</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>79.494</td>
<td>9.292</td>
<td>0.000</td>
</tr>
<tr>
<td>Text*</td>
<td>65.865</td>
<td>142.449</td>
<td></td>
</tr>
<tr>
<td>Type of Referencing Device</td>
<td>-0.782</td>
<td>4.782</td>
<td>0.870</td>
</tr>
<tr>
<td>Text X Type of Referencing Device*</td>
<td>50.575</td>
<td>102.919</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>-0.682</td>
<td>0.161</td>
<td>0.000</td>
</tr>
<tr>
<td>Baseline X Type of Referencing Device</td>
<td>-0.030</td>
<td>0.144</td>
<td>0.836</td>
</tr>
<tr>
<td>Error*</td>
<td>148.474</td>
<td>29.402</td>
<td></td>
</tr>
</tbody>
</table>

* = random effects
ANCova Results for Reading Comprehension

The ANCOVA test for the dependent variable of reading comprehension as measured by written recall protocol scores produced the observed and predicted means shown in Table 7 and the parameter estimates listed in Table 8.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
<th>Predicted Mean</th>
<th>Predicted SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic</td>
<td>53.2037</td>
<td>28.1354</td>
<td>49.4839</td>
<td>13.9360</td>
</tr>
<tr>
<td>Experimental</td>
<td>51.7778</td>
<td>25.9315</td>
<td>55.4976</td>
<td>13.9360</td>
</tr>
</tbody>
</table>

N = 108, cell size = 54

Hypothesis 4A: There is no significant difference due to type of referencing device as measured by recall protocol score. Table 7 shows that the observed average recall protocol score was 53.2037 for the text with the authentic type of referencing device and was 51.7778 for the text with the experimental type of referencing device. This difference was not found to be significant by the ANCOVA, p > .0.354 (Table 8), thus leading to a retention of the null hypothesis of no difference between the two types of referencing devices.

A review of the predicted means for type of referencing device in Table 7 reveals a reading comprehension pattern different from the one obtained for the
observed means. The means predicted by the model reveal that on the average subjects’ predicted recall scores were lower (49.4389) for texts that contain anaphoric expressions than for texts that contain repeated nouns and noun phrases (55.4976). Although this finding is not statistically significant, it suggests that greater comprehension demands are being placed on the reader when interacting with texts that contain anaphoric references.

Hypothesis 4B: There is no significant difference due to baseline German language ability as measured by recall protocol score. An examination of Table 8 shows that subjects’ recall protocol scores differed significantly according to their baseline German language ability. This finding led to a rejection of the null hypothesis, p < 0.000. These results indicate that subjects’ baseline German language ability is a factor in their recall protocol scores. The estimated coefficient of 1.036 indicates that, holding all other factors in the model constant, an increase of one point in baseline German language ability score would yield an average increase of 1.036 points in recall protocol score. In practical terms, this result means that a subject with a very high baseline German language ability score of 60 was predicted to have a recall protocol score that was 41.44 points higher than the recall protocol score of a subject with a relatively low baseline German language ability score of 20.

Hypothesis 4C: There is no significant difference due to text as measured by recall protocol score. The likelihood ratio test to determine the effect of text on recall protocol score indicated that the two texts were significantly different from each other, p < 0.000. The null hypothesis of no effect due to text was,
therefore, rejected. These results show that subjects’ recall protocol scores differ significantly from text to text and that the estimate of that variance for the population of texts is 359.003 (Table 8).

Hypothesis 4D: There is no significant interaction between type of referencing device and baseline German language ability as measured by recall protocol score. The ANCOVA test failed to detect a significant interaction between type of referencing device and subjects’ baseline German language ability as measured by recall protocol score, \( p > 0.365 \) (Table 8). This result led to a retention of the null hypothesis of no interaction.

Hypothesis 4E: There is no significant interaction between type of referencing device and text as measured by recall protocol score. The results of the likelihood ratio test revealed no significant interaction between type of referencing device and text for recall protocol score, \( p > 0.983 \). Thus, the null hypothesis of no interaction was retained.
Table 8
Analysis of Covariance Parameter Estimates
for Recall Protocol Scores

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SD</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject*</td>
<td>148.570</td>
<td>46.679</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>12.523</td>
<td>14.503</td>
<td>0.388</td>
</tr>
<tr>
<td>Text*</td>
<td>359.003</td>
<td>511.748</td>
<td></td>
</tr>
<tr>
<td>Type of Referencing Device</td>
<td>-3.007</td>
<td>3.247</td>
<td>0.354</td>
</tr>
<tr>
<td>Text X Type of Referencing Device*</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>1.036</td>
<td>0.145</td>
<td>0.000</td>
</tr>
<tr>
<td>Baseline X Type of Referencing Device</td>
<td>-0.133</td>
<td>0.146</td>
<td>0.365</td>
</tr>
<tr>
<td>Error*</td>
<td>153.093</td>
<td>30.317</td>
<td></td>
</tr>
</tbody>
</table>

* = random effects
Phase II of the Study

The results of the four dependent variables in Phase I of the study measured subjects’ performance on the two texts as a whole and, as such, were global indices of their processing and comprehension abilities. In order to assess how readers of German process specific anaphoric references within texts, subjects’ fixation behaviors for blocks of text containing either anaphoric references or substituted nouns or noun phrases were subsequently analyzed. Analysis of Covariance (ANCOVA) was also used for statistical analysis in Phase II of the study. In contrast to Phase I of the study, in which repeated measures and random variable effects had to be accounted for in the ANCOVA tests, Phase II allowed for a less complex ANCOVA procedure to be used. The repeated measures factor and the random effect due to text were eliminated from the design in Phase II, because each text had to be analyzed independently. The covariate, baseline German language ability, was treated as a continuous, quantitative variable. Type of referencing device was treated as a fixed, categorical, independent variable with two levels: (a) authentic and (b) experimental.

Two quantitative, dependent variables were investigated: (a) average fixation duration in blocks and (b) percentage of total fixation time in blocks. The values of these two dependent variables were obtained with a Micromeasurements System 1200 Eye Monitor interfaced with an IBM PCXT personal computer. The dependent variables were measured in specific blocks of a matrix mapped onto both versions of each text (Appendix L and M). Block size was determined by the size (length) of the anaphoric expressions in the authentic text. Only those blocks containing anaphoric references in the
authentic text versions were compared with the corresponding blocks in the experimental text versions containing substitute nouns or noun phrases.

Separate one-way ANCOVA tests were run for each block of interest in each text for each dependent variable. A SAS statistical program that could accommodate fixed effect variables and one covariate was used to test for main and interaction effects. This statistical program generates F-test values and computes their significance levels. These statistics are reported in Phase II of the study.

The two dependent variables, average fixation duration in blocks and percentage of total fixation time in blocks, are presented in separate sections. Within those sections the findings for each text and for each block of interest are reported separately, beginning with a report on the tests of significance for the Type of Referencing Device X Baseline German Language Ability interaction and followed by a report on the tests for the main effects of type of referencing device and baseline German language ability.

**ANCOVA Results for Average Fixation Duration in Blocks**

In this section the average time that subjects spent fixating a block of the authentic text version containing an anaphoric expression will be compared to the average time that they spent fixating a block of the experimental text version containing a noun or noun phrase.

**Expository Text: “Der Mechanisierte Hof”**

Copies of both the authentic and experimental versions of the expository text, “Der mechanisierte Hof,” with the block matrix mapped onto the text are
provided in Appendix L. Only those blocks containing anaphoric expressions or substitute nouns or noun phrases were selected for further investigation. In the text, "Der mechanisierte Hof," seven blocks containing anaphoric references were chosen for analysis. The average fixation durations for each block of interest are examined independently in the following section.

Hypothesis 5A (Block 1): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 1. In Block 1 of the text, "Der mechanisierte Hof," the nominative case plural personal pronoun "sie" [they] in the authentic text version was compared to the noun phrase "die Maschinen" [the machines] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was rejected for average fixation duration in Block 1, $F(1, 50) = 7.22, p < 0.0098$ (Table 10).

Figure 1 portrays the nature of the interaction between type of referencing device and baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This figure reveals that the predicted average fixation durations for Block 1 of the authentic text version increased slightly as baseline German language ability scores increased. This result indicates that low ability students required slightly shorter average fixation durations to process the less explicit pronoun reference than did students with high baseline German language skills. In contrast, the results for the experimental text version, containing the repeated noun phrase "die Maschinen" [the machines], reveal that predicted average fixation durations decreased sharply as students' baseline German language ability scores
increased. This finding suggests that the repeated noun phrase was processed faster by high ability subjects than by subjects with low baseline German language skills.

Table 9
Means and Standard Deviations of Average Fixation Duration in Block 1 of "Der Mechanisierte Hof" as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic^1</td>
<td>0.2529</td>
<td>0.1559</td>
</tr>
<tr>
<td>Experimental^2</td>
<td>0.2746</td>
<td>0.1540</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2646</td>
<td>0.1538</td>
</tr>
</tbody>
</table>

N = 54, cell size^1 = 25; cell size^2 = 29

Hypothesis 5B (Block 1): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 1. Table 9 reveals that subjects fixated an average of 0.2529 msec. in Block 1 of the authentic text and 0.2746 msec. in the same block of the experimental text. The ANCOVA detected a significant difference between these two groups for Block 1 of "Der mechanisierte Hof," F = 7.10, p < 0.0104, leading to a rejection of the null hypothesis (Table 10).

The nature of the main effect due to type of referencing device as indicated by the average fixation durations predicted by the ANCOVA is shown in
Figure 1: Interaction Between Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Average Fixation Duration in Block 1 of "Der Mechanisierte Hof"
Figure 1. This main effect must be interpreted cautiously, however, in light of the significant interaction. The figure reveals that subjects' predicted average fixation durations were greater for Block 1 of the experimental text version of "Der mechanisierte Hof," which contained the repeated noun phrase "die Maschinen" [the machines], than for Block 1 of the authentic version, which contained the gendered anaphoric reference "sie" [it]. This finding indicates that fixations of longer duration were needed to process the unambiguous repeated noun phrase than were needed to interpret the less explicit anaphoric reference. The significant interaction reported above, however, shows that the predicted average fixation durations also varied according to subjects' baseline German language ability. No definitive conclusions, therefore, can be draw about the effect due solely to type of referencing device.

Hypothesis 5C (Block 1): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 1. The results of the ANCOVA reveal significant differences among subjects' average fixation durations in Block 1 of "Der mechanisierte Hof" as a function of their baseline German language ability, $F = 4.90, p < 0.0314$ (Table 10). This finding led to a rejection of the null hypothesis of no difference.

Figure 1 illustrates the nature of the main effect due to baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This main effect must also be interpreted cautiously because of the significant interaction. The figure demonstrates that as baseline German language ability scores increased, the predicted average fixation durations tended to decrease. This result indicates that as readers became more
proficient in their knowledge of the language they needed a lower percentage of total fixation time to process the information in Block 1 of the text. The significant interaction reported above, however, shows that the predicted average fixation durations also varied according to the text version that was read. No definitive conclusions, therefore, can be drawn about the effect due solely to baseline German language ability.

Table 10
Analysis of Covariance of Average Fixation Duration in Block 1 of "Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.1405</td>
<td>0.0971</td>
<td>7.10</td>
<td>0.0104</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0971</td>
<td>0.0971</td>
<td>4.90</td>
<td>0.0314</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.1428</td>
<td>0.1428</td>
<td>7.22</td>
<td>0.0098</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.9898</td>
<td>0.0198</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.2540</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5A (Block 2): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 2. In Block 2 of the text, "Der mechanisierte Hof," the nominative case plural personal pronoun "wir" [we] in the authentic text
version was compared to the noun phrase “die Bauern” [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was rejected for average fixation duration in Block 2, $F(1, 50) = 7.58, p < 0.0082$ (Table 12).

Figure 2 portrays the nature of the interaction between type of referencing device and baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This figure reveals that the predicted average fixation durations for Block 2 of the authentic text version decreased slightly as baseline German language ability scores increased. This result indicates that high ability students required slightly shorter average fixation durations to process the less explicit pronoun reference than did students with low baseline German language skills. In contrast, the results for the experimental text version, containing the repeated noun phrase “die Bauern” [the farmers], reveal that predicted average fixation durations decreased sharply as students’ baseline German language ability scores increased. This finding suggests that the repeated noun phrase was processed faster by high ability subjects than by subjects with low baseline German language skills.

Hypothesis 5B (Block 2): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 2. An analysis of Table 11 reveals that subjects spent an average of 0.2811 msec. in Block 2 of the authentic text version and 0.2903 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 2 of “Der mechanisierte Hof,” $F = 3.77$, 
Figure 2: Interaction Between Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Average Fixation Duration in Block 2 of "Der Mechanisierte Hof"
\( p > 0.0579 \), leading to a retention of the null hypothesis (Table 12).

**Table 11**

Means and Standard Deviations of Average Fixation Duration as a Function of Type of Referencing Device in Block 2 of “Der Mechaniserte Hof”

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>0.2811</td>
<td>0.1653</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>0.2903</td>
<td>0.1522</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2861</td>
<td>0.1569</td>
</tr>
</tbody>
</table>

N = 54, cell size\(^1\) = 25; cell size\(^2\) = 29

Hypothesis 5C (Block 2): **There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 2.**

The results of the ANCOVA revealed significant differences among subjects’ average fixation durations in Block 2 of “Der mechanisierte Hof” as a function of their baseline German language ability, \( F = 4.19, p < 0.0459 \) (Table 12). This finding led to a rejection of the null hypothesis of no difference.

Figure 2 illustrates the nature of the main effect due to baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This main effect must also be interpreted cautiously because of the significant interaction. The figure demonstrates that as baseline German
language ability scores increased, the predicted average fixation durations tended to decrease. This result indicates that as readers became more proficient in their knowledge of the language they needed a lower percentage of total fixation time to process the information in Block 2 of the text. The significant interaction reported above, however, shows that the predicted average fixation durations also varied according to the text version that was read. No definitive conclusions, therefore, can be draw about the effect due solely to baseline German language ability.

Table 12
Analysis of Covariance of Average Fixation Duration in Block 2 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S S</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0785</td>
<td>0.0785</td>
<td>3.77</td>
<td>0.0579</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.1579</td>
<td>0.1579</td>
<td>7.58</td>
<td>0.0082</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0873</td>
<td>0.0873</td>
<td>4.19</td>
<td>0.0459</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.0412</td>
<td>0.0208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.3055</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 3): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 3. In Block 3 of the text, “Der mechanisierte Hof,” the nominative case singular, masculine personal pronoun “er” [it] in the authentic text version was compared to the noun phrase “der Traktor” [the tractor] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 3, $F(1, 50) = 2.10, p > 0.1539$ (Table 14).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2972</td>
<td>0.2266</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2562</td>
<td>0.1350</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2752</td>
<td>0.1825</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5B (Block 3): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 3. An analysis of Table 13 reveals that subjects spent an average of 0.2972 msec.
Block 3 of the authentic text version and 0.2562 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 3 of “Der mechanisierte Hof,” $F = 1.03, p > 0.1539$, leading to a retention of the null hypothesis (Table 14).

Hypothesis 5C (Block 3): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 3. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 3 of “Der mechanisierte Hof” as a function of their baseline German language ability, $F = 1.94, p > 0.1700$ (Table 14). This finding led to a retention of the null hypothesis of no difference.

Table 14
Analysis of Covariance of Average Fixation Duration in Block 3 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0330</td>
<td>0.0330</td>
<td>1.03</td>
<td>0.3154</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0622</td>
<td>0.0622</td>
<td>1.94</td>
<td>0.1700</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0672</td>
<td>0.0672</td>
<td>2.10</td>
<td>0.1539</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.6033</td>
<td>0.0321</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.7650</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 4): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 4. In Block 4 of the text, “Der mechanisierte Hof,” the dative case plural personal pronoun “uns” [us] in the authentic text version was compared to the noun phrase “den Bauern” [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 4, $F (1, 50) = 1.43, p > 0.2378$ (Table 16).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2660</td>
<td>0.1537</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2168</td>
<td>0.1079</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2400</td>
<td>0.1321</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1$ = 25; cell size$^2$ = 29

Hypothesis 5B (Block 4): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 4. An analysis of Table 15 reveals that subjects spent an average of 0.2660 msec. in
Block 4 of the authentic text version and 0.2168 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 4 of "Der mechanisierte Hof," $F = 0.36$, $p > 0.5538$, leading to a retention of the null hypothesis (Table 16).

Hypothesis 5C (Block 4): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 4. The results of the ANCOVA revealed no significant differences among subjects' average fixation durations in Block 4 of "Der mechanisierte Hof" as a function of their baseline German language ability, $F = 0.35$, $p < 0.5590$ (Table 16). This finding led to a retention of the null hypothesis of no difference.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0061</td>
<td>0.0061</td>
<td>0.36</td>
<td>0.5538</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0060</td>
<td>0.0060</td>
<td>0.35</td>
<td>0.5590</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0248</td>
<td>0.0248</td>
<td>1.43</td>
<td>0.2378</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.8604</td>
<td>0.0172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.9254</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 5): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 5. In Block 5 of the text, “Der mechanisierte Hof,” the nominative case singular, feminine personal pronoun “sie” [it] in the authentic text version was compared to the noun phrase “die Mechanisierung” [the mechanization] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 5, \( F(1, 50) = 2.10, p > 0.1538 \) (Table 18).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>0.1922</td>
<td>0.1473</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>0.2671</td>
<td>0.1185</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2324</td>
<td>0.1366</td>
</tr>
</tbody>
</table>

N = 54, cell size\(^1\) = 25; cell size\(^2\) = 29

Hypothesis 5B (Block 5): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 5. An
analysis of Table 17 reveals that subjects spent an average of 0.1922 msec. in Block 5 of the authentic text version and 0.2671 msec. in the same block of the experimental text version. The ANCOVA detected a significant difference between these two groups for Block 5 of “Der mechanisierte Hof,” \( F = 4.42, \ p < 0.0407 \), leading to a rejection of the null hypothesis (Table 18).

Figure 3 illustrates the nature of the main effect due to type of referencing device as indicated by the average fixation durations predicted by the ANCOVA. This figure indicates that regardless of their baseline German language ability, subjects’ predicted average fixation durations were much greater for the experimental text version of “Der mechanisierte Hof,” containing the gendered anaphoric reference “sie” [it], than they were for the authentic version, which contained the repeated noun phrase “die Mechanisierung” [the mechanization]. This finding reveals that shorter average fixation durations were needed to process the less explicit but shorter anaphoric reference than were needed to interpret the unambiguous but lengthy repeated noun phrase.

Hypothesis 5C (Block 5): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 5. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 5 of “Der mechanisierte Hof” as a function of their baseline German language ability, \( F = 3.77, \ p > 0.0577 \) (Table 18). This finding led to a retention of the null hypothesis of no difference.
Figure 3: Main Effect of Baseline German Language Ability as Measured by Predicted Average Fixation Duration in Block 5 of "Der Mechanisierte Hof"
Table 18
Analysis of Covariance of Average Fixation Duration in Block 5 of “Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0717</td>
<td>0.0717</td>
<td>4.42</td>
<td>0.0407</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0612</td>
<td>0.0612</td>
<td>3.77</td>
<td>0.0577</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0340</td>
<td>0.0340</td>
<td>2.10</td>
<td>0.1538</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.8116</td>
<td>0.0162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.9892</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5A (Block 6): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 6. In Block 6 of the text, “Der mechanisierte Hof,” the accusative case plural personal pronoun “uns” [us] in the authentic text version was compared to the noun phrase “die Bauern” [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 6, F(1, 50) = 0.03, p > 0.8697 (Table 20).

Hypothesis 5B (Block 6): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 6. An
analysis of Table 19 reveals that subjects spent an average of 0.2221 msec. in Block 6 of the authentic text version and 0.1655 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 6 of “Der mechanisierte Hof,” \( F = 0.21, p > 0.6504 \), leading to a retention of the null hypothesis (Table 20).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>0.2221</td>
<td>0.1261</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>0.1655</td>
<td>0.1235</td>
</tr>
<tr>
<td>Overall</td>
<td>0.1917</td>
<td>0.1268</td>
</tr>
</tbody>
</table>

\( N = 54, \text{ cell size}^1 = 25; \text{ cell size}^2 = 29 \)

Hypothesis 5C (Block 6): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 6. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 6 of “Der mechanisierte Hof” as a function of their baseline German language ability, \( F = 0.30, p > 0.5864 \) (Table 20). This finding led to a retention of the null hypothesis of no difference.
Hypothesis 5A (Block 7): **There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 7.** In Block 7 of the text, “Der mechanisierte Hof,” the nominative case plural personal pronoun “wir” [we] in the authentic text version was compared to the noun phrase “die Bauern” [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 7, \( F(1, 50) = 0.13, p > 0.7198 \) (Table 22).

Hypothesis 5B (Block 7): **There is no significant difference due to type of referencing device as measured by average fixation duration in Block 7.** An
analysis of Table 21 reveals that subjects spent an average of 0.2696 msec. in Block 7 of the authentic text version and 0.2050 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 7 of “Der mechanisierte Hof,” $F = 0.69, p > 0.4107$, leading to a retention of the null hypothesis (Table 22).

Table 21
Means and Standard Deviations of Average Fixation Duration in Block 7 of “Der Mechanisierte Hof” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2696</td>
<td>0.2306</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2050</td>
<td>0.1279</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2349</td>
<td>0.1838</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5C (Block 7): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 7. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 7 of “Der mechanisierte Hof” as a function of their baseline German language ability, $F = 1.55, p > 0.2196$ (Table 22). This finding led to a retention of the null hypothesis of no difference.
Table 22
Analysis of Covariance of Average Fixation Duration in Block 7 of "Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0231</td>
<td>0.0231</td>
<td>0.69</td>
<td>0.4107</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0519</td>
<td>0.0519</td>
<td>1.55</td>
<td>0.2196</td>
</tr>
<tr>
<td>Referencing Device X Baseline</td>
<td>1</td>
<td>0.0044</td>
<td>0.0044</td>
<td>0.13</td>
<td>0.7198</td>
</tr>
<tr>
<td>Ability</td>
<td>50</td>
<td>1.680</td>
<td>0.0336</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Total                         | 53 | 1.7903|      |     |        |

Table 23 shows a summary of the results of the analyses of covariance for average fixation duration in blocks for "Der mechanisierte Hof."
<table>
<thead>
<tr>
<th>Block</th>
<th>Interaction</th>
<th>Type of Referencing Device</th>
<th>Baseline German Language Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>(sie/die Maschinen)</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Block 2</td>
<td>(wir/die Bauern)</td>
<td>S</td>
<td>NS</td>
</tr>
<tr>
<td>Block 3</td>
<td>(er/der Traktor)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 4</td>
<td>(uns/den Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 5</td>
<td>(sie/die Mechanisierung)</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>Block 6</td>
<td>(uns/die Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 7</td>
<td>(wir/die Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

S (significant), NS (nonsignificant)

**Literary Text: “Im Volksgarten”**

Copies of both the authentic and experimental versions of the literary text, “Im Volksgarten,” with the block matrix mapped onto the text are provided in Appendix M. Only those blocks containing anaphoric expressions or substitute nouns or noun phrases were selected for further investigation. In the text, “Im Volksgarten,” ten blocks containing anaphoric references were chosen for analysis. The average fixation durations for each block of interest are examined independently in the following section.
Hypothesis 5A (Block 1): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 1. In Block 1 of the text, "Im Volksgarten," the dative case singular, masculine personal pronoun "ihn" [it] in the authentic text version was compared to the noun phrase "dem Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 1, $F(1, 50) = 0.00, p > 0.9875$ (Table 25).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2352</td>
<td>0.0811</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2020</td>
<td>0.0933</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2200</td>
<td>0.0886</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1$ = 25; cell size$^2$ = 29

Hypothesis 5B (Block 1): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 1. An analysis of Table 24 reveals that subjects spent an average of 0.2352 msec. in
Block 1 of the authentic text version and 0.2020 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 1 of “Im Volksgarten,” $F = 0.22$, $p > 0.6412$, leading to a retention of the null hypothesis (Table 25).

Hypothesis 5C (Block 1): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 1. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 1 of “Im Volksgarten” as a function of their baseline German language ability, $F = 2.03$, $p > 0.1603$ (Table 25). This finding led to a retention of the null hypothesis of no difference.

Table 25
Analysis of Covariance of Average Fixation Duration in Block 1 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.22</td>
<td>0.6412</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0157</td>
<td>0.0157</td>
<td>2.03</td>
<td>0.1603</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.00</td>
<td>0.9875</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.3857</td>
<td>0.0077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.4163</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 2): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 2. In Block 2 of the text, “Im Volksgarten,” the nominative case singular, feminine personal pronoun “sie” [she] in the authentic text version was compared to the noun “Anna” [Anna] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 2, $F(1, 50) = 0.396, p > 0.0522$ (Table 27).

<table>
<thead>
<tr>
<th>Table 26</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Means and Standard Deviations of Average Fixation Duration in Block 2 of “Im Volksgarten” as a Function of Type of Referencing Device</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>0.2736</td>
<td>0.1854</td>
</tr>
<tr>
<td>Experimental²</td>
<td>0.1090</td>
<td>0.1296</td>
</tr>
<tr>
<td>Overall</td>
<td>0.1974</td>
<td>0.1806</td>
</tr>
</tbody>
</table>

$N = 54$, cell size¹ = 25; cell size² = 29

Hypothesis 5B (Block 2): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 2. An analysis of Table 26 reveals that subjects spent an average of 0.2736 msec. in
Block 2 of the authentic text version and 0.1090 msec. in the same block of the experimental text version. The ANCOVA detected a significant difference between these two groups for Block 2 of “Im Volksgarten,” $F = 11.09, p < 0.0016$, leading to a rejection of the null hypothesis (Table 27).

The nature of the main effect due to type of referencing device as indicated by the average fixation durations predicted by the ANCOVA is shown in Figure 4. This figure indicates that regardless of their baseline German language ability, subjects predicted average fixation durations were much greater for the authentic text version of “Im Volksgarten,” containing the gendered anaphoric reference “sie” [she], than they were for the experimental version, which contained the repeated noun “Anna.” This finding reveals that longer fixations were needed to process the less explicit anaphoric reference than were needed to interpret the unambiguous repeated noun.

Hypothesis 5C (Block 2): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 2. The results of the ANCOVA revealed significant differences among subjects’ average fixation durations in Block 2 of “Im Volksgarten” as a function of their baseline German language ability, $F = 11.32, p < 0.0015$ (Table 27). This finding led to a rejection of the null hypothesis of no difference.

Figure 4 illustrates the nature of the main effect due to baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This figure demonstrates that as baseline German language ability scores increased, the predicted average fixation durations decreased for both text versions. This result indicates that as readers became more proficient in
their knowledge of the language they required fixations of shorter duration to process the information in Block 2 of the text.

Table 27
Analysis of Covariance of Average Fixation Duration in Block 2 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S S</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.2284</td>
<td>0.2284</td>
<td>11.09</td>
<td>0.0016</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.2331</td>
<td>0.2331</td>
<td>11.32</td>
<td>0.0015</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0815</td>
<td>0.0815</td>
<td>3.96</td>
<td>0.0522</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.0298</td>
<td>0.0206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.7292</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4: Main Effects of Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Average Fixation Duration in Block 2 of "Im Volksgarten"
Hypothesis 5A (Block 3): *There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 3.* In Block 3 of the text, "Im Volksgarten," the accusative case singular, masculine personal pronoun "ihn" [it] in the authentic text version was compared to the noun phrase "den Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 3, $F(1, 50) = 3.03, p > 0.0879$ (Table 29).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>0.2424</td>
<td>0.0980</td>
</tr>
<tr>
<td>Experimental²</td>
<td>0.2349</td>
<td>0.0700</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2389</td>
<td>0.0855</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29

Hypothesis 5B (Block 3): *There is no significant difference due to type of referencing device as measured by average fixation duration in Block 3.* An analysis of Table 28 reveals that subjects spent an average of 0.2424 msec. in
Block 3 of the authentic text version and 0.2349 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 3 of "Im Volksgarten," $F = 2.92, p > 0.0938$, leading to a retention of the null hypothesis (Table 29).

Hypothesis 5C (Block 3): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 3. The results of the ANCOVA revealed no significant differences among subjects' average fixation durations in Block 3 of "Im Volksgarten" as a function of their baseline German language ability, $F = 3.38$, $p > 0.0718$ (Table 29). This finding led to a retention of the null hypothesis of no difference.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0198</td>
<td>0.0198</td>
<td>2.92</td>
<td>0.0938</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0230</td>
<td>0.0230</td>
<td>3.38</td>
<td>0.0718</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0206</td>
<td>0.0206</td>
<td>3.03</td>
<td>0.0879</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.3400</td>
<td>0.0068</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.3874</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 4): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 4. In Block 4 of the text, "Im Volksgarten," the accusative case singular, masculine personal pronoun "ihn" [it] in the authentic text version was compared to the noun phrase "den Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 4, $F(1, 50) = 0.05, p > 0.8230$ (Table 31).

Table 30
Means and Standard Deviations of Average Fixation Duration in Block 4 of "Im Volksgarten" as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2649</td>
<td>0.1224</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2334</td>
<td>0.1376</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2504</td>
<td>0.1294</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1$ = 25; cell size$^2$ = 29

Hypothesis 5B (Block 4): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 4. An analysis of Table 30 reveals that subjects spent an average of 0.2649 msec. in
Block 4 of the authentic text version and 0.2334 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 4 of “Im Volksgarten,” $F = 0.01$, $p > 0.9263$, leading to a retention of the null hypothesis (Table 31).

Hypothesis 5C (Block 4): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 4. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 4 of “Im Volksgarten” as a function of their baseline German language ability, $F = 1.88$, $p > 0.1765$ (Table 31). This finding led to a retention of the null hypothesis of no difference.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.01</td>
<td>0.9263</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0317</td>
<td>0.0317</td>
<td>1.88</td>
<td>0.1765</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.05</td>
<td>0.8230</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.8421</td>
<td>0.0168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.8873</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 5A (Block 5): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 5. In Block 5 of the text, “Im Volksgarten,” the nominative case singular, feminine personal pronoun “sie” [she] in the authentic text version was compared to the noun “Anna” [Anna] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 5, \( F(1, 50) = 1.48, p > 0.2294 \) (Table 33).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>0.3218</td>
<td>0.2369</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>0.2692</td>
<td>0.1314</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2974</td>
<td>0.1954</td>
</tr>
</tbody>
</table>

\( N = 54, \) cell size\(^1\) = 25; cell size\(^2\) = 29

Hypothesis 5B (Block 5): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 5. An analysis of Table 32 reveals that subjects spent an average of 0.3218 msec. in
Block 5 of the authentic text version and 0.2692 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 5 of “Im Volksgarten,” $F = 2.17, p > 0.1468$, leading to a retention of the null hypothesis (Table 33).

Table 33
Analysis of Covariance of Average Fixation Duration in Block 5 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0806</td>
<td>0.0806</td>
<td>2.17</td>
<td>0.1468</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0657</td>
<td>0.0657</td>
<td>1.77</td>
<td>0.1894</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0550</td>
<td>0.0550</td>
<td>1.48</td>
<td>0.2294</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.8564</td>
<td>0.0371</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>2.0233</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5C (Block 5): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 5. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 5 of “Im Volksgarten” as a function of their baseline German language ability, $F = 1.77, p > 0.1894$ (Table 33). This finding led to a retention of the null hypothesis of no difference.
Hypothesis 5A (Block 6): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 6. In Block 6 of the text, “Im Volksgarten,” the dative case singular, masculine personal pronoun “ihm” [it] in the authentic text version was compared to the noun phrase “dem Ballon” [the Balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 6, $F (1, 50) = 0.55, p > 0.4612$ (Table 35).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2336</td>
<td>0.1123</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2524</td>
<td>0.1112</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2423</td>
<td>0.1111</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5B (Block 6): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 6. An analysis of Table 34 reveals that subjects spent an average of 0.2336 msec. in
Block 6 of the authentic text version and 0.2524 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 6 of "Im Volksgarten," $F = 0.20$, $p > 0.6575$, leading to a retention of the null hypothesis (Table 35).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.20</td>
<td>0.6575</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0146</td>
<td>0.0146</td>
<td>1.16</td>
<td>0.2862</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0069</td>
<td>0.0069</td>
<td>0.55</td>
<td>0.4612</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.6268</td>
<td>0.0125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.6545</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5C (Block 6): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 6. The results of the ANCOVA revealed no significant differences among subjects' average fixation durations in Block 6 of "Im Volksgarten" as a function of their baseline German language ability, $F = 1.16$, $p > 0.2862$ (Table 35). This finding led to a retention of the null hypothesis of no difference.
Hypothesis 5A (Block 7): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 7. In Block 7 of the text, “Im Volksgarten,” the nominative case singular, masculine personal pronoun “er” [it] in the authentic text version was compared to the noun phrase “der Ballon” [the Balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 7, $F(1, 50) = 0.12, p > 0.7340$ (Table 37).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2584</td>
<td>0.1282</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.2838</td>
<td>0.1924</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2702</td>
<td>0.1601</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5B (Block 7): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 7. An analysis of Table 36 reveals that subjects spent an average of 0.2584 msec. in
Block 7 of the authentic text version and 0.2838 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 7 of "Im Volksgarten," $F = 0.31$, $p > 0.5818$, leading to a retention of the null hypothesis (Table 37).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0080</td>
<td>0.0080</td>
<td>0.31</td>
<td>0.5818</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0537</td>
<td>0.0537</td>
<td>2.08</td>
<td>0.1558</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0030</td>
<td>0.0030</td>
<td>0.12</td>
<td>0.7340</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.2940</td>
<td>0.0259</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.3579</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5C (Block 7): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 7.

The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 7 of "Im Volksgarten" as a function of their baseline German language ability, $F = 2.08$, $p > 0.1558$ (Table 37). This finding led to a retention of the null hypothesis of no difference.
Hypothesis 5A (Block 8): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 8. In Block 8 of the text, "Im Volksgarten," the accusative case singular, masculine personal pronoun "ihn" [it] in the authentic text version was compared to the noun phrase "den Ballon" [the Balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 8, $F(1, 50) = 0.12, p > 0.7347$ (Table 39).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.2344</td>
<td>0.1250</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.1901</td>
<td>0.0731</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2139</td>
<td>0.1057</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5B (Block 8): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 8. An analysis of Table 38 reveals that subjects spent an average of 0.2344 msec. in
Block 8 of the authentic text version and 0.1901 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 8 of “Im Volksgarten,” $F = 0.05$, $p > 0.8190$, leading to a retention of the null hypothesis (Table 39).

Table 39
Analysis of Covariance of Average Fixation Duration in Block 8 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.05</td>
<td>0.8190</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0160</td>
<td>0.0160</td>
<td>1.46</td>
<td>0.2330</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0013</td>
<td>0.0013</td>
<td>0.12</td>
<td>0.7347</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.5492</td>
<td>0.0110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.5923</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5C (Block 8): **There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 8.**

The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 8 of “Im Volksgarten” as a function of their baseline German language ability, $F = 1.46$, $p > 0.2330$ (Table 39). This finding led to a retention of the null hypothesis of no difference.
Hypothesis 5A (Block 9): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 9. In Block 9 of the text, “Im Volksgarten,” the dative case singular, feminine personal pronoun “ihr” [she] in the authentic text version was compared to the noun phrase “dem Mädchen” [the girl] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 9, \( F(1, 50) = 0.99, p > 0.0511 \) (Table 41).

Table 40
Means and Standard Deviations of Average Fixation Duration in Block 9 of “Im Volksgarten” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>0.2855</td>
<td>0.1996</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>0.1513</td>
<td>0.0902</td>
</tr>
<tr>
<td>Overall</td>
<td>0.2234</td>
<td>0.1712</td>
</tr>
</tbody>
</table>

\( N = 54, \) cell size\(^1\) = 25; cell size\(^2\) = 29

Hypothesis 5B (Block 9): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 9. An analysis of Table 40 reveals that subjects spent an average of 0.2855 msec. in
Block 9 of the authentic text version and 0.1513 msec. in the same block of the experimental text version. The ANCOVA detected a significant difference between these two groups for Block 9 of “Im Volksgarten,” $F = 9.33$, $p < 0.0036$, leading to a rejection of the null hypothesis (Table 41).

The nature of the main effect due to type of referencing device as indicated by the average fixation durations predicted by the ANCOVA is shown in Figure 5. This figure indicates that regardless of their baseline German language ability, subjects’ predicted average fixation durations were much greater for the authentic text version of “Im Volksgarten,” containing the gendered anaphoric reference “ihr” [her], than they were for the experimental version, which contained the repeated noun phrase “dem Mädchen” [the girl]. This finding reveals that longer fixations were needed to process the less explicit anaphoric reference than were needed to interpret the unambiguous repeated noun phrase.

Hypothesis 5C (Block 9): **There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 9.** The results of the ANCOVA revealed significant differences among subjects’ average fixation durations in Block 9 of “Im Volksgarten” as a function of their baseline German language ability, $F = 7.82$, $p < 0.0073$ (Table 41). This finding led to a rejection of the null hypothesis of no difference.

Figure 5 illustrates the nature of the main effect due to baseline German language ability as indicated by the average fixation durations predicted by the ANCOVA. This figure demonstrates that as baseline German language ability scores increased, the predicted average fixation durations decreased for both
text versions. This result indicates that as readers became more proficient in their knowledge of the language they required fixations of shorter duration to process the information in Block 9 of the text.

Table 41
Analysis of Covariance of Average Fixation Duration in Block 9 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.1952</td>
<td>0.1952</td>
<td>9.33</td>
<td>0.0036</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.1636</td>
<td>0.1636</td>
<td>7.82</td>
<td>0.0073</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0835</td>
<td>0.0835</td>
<td>3.99</td>
<td>0.0511</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>1.0457</td>
<td>0.0209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>1.5525</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5: Main Effects of Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Average Fixation Duration in Block 9 of "Im Volksgarten"
Hypothesis 5A (Block 10): There is no significant interaction between type of referencing device and baseline German language ability as measured by average fixation duration in Block 10. In Block 10 of the text, "Im Volksgarten," the accusative case singular, masculine demonstrative pronoun "diesen" [this (one)] in the authentic text version was compared to the noun phrase "diesen Ballon" [this Balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for average fixation duration in Block 10, $F(1, 50) = 0.49, p > 0.4867$ (Table 43).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.1112</td>
<td>0.1118</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.0996</td>
<td>0.0903</td>
</tr>
<tr>
<td>Overall</td>
<td>0.1059</td>
<td>0.1017</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 5B (Block 10): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 10. An
analysis of Table 42 reveals that subjects spent an average of 0.1112 msec. in Block 10 of the authentic text version and 0.0996 msec. in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 10 of "Im Volksgarten," $F = 0.25, p > 0.6212$, leading to a retention of the null hypothesis (Table 43).

### Table 43
Analysis of Covariance of Average Fixation Duration in Block 10 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.25</td>
<td>0.6212</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.06</td>
<td>0.8047</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0053</td>
<td>0.0053</td>
<td>0.49</td>
<td>0.4867</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>0.5397</td>
<td>0.0108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>0.5477</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 5C (Block 10): There is no significant difference due to baseline German language ability as measured by average fixation duration in Block 10. The results of the ANCOVA revealed no significant differences among subjects’ average fixation durations in Block 10 of "Im Volksgarten" as a function of their baseline German language ability, $F = 0.06, p > 0.8047$ (Table 43). This finding
led to a retention of the null hypothesis of no difference.

Table 44 shows a summary of the results of the analyses of covariance for average fixation duration in blocks for “Im Volksgarten.”

<table>
<thead>
<tr>
<th>Block</th>
<th>Interaction</th>
<th>Type of Referencing Device</th>
<th>Baseline German Language Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>(ihm/dem Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 2</td>
<td>(sie/Anna)</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>Block 3</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 4</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 5</td>
<td>(sie/Anna)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 6</td>
<td>(ihn/dem Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 7</td>
<td>(er/der Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 8</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 9</td>
<td>(ihr/dem Mädchen)</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>Block 10</td>
<td>(diesen/diesen Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

S (significant), NS (nonsignificant)
ANCOVA Results for Percentage of Total Fixation Time in Blocks

In this section the percentage of total fixation time that subjects spent fixating a block of the authentic text version containing an anaphoric expression will be compared to the percentage of total fixation time that they spent fixating a block of the experimental text version containing a noun or noun phrase.

Expository Text: “Der Mechanisierte Hof”

Copies of both the authentic and experimental versions of the expository text, “Der mechanisierte Hof,” with the block matrix mapped onto the text are provided in Appendix L. Only those blocks containing anaphoric expressions or substitute nouns or noun phrases were selected for further investigation. In the text, “Der mechanisierte Hof,” seven blocks containing anaphoric references were chosen for analysis. The percentage of total fixation times for each block of interest are examined independently in the following section.

Hypothesis 6A (Block 1): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 1. In Block 1 of the text, “Der mechanisierte Hof,” the nominative case plural personal pronoun “sie” [they] in the authentic text version was compared to the noun phrase “die Maschinen” [the machines] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was rejected for percentage of total fixation time in Block 1, \( F(1, 50) = 5.78, p < 0.0200 \) (Table 46).
Figure 6 portrays the nature of the interaction between type of referencing device and baseline German language ability as indicated by the percentage of total fixation times predicted by the ANCOVA. This figure reveals that the predicted percentage of total fixation time for the repeated noun phrase “die Maschinen” [the machines] in the experimental text version decreased sharply as subjects’ baseline German language ability scores increased. This result indicates that the unambiguous noun phrase was processed more quickly by high ability students than by those with low baseline German language abilities. In contrast, the results for the experimental text version indicated that the predicted percentage of total fixation time increased sharply as students’ baseline German language ability scores increased. This finding suggests that high ability subjects needed more time to process the less explicit pronoun reference, “sie” [they], than low ability subjects did.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>1.9922</td>
<td>1.5334</td>
</tr>
<tr>
<td>Experimental²</td>
<td>1.7325</td>
<td>1.2287</td>
</tr>
<tr>
<td>Overall</td>
<td>1.8527</td>
<td>1.3709</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29
Hypothesis 6B (Block 1): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 1. An analysis of Table 45 reveals that subjects spent an average of 1.9922% of their total fixation time in Block 1 of the authentic text version and 1.7325% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 1 of “Der mechanisierte Hof,” \( F = 3.87, p > 0.0547 \), leading to a retention of the null hypothesis (Table 46).

Table 46
Analysis of Covariance of Percentage of Total Fixation Time in Block 1 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>S S</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>6.8367</td>
<td>6.8367</td>
<td>3.87</td>
<td>0.0547</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0502</td>
<td>0.0502</td>
<td>0.03</td>
<td>0.8668</td>
</tr>
<tr>
<td>Referencing Device X Baseline</td>
<td>1</td>
<td>10.2007</td>
<td>10.2007</td>
<td>5.78</td>
<td>0.0200</td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>88.2984</td>
<td>1.7660</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>99.6088</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6: Interaction Between Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Percentage of Total Fixation Time in Block 1 of “Der Mechanisierte Hof”
Hypothesis 6C (Block 1): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 1. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 1 of "Der mechanisierte Hof" as a function of their baseline German language ability, $F = 0.03$, $p > 0.8668$ (Table 46). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 2): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 2. In Block 2 of the text, "Der mechanisierte Hof," the nominative case plural personal pronoun "wir" [we] in the authentic text version was compared to the noun phrase "die Bauern" [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 2, $F (1, 50) = 2.03$, $p > 0.1603$ (Table 48).

Hypothesis 6B (Block 2): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 2. An analysis of Table 47 reveals that subjects spent an average of 2.2735% of their total fixation time in Block 2 of the authentic text version and 1.9127% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 2 of "Der mechanisierte Hof," $F = 1.07$, $p > 0.3051$, leading to a retention of the null hypothesis (Table 48).
Table 47  
Means and Standard Deviations of Percentage of Total Fixation Time in Block 2 of “Der Mechanisierte Hof” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic^1</td>
<td>2.2735</td>
<td>2.0900</td>
</tr>
<tr>
<td>Experimental^2</td>
<td>1.9127</td>
<td>1.1039</td>
</tr>
<tr>
<td>Overall</td>
<td>2.0800</td>
<td>1.6293</td>
</tr>
</tbody>
</table>

N = 54, cell size^1 = 25; cell size^2 = 29

Hypothesis 6C (Block 2): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 2. The results of the ANCOVA revealed no significant differences among subjects’ percentage of total fixation times in Block 2 of “Der mechanisierte Hof” as a function of their baseline German language ability, F = 2.03, p > 0.1603 (Table 48). This finding led to a retention of the null hypothesis of no difference.
### Table 48
Analysis of Covariance of Percentage of Total Fixation Time in Block 2 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>2.8573</td>
<td>2.8573</td>
<td>1.07</td>
<td>0.3051</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.7600</td>
<td>0.7600</td>
<td>0.29</td>
<td>0.5954</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>5.4046</td>
<td>5.4046</td>
<td>2.03</td>
<td>0.1603</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>133.0427</td>
<td>2.6609</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>53</td>
<td>140.6972</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 3): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 3. In Block 3 of the text, “Der mechanisierte Hof,” the nominative singular, masculine personal pronoun “er” [it] in the authentic text version was compared to the noun phrase “der Traktor” [the tractor] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 3, $F (1, 50) = 0.00$, $p > 0.9571$ (Table 50).
Hypothesis 6B (Block 3): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 3. An analysis of Table 49 reveals that subjects spent an average of 1.7547% of their total fixation time in Block 3 of the authentic text version and 1.6938% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 3 of "Der mechanisierte Hof," $F = 0.01, p > 0.9093$, leading to a retention of the null hypothesis (Table 50).

Table 49
Means and Standard Deviations of Percentage of Total Fixation Time in Block 3 of "Der Mechanisierte Hof" as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>1.7547</td>
<td>1.3332</td>
</tr>
<tr>
<td>Experimental²</td>
<td>1.6938</td>
<td>0.8119</td>
</tr>
<tr>
<td>Overall</td>
<td>1.7220</td>
<td>1.0743</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29

Hypothesis 6C (Block 3): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 3. The results of the ANCOVA revealed no significant differences among
subjects' percentage of total fixation times in Block 3 of "Der mechanisierte Hof" as a function of their baseline German language ability, $E = 0.34$, $p > 0.5636$ (Table 50). This finding led to a retention of the null hypothesis of no difference.

![Table 50](image)

Analysis of Covariance of Percentage of Total Fixation Time in Block 3 of "Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0159</td>
<td>0.0159</td>
<td>0.01</td>
<td>0.9093</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.4102</td>
<td>0.4102</td>
<td>0.34</td>
<td>0.5636</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0035</td>
<td>0.0035</td>
<td>0.00</td>
<td>0.9571</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>60.6917</td>
<td>1.2138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>61.1628</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 4): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 4. In Block 4 of the text, "Der mechanisierte Hof," the dative case plural personal pronoun "uns" [us] in the authentic text version was compared to the noun phrase "den Bauern" [the farmers] that appeared in the experimental text version. The null hypothesis of
no interaction between the two independent variables was retained for percentage of total fixation time in Block 4, \( F(1, 50) = 3.18, p > 0.0806 \) (Table 52).

Hypothesis 6B (Block 4): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 4. An analysis of Table 51 reveals that subjects spent an average of 1.8716% of their total fixation time in Block 4 of the authentic text version and 1.1720% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 4 of “Der mechanisierte Hof,” \( F = 0.49, p > 0.4866 \), leading to a retention of the null hypothesis (Table 52).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^1)</td>
<td>1.8716</td>
<td>1.3096</td>
</tr>
<tr>
<td>Experimental(^2)</td>
<td>1.1720</td>
<td>0.6322</td>
</tr>
<tr>
<td>Overall</td>
<td>1.4959</td>
<td>1.0544</td>
</tr>
</tbody>
</table>

\( N = 54 \), cell size\(^1\) = 25; cell size\(^2\) = 29
Table 52
Analysis of Covariance of Percentage of Total Fixation Time
in Block 4 of "Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.4328</td>
<td>0.4328</td>
<td>0.49</td>
<td>0.4866</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>6.0471</td>
<td>6.0471</td>
<td>6.86</td>
<td>0.0116</td>
</tr>
<tr>
<td>Referencing Device X</td>
<td>1</td>
<td>2.8016</td>
<td>2.8016</td>
<td>3.18</td>
<td>0.0806</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>44.0463</td>
<td>0.8809</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>58.9237</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 4): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 4. The results of the ANCOVA revealed significant differences among subjects’ percentage of total fixation times in Block 4 of “Der mechanisierte Hof” as a function of their baseline German language ability, F = 6.86, p < 0.0116 (Table 52). This finding led to a rejection of the null hypothesis of no difference.

Figure 7 illustrates the nature of the main effect due to baseline German language ability as indicated by the percentage of total fixation times predicted by the ANCOVA. This figure demonstrates that as baseline German language ability scores increased, the predicted average fixation durations also increased for both text versions. This result indicates that as readers became more
Figure 7: Main Effect of Baseline German Language Ability as Measured by Predicted Percentage of Total Fixation Time in Block 4 of "Der Mechanisierte Hof"
proficient in their knowledge of the language they required a greater percentage of total fixation time to process the information in Block 4 of the text.

Hypothesis 6A (Block 5): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 5. In Block 5 of the text, "Der mechanisierte Hof," the nominative case singular, feminine personal pronoun "sie" [it] in the authentic text version was compared to the noun phrase "die Mechanisierung" [the mechanization] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 5, $F(1, 50) = 0.00$, $p > 0.9878$ (Table 54).

| Table 53 |
| Means and Standard Deviations of Percentage of Total Fixation Time in Block 5 of "Der Mechanisierte Hof" as a Function of Type of Referencing Device |

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>0.8600</td>
<td>0.8278</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>1.2438</td>
<td>0.6020</td>
</tr>
<tr>
<td>Overall</td>
<td>1.0661</td>
<td>0.7342</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$
Hypothesis 6B (Block 5): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 5. An analysis of Table 53 reveals that subjects spent an average of 0.8600% of their total fixation time in Block 5 of the authentic text version and 1.2438% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 5 of “Der mechanisierte Hof,” $F = 0.56, p > 0.4577$, leading to a retention of the null hypothesis (Table 54).

### Table 54
Analysis of Covariance of Percentage of Total Fixation Time in Block 5 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.2908</td>
<td>0.2908</td>
<td>0.56</td>
<td>0.4577</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.6319</td>
<td>0.6319</td>
<td>1.22</td>
<td>0.2752</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.00</td>
<td>0.9878</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>25.9581</td>
<td>0.5192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>28.5701</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 5): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in
Block 5. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 5 of "Der mechanisierte Hof" as a function of their baseline German language ability, $F = 1.22$, $p > 0.2752$ (Table 54). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 6): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 6. In Block 6 of the text, "Der mechanisierte Hof," the accusative case plural personal pronoun "uns" [us] in the authentic text version was compared to the noun phrase "die Bauern" [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 6, $F(1, 50) = 0.22$, $p > 0.6438$ (Table 56).

Hypothesis 6B (Block 6): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 6. An analysis of Table 55 reveals that subjects spent an average of 0.6804% of their total fixation time in Block 6 of the authentic text version and 1.2003% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 6 of "Der mechanisierte Hof," $F = 1.87$, $p > 0.1776$, leading to a retention of the null hypothesis (Table 56).
Table 55
Means and Standard Deviations of Percentage of Total Fixation Time in Block 6 of "Der Mechanisierte Hof" as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic</td>
<td>1.2003</td>
<td>0.8570</td>
</tr>
<tr>
<td>Experimental</td>
<td>0.6804</td>
<td>0.5975</td>
</tr>
<tr>
<td>Overall</td>
<td>0.9211</td>
<td>0.7679</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1$ = 25; cell size$^2$ = 29

Hypothesis 6C (Block 6): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 6. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 6 of "Der mechanisierte Hof" as a function of their baseline German language ability, $F = 1.63$, $p > 0.2073$ (Table 56). This finding led to a retention of the null hypothesis of no difference.
Table 56
Analysis of Covariance of Percentage of Total Fixation Time in Block 6 of "Der Mechanisierte Hof" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.9945</td>
<td>0.9945</td>
<td>1.87</td>
<td>0.1776</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.8662</td>
<td>0.8662</td>
<td>1.63</td>
<td>0.2078</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.1151</td>
<td>0.1151</td>
<td>0.22</td>
<td>0.6436</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>26.5919</td>
<td>0.5318</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>31.2517</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 7): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 7. In Block 7 of the text, "Der mechanisierte Hof," the nominative case plural personal pronoun "wir" [we] in the authentic text version was compared to the noun phrase "die Bauern" [the farmers] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 7, \( F(1, 50) = 0.05, p > 0.8169 \) (Table 58).
Hypothesis 6B (Block 7): **There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 7.**

An analysis of Table 57 reveals that subjects spent an average of 1.4054% of their total fixation time in Block 7 of the authentic text version and 0.9499% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 7 of "Der mechanisierte Hof," $F = 0.22, p > 0.6427$, leading to a retention of the null hypothesis (Table 58).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>1.4054</td>
<td>1.0920</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.9499</td>
<td>0.6095</td>
</tr>
<tr>
<td>Overall</td>
<td>1.1608</td>
<td>0.8881</td>
</tr>
</tbody>
</table>

N = 54, cell size$^1$ = 25; cell size$^2$ = 29

Hypothesis 6C (Block 7): **There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 7.** The results of the ANCOVA revealed no significant differences among
subjects’ percentage of total fixation times in Block 7 of “Der mechanisierte Hof” as a function of their baseline German language ability, $F = 0.72, p > 0.4004$ (Table 58). This finding led to a retention of the null hypothesis of no difference.

Table 58
Analysis of Covariance of Percentage of Total Fixation Time in Block 7 of “Der Mechanisierte Hof” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.1675</td>
<td>0.1675</td>
<td>0.22</td>
<td>0.6427</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.5531</td>
<td>0.5531</td>
<td>0.72</td>
<td>0.4004</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0417</td>
<td>0.0417</td>
<td>0.05</td>
<td>0.8169</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>38.4440</td>
<td>0.7689</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>41.8060</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 59 shows a summary of the results of the analyses of covariance for percentage of total fixation time in blocks for “Der mechanisierte Hof.”
Table 59  
Summary Results of ANCOVAs for Percentage of Total Fixation Time in Blocks of “Der Mechanisierte Hof”

<table>
<thead>
<tr>
<th>Block</th>
<th>Interaction</th>
<th>Type of Referencing Device</th>
<th>Baseline German Language Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>(sie/die Maschinen)</td>
<td>S</td>
<td>NS</td>
</tr>
<tr>
<td>Block 2</td>
<td>(wir/die Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 3</td>
<td>(er/der Traktor)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 4</td>
<td>(uns/den Bauern)</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>Block 5</td>
<td>(sie/die Mechanisierung)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 6</td>
<td>(uns/die Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 7</td>
<td>(wir/die Bauern)</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

S (significant), NS (nonsignificant)

**Literary Text: “Im Volksgarten”**

Copies of both the authentic and experimental versions of the literary text, “Im Volksgarten,” with the block matrix mapped onto the text are provided in Appendix M. Only those blocks containing anaphoric expressions or substitute nouns or noun phrases were selected for further investigation. In the text, “Im Volksgarten,” ten blocks containing anaphoric references were chosen for analysis. The percentage of total fixation times for each block of interest are examined independently in the following section.
Hypothesis 6A (Block 1): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 1. In Block 1 of the text, "Im Volksgarten," the dative case singular, masculine personal pronoun "ihm" [it] in the authentic text version was compared to the noun phrase "dem Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 1, $F(1, 50) = 0.03, p > 0.8660$ (Table 61).

Table 60
Means and Standard Deviations of Percentage of Total Fixation Time in Block 1 of "Im Volksgarten" as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>2.4183</td>
<td>1.7423</td>
</tr>
<tr>
<td>Experimental²</td>
<td>1.7821</td>
<td>0.9854</td>
</tr>
<tr>
<td>Overall</td>
<td>2.1237</td>
<td>1.4649</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29

Hypothesis 6B (Block 1): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 1.
An analysis of Table 60 reveals that subjects spent an average of 2.4183% of their total fixation time in Block 1 of the authentic text version and 1.7821% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 1 of “Im Volksgarten,” $F = 0.54, p > 0.4639$, leading to a retention of the null hypothesis (Table 61).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>1.1793</td>
<td>1.1793</td>
<td>0.54</td>
<td>0.4639</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.00</td>
<td>0.9944</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0623</td>
<td>0.0623</td>
<td>0.03</td>
<td>0.8660</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>108.2325</td>
<td>2.1647</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>113.7297</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 1): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 1. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 1 of “Im Volksgarten” as a
function of their baseline German language ability, $F = 0.00$, $p > 0.9944$ (Table 61). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 2): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 2. In Block 2 of the text, "Im Volksgarten," the nominative case singular, feminine personal pronoun "sie" [she] in the authentic text version was compared to the noun "Anna" [Anna] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 2, $F (1, 50) = 0.21$, $p > 0.6495$ (Table 63).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>1.9500</td>
<td>1.5729</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>0.4477</td>
<td>0.5377</td>
</tr>
<tr>
<td>Overall</td>
<td>1.2545</td>
<td>1.4176</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$
Table 63
Analysis of Covariance of Percentage of Total Fixation Time in Block 2 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>6.2850</td>
<td>6.2850</td>
<td>4.27</td>
<td>0.0440</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>2.1960</td>
<td>2.1960</td>
<td>1.49</td>
<td>0.2276</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.3075</td>
<td>0.3075</td>
<td>0.21</td>
<td>0.6495</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>73.5796</td>
<td>1.4716</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>106.5132</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6B (Block 2): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 2. An analysis of Table 62 reveals that subjects spent an average of 1.9500% of their total fixation time in Block 2 of the authentic text version and 0.4477% of their total fixation time in the same block of the experimental text version. The ANCOVA detected a significant difference between these two groups for Block 2 of "Im Volksgarten," $F = 4.27, p < 0.0440$, leading to a rejection of the null hypothesis (Table 63).

Figure 8 illustrates the nature of the main effect due to type of referencing device as indicated by the percentage of total fixation times predicted by the ANCOVA. This figure indicates that regardless of their baseline German
Figure 8: Main Effect of Type of Referencing Device as Measured By Predicted Percentage of Total Fixation Time in Block 2 of "Im Volksgarten"
language ability, subjects allocated a greater percentage of their total fixation time to Block 2 of the authentic text version of “Im Volksgarten,” which contained the gendered anaphoric reference “sie” [she], than they did to Block 2 of the experimental version, which contained the repeated noun “Anna.” This finding reveals that more fixation time was needed to process the less explicit anaphoric reference than was needed to interpret the unambiguous repeated noun.

Hypothesis 6C (Block 2): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 2. The results of the ANCOVA revealed no significant differences among subjects’ percentage of total fixation times in Block 2 of “Im Volksgarten” as a function of their baseline German language ability, $F = 1.49, p > 0.2276$ (Table 63). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 3): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 3. In Block 3 of the text, “Im Volksgarten,” the accusative case singular, masculine personal pronoun “ihn” [it] in the authentic text version was compared to the noun phrase “den Ballon” [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 3, $F(1, 50) = 3.59, p > 0.0638$ (Table 65).
Table 64
Means and Standard Deviations of Percentage of Total Fixation Time in Block 3 of “Im Volksgarten” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>1.9912</td>
<td>0.9949</td>
</tr>
<tr>
<td>Experimental²</td>
<td>1.9288</td>
<td>1.0081</td>
</tr>
<tr>
<td>Overall</td>
<td>1.9623</td>
<td>0.9920</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29

Hypothesis 6B (Block 3): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 3. An analysis of Table 64 reveals that subjects spent an average of 1.9912% of their total fixation time in Block 3 of the authentic text version and 1.9288% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 3 of “Im Volksgarten,” F = 3.31, p > 0.0747, leading to a retention of the null hypothesis (Table 65).

Hypothesis 6C (Block 3): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 3. The results of the ANCOVA revealed no significant differences among
subjects' percentage of total fixation times in Block 3 of "Im Volksgarten" as a function of their baseline German language ability, $F = 2.67$, $p > 0.1083$ (Table 65). This finding led to a retention of the null hypothesis of no difference.

**Table 65**
Analysis of Covariance of Percentage of Total Fixation Time in Block 3 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>3.0435</td>
<td>3.0435</td>
<td>3.31</td>
<td>0.0747</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>2.4552</td>
<td>2.4552</td>
<td>2.67</td>
<td>0.1083</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>3.2996</td>
<td>3.2996</td>
<td>3.59</td>
<td>0.0638</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>45.9114</td>
<td>0.9182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>52.1532</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 4): *There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 4.* In Block 4 of the text, "Im Volksgarten," the accusative case singular, masculine personal pronoun "ihn" [it] in the authentic text version was compared to the noun phrase "den Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for
percentage of total fixation time in Block 4, $F (1, 50) = 0.37, p > 0.5454$ (Table 67).

Hypothesis 6B (Block 4): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 4. An analysis of Table 66 reveals that subjects spent an average of 1.8208% of their total fixation time in Block 4 of the authentic text version and 1.5658% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 4 of “Im Volksgarten,” $F = 0.76, p > 0.3862$, leading to a retention of the null hypothesis (Table 67).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>1.8208</td>
<td>1.0120</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>1.5658</td>
<td>1.1848</td>
</tr>
<tr>
<td>Overall</td>
<td>1.7028</td>
<td>1.0924</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$
Hypothesis 6C (Block 4): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 4. The results of the ANCOVA revealed no significant differences among subjects’ percentage of total fixation times in Block 4 of “Im Volksgarten” as a function of their baseline German language ability, $F = 0.00$, $p > 0.9635$ (Table 67). This finding led to a retention of the null hypothesis of no difference.

Table 67
Analysis of Covariance of Percentage of Total Fixation Time in Block 4 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.9461</td>
<td>3.0435</td>
<td>0.76</td>
<td>0.3862</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0026</td>
<td>2.4552</td>
<td>0.00</td>
<td>0.9635</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.4589</td>
<td>3.2996</td>
<td>0.37</td>
<td>0.5454</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>61.9106</td>
<td>1.2382</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>63.2428</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 5): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 5. In Block 5 of the text, “Im Volksgarten,” the nominative case singular, feminine personal pronoun “sie”
[she] in the authentic text version was compared to the noun “Anna” [Anna] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 5, $F(1, 50) = 0.65$, $p > 0.4224$ (Table 69).

Hypothesis 6B (Block 5): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 5.

An analysis of Table 68 reveals that subjects spent an average of 2.0162% of their total fixation time in Block 5 of the authentic text version and 1.9971% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 5 of “Im Volksgarten,” $F = 0.55$, $p > 0.4633$, leading to a retention of the null hypothesis (Table 69).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>1.9971</td>
<td>2.3330</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>2.0162</td>
<td>1.3600</td>
</tr>
<tr>
<td>Overall</td>
<td>2.0059</td>
<td>1.9269</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$
Hypothesis 6C (Block 5): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 5. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 5 of "Im Volksgarten" as a function of their baseline German language ability, $F = 0.00, p > 0.9692$ (Table 69). This finding led to a retention of the null hypothesis of no difference.

Table 69
Analysis of Covariance of Percentage of Total Fixation Time in Block 5 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>2.1223</td>
<td>2.1223</td>
<td>0.55</td>
<td>0.4633</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0058</td>
<td>0.0058</td>
<td>0.00</td>
<td>0.9692</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>2.5420</td>
<td>2.5420</td>
<td>0.65</td>
<td>0.4224</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>194.2416</td>
<td>3.8848</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>196.7899</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6A (Block 6): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 6. In Block 6 of the text, "Im Volksgarten," the dative case singular, masculine personal pronoun "ihm" [it] in
the authentic text version was compared to the noun phrase “dem Ballon” [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 6, $F(1, 50) = 0.01$, $p > 0.9103$ (Table 71).

Table 70
Means and Standard Deviations of Percentage of Total Fixation Time in Block 6 of “Im Volksgarten” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>2.2171</td>
<td>1.3608</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>1.9714</td>
<td>1.0487</td>
</tr>
<tr>
<td>Overall</td>
<td>2.1033</td>
<td>1.2213</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$

Hypothesis 6B (Block 6): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 6. An analysis of Table 70 reveals that subjects spent an average of 2.2171% of their total fixation time in Block 6 of the authentic text version and 1.9714% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for
Block 6 of "Im Volksgarten," $F = 0.03$, $p > 0.8675$, leading to a retention of the null hypothesis (Table 71).

Hypothesis 6C (Block 6): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 6. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 6 of "Im Volksgarten" as a function of their baseline German language ability, $F = 0.25$, $p > 0.6203$ (Table 71). This finding led to a retention of the null hypothesis of no difference.

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0438</td>
<td>0.0438</td>
<td>0.03</td>
<td>0.8675</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.3868</td>
<td>0.3868</td>
<td>0.25</td>
<td>0.6203</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0200</td>
<td>0.0200</td>
<td>0.01</td>
<td>0.9103</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>77.8258</td>
<td>1.5565</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>79.0576</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 6A (Block 7): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 7. In Block 7 of the text, "Im Volksgarten," the nominative case singular, masculine personal pronoun "er" [it] in the authentic text version was compared to the noun phrase "der Ballon" [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 7, $F (1, 50) = 0.68, p > 0.4126$ (Table 73).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>2.0794</td>
<td>0.8530</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>1.6178</td>
<td>1.3336</td>
</tr>
<tr>
<td>Overall</td>
<td>1.8315</td>
<td>1.1153</td>
</tr>
</tbody>
</table>

$N = 54,$ cell size$^1 = 25; $cell size$^2 = 29$

Hypothesis 6B (Block 7): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 7.
An analysis of Table 72 reveals that subjects spent an average of 2.0794% of their total fixation time in Block 7 of the authentic text version and 1.6178% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 7 of "Im Volksgarten," $F = 1.73$, $p > 0.1944$, leading to a retention of the null hypothesis (Table 73).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>2.1348</td>
<td>2.1348</td>
<td>1.73</td>
<td>0.1944</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.4399</td>
<td>0.4399</td>
<td>0.36</td>
<td>0.5531</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.8423</td>
<td>0.8423</td>
<td>0.68</td>
<td>0.4126</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>61.6849</td>
<td>1.2337</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>65.9211</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 7): *There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 7.* The results of the ANCOVA revealed no significant differences among subjects’ percentage of total fixation times in Block 7 of "Im Volksgarten" as a
function of their baseline German language ability, $F = 0.36, p > 0.5531$ (Table 73). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 8): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 8. In Block 8 of the text, “Im Volksgarten,” the accusative case singular, masculine personal pronoun “ihn” [it] in the authentic text version was compared to the noun phrase “den Ballon” [the balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 8, $F (1, 50) = 0.02, p > 0.8998$ (Table 75).

Table 74
Means and Standard Deviations of Percentage of Total Fixation Time in Block 8 of “Im Volksgarten” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic$^1$</td>
<td>1.5796</td>
<td>0.9527</td>
</tr>
<tr>
<td>Experimental$^2$</td>
<td>1.2628</td>
<td>0.7251</td>
</tr>
<tr>
<td>Overall</td>
<td>1.4329</td>
<td>0.8620</td>
</tr>
</tbody>
</table>

$N = 54$, cell size$^1 = 25$; cell size$^2 = 29$
Hypothesis 6B (Block 8): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 8. An analysis of Table 74 reveals that subjects spent an average of 1.5796% of their total fixation time in Block 8 of the authentic text version and 1.2628% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 8 of "Im Volksgarten," $F = 0.13$, $p > 0.7241$, leading to a retention of the null hypothesis (Table 75).

Table 75
Analysis of Covariance of Percentage of Total Fixation Time in Block 8 of "Im Volksgarten" by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.0903</td>
<td>0.0903</td>
<td>0.13</td>
<td>0.7241</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>2.1726</td>
<td>2.1726</td>
<td>3.03</td>
<td>0.0879</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.02</td>
<td>0.8998</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>35.8623</td>
<td>0.7172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>39.3829</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 8): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in
Block 8. The results of the ANCOVA revealed no significant differences among subjects’ percentage of total fixation times in Block 8 of “Im Volksgarten” as a function of their baseline German language ability, $F = 3.03, p > 0.0879$ (Table 75). This finding led to a retention of the null hypothesis of no difference.

Hypothesis 6A (Block 9): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 9. In Block 9 of the text, “Im Volksgarten,” the dative case singular, feminine personal pronoun “ihre” [she] in the authentic text version was compared to the noun phrase “dem Mädchen” [the girl] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was rejected for percentage of total fixation time in Block 9, $F (1, 50) = 8.17, p < 0.0062$ (Table 77).

Figure 9 portrays the nature of the interaction between type of referencing device and baseline German language ability as indicated by the percentage of total fixation times predicted by the ANCOVA. This figure reveals that the predicted percentage of total fixation time for the repeated noun phrase “dem Mädchen” [the girl] in the experimental text version was equally low for subjects at all levels of baseline German language ability. This result indicates that the unambiguous noun phrase was quickly processed by both low and high ability students. In contrast, the gendered pronoun “ihre” [her] required a much greater predicted percentage of total fixation time from readers with low baseline German language ability scores than from those with high ability scores. This finding suggests that the less explicit pronoun reference was more difficult for
the low ability subjects to process than for subjects with high baseline German language skills.

Table 76
Means and Standard Deviations of Percentage of Total Fixation Time in Block 9 of “Im Volksgarten” as a Function of Type of Referencing Device

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic¹</td>
<td>1.6787</td>
<td>1.4567</td>
</tr>
<tr>
<td>Experimental²</td>
<td>0.9048</td>
<td>0.7909</td>
</tr>
<tr>
<td>Overall</td>
<td>1.3204</td>
<td>1.2474</td>
</tr>
</tbody>
</table>

N = 54, cell size¹ = 25; cell size² = 29

Hypothesis 6B (Block 9): There is no significant difference due to type of referencing device as measured by average fixation duration in Block 9. An analysis of Table 76 reveals that subjects spent an average of 1.6787% of their total fixation time in Block 9 of the authentic text version and 0.9048% of their total fixation time in the same block of the experimental text version. The ANCOVA detected a significant difference between these two groups for Block 9 of “Im Volksgarten,” $F = 12.91$, $p < 0.0007$, leading to a rejection of the null hypothesis (Table 77).
Figure 9: Interaction Between Type of Referencing Device and Baseline German Language Ability as Measured by Predicted Percentage of Total Fixation Time in Block 9 of "Im Volksgarten"
The nature of the main effect due to type of referencing device as indicated by the percentage of total fixation times predicted by the ANCOVA is shown in Figure 9. This main effect must be interpreted cautiously, however, in light of the significant interaction. The figure indicates that subjects allocated a greater percentage of their total fixation time to Block 9 of the authentic text version of "Im Volksgarten," which contained the gendered anaphoric reference "ihr" [her], than they did to Block 9 of the experimental version, which contained the repeated noun "dem Mädchen" [the girl]. This finding indicates that more fixation time was needed to process the less explicit anaphoric reference than was needed to interpret the unambiguous repeated noun phrase. The significant interaction reported above, however, shows that the predicted percentage of fixation time also varied according to subjects' baseline German language ability. No definitive conclusions, therefore, can be drawn about the effect due solely to type of referencing device.

Hypothesis 6C (Block 9): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 9. The results of the ANCOVA revealed significant differences among subjects' percentage of total fixation times in Block 9 of "Im Volksgarten" as a function of their baseline German language ability, $F = 7.64, p < 0.0080$ (Table 77). This finding led to a rejection of the null hypothesis of no difference.

Figure 9 illustrates the nature of the main effect due to baseline German language ability as indicated by the percentage of total fixation times predicted by the ANCOVA. This main effect must also be interpreted cautiously because of the significant interaction. The figure demonstrates that as baseline German
language ability scores increased, the predicted percentage of total fixation time tended to decrease. This result indicates that as readers became more proficient in their knowledge of the language they needed a lower percentage of total fixation time to process the information in Block 9 of the text. The significant interaction reported above, however, shows that the predicted percentage of fixation time also varied according to the text version that was read. No definitive conclusions, therefore, can be drawn about the effect due solely to baseline German language ability.

Table 77
Analysis of Covariance of Percentage of Total Fixation Time in Block 9 of “Im Volksgarten” by Type of Referencing Device

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>14.3425</td>
<td>14.3425</td>
<td>12.91</td>
<td>0.0007</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>8.4872</td>
<td>8.4872</td>
<td>7.64</td>
<td>0.0080</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>9.0694</td>
<td>9.0694</td>
<td>8.17</td>
<td>0.0062</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>55.5338</td>
<td>1.1107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>82.4705</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 6A (Block 10): There is no significant interaction between type of referencing device and baseline German language ability as measured by percentage of total fixation time in Block 10. In Block 10 of the text, "Im Volksgarten," the accusative case singular, masculine demonstrative pronoun "diesen" [this (one)] in the authentic text version was compared to the noun phrase "diesen Ballon" [this balloon] that appeared in the experimental text version. The null hypothesis of no interaction between the two independent variables was retained for percentage of total fixation time in Block 10, \( F (1, 50) = 2.23, p > 0.1413 \) (Table 79).

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authentic(^{1})</td>
<td>0.4931</td>
<td>0.6051</td>
</tr>
<tr>
<td>Experimental(^{2})</td>
<td>0.4461</td>
<td>0.4688</td>
</tr>
<tr>
<td>Overall</td>
<td>0.4714</td>
<td>0.5418</td>
</tr>
</tbody>
</table>

N = 54, cell size\(^{1}\) = 25; cell size\(^{2}\) = 29

Hypothesis 6B (Block 10): There is no significant difference due to type of referencing device as measured by percentage of total fixation time in Block 10.
An analysis of Table 78 reveals that subjects spent an average of 0.4931% of their total fixation time in Block 10 of the authentic text version and 0.4461% of their total fixation time in the same block of the experimental text version. The ANCOVA failed to detect a significant difference between these two groups for Block 10 of "Im Volksgarten," $F = 1.60, p > 0.2120$, leading to a retention of the null hypothesis (Table 79).

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Referencing Device</td>
<td>1</td>
<td>0.4718</td>
<td>0.4718</td>
<td>1.60</td>
<td>0.2120</td>
</tr>
<tr>
<td>Baseline Ability</td>
<td>1</td>
<td>0.0700</td>
<td>0.0700</td>
<td>0.24</td>
<td>0.6285</td>
</tr>
<tr>
<td>Referencing Device X Baseline Ability</td>
<td>1</td>
<td>0.6597</td>
<td>0.6597</td>
<td>2.23</td>
<td>0.1413</td>
</tr>
<tr>
<td>Within Cell</td>
<td>50</td>
<td>14.7619</td>
<td>0.2952</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>53</td>
<td>15.5555</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Hypothesis 6C (Block 10): There is no significant difference due to baseline German language ability as measured by percentage of total fixation time in Block 10. The results of the ANCOVA revealed no significant differences among subjects' percentage of total fixation times in Block 10 of "Im Volksgarten" as a
function of their baseline German language ability, $F = 0.24, p > 0.6285$ (Table 79). This finding led to a retention of the null hypothesis of no difference.

Table 80 shows a summary of the results of the analyses of covariance for percentage of total fixation time in blocks for "Im Volksgarten."

<table>
<thead>
<tr>
<th>Block</th>
<th>Interaction</th>
<th>Type of Referencing Device</th>
<th>Baseline German Language Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>(ihm/dem Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 2</td>
<td>(sie/Anna)</td>
<td>NS</td>
<td>S</td>
</tr>
<tr>
<td>Block 3</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 4</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 5</td>
<td>(sie/Anna)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 6</td>
<td>(ihn/dem Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 7</td>
<td>(er/der Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 8</td>
<td>(ihn/den Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Block 9</td>
<td>(ihr/dem Mädchen)</td>
<td>S</td>
<td>S</td>
</tr>
<tr>
<td>Block 10</td>
<td>(diesen/diesen Ballon)</td>
<td>NS</td>
<td>NS</td>
</tr>
</tbody>
</table>

S (significant), NS (nonsignificant)
Phase III of the Study

Phases I and II of the study assessed subjects' processing and comprehension of two short German texts. In Phase III of the study, students' performance on a longer German text was analyzed. Readers' comprehension of the longer text was measured via written immediate recall protocols. Their ability to identify correctly the coreferents of specific anaphoric expressions within the text was also evaluated. The relationships between coreferential tie comprehension, overall comprehension, and baseline German language ability were examined using the Pearson product-moment correlation procedure. In addition, students' coreferential tie errors were analyzed qualitatively to determine how discourse features, selection strategies, and baseline language ability influence the resolution process.

Correlational Analysis

Hypothesis 7A: There is no significant relationship between coreferential tie comprehension and overall comprehension. A Pearson product-moment procedure produced a correlation coefficient of 0.6161, p < 0.0001, between coreferential tie comprehension and overall comprehension (Table 81). This finding revealed that there is a significant relationship between coreferential tie comprehension and overall comprehension for readers of German and led to a rejection of the null hypothesis. The positive correlation of 0.6161 indicates that the ability to interpret anaphora correctly accounts for 37.957% of the variance in the subjects' overall comprehension scores. The high degree of common variance between the two variables suggests that subjects' coreferential tie comprehension scores are good predictors of their overall comprehension.
scores.

Table 81
Pearson Product-Moment Correlations between Coreferential Tie Comprehension, Overall Comprehension, and Baseline German Language Ability

<table>
<thead>
<tr>
<th></th>
<th>Coreferential Tie Comprehension</th>
<th>Overall Comprehension</th>
<th>Baseline German Language Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coreferential Tie Comprehension</td>
<td>1.0000</td>
<td>0.6161</td>
<td>0.7833</td>
</tr>
<tr>
<td>Overall Comprehension</td>
<td></td>
<td>1.0000</td>
<td>0.4998</td>
</tr>
<tr>
<td>Baseline German Language Ability</td>
<td></td>
<td></td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Hypothesis 7B: There is no significant relationship between coreferential tie comprehension and baseline German language ability. A Pearson product-moment procedure produced a correlation coefficient of 0.7833, \( p < 0.0001 \), between coreferential tie comprehension and baseline German language ability (Table 81). This finding revealed that there is a significant relationship between coreferential tie comprehension and baseline German language ability for readers of German and led to a rejection of the null hypothesis. The positive correlation of 0.7833 indicates that subjects' baseline German
language ability accounts for 61.356% of the variance in the their coreferential tie comprehension scores. The high degree of common variance between the two variables suggests that subjects’ baseline German language ability scores are good predictors of their coreferential tie comprehension scores.

Hypothesis 7C: There is no significant relationship between overall comprehension and baseline German language ability. A Pearson product-moment procedure produced a correlation coefficient of 0.4998, p < 0.0001, between overall comprehension and baseline German language ability (Table 81). This finding revealed that there is a significant relationship between subjects’ overall comprehension of a German text and their baseline German language ability and led to a rejection of the null hypothesis. The positive correlation of 0.4998 indicates that subjects’ baseline German language ability accounts for 24.980% of the variance in their overall comprehension scores. The moderate degree of common variance between the two variables suggests that subjects’ baseline German language ability scores are of some value in predicting their overall comprehension scores.

Qualitative Analysis

The coreferential tie errors made by students were examined to determine how specific discourse features in the text and how baseline German language ability might influence subjects’ selection of antecedents. The percentages of subjects who gave incorrect responses on the coreferent selection task are illustrated in Figures 10 and 11. These bar graphs indicate which of the 31 anaphoric references investigated in the longer German text were the most
problematic for readers. The following analysis focuses on the types of coreferential tie errors made most frequently by subjects. Particular emphasis was given to the kinds of selection strategies students applied incorrectly or failed to apply at all in their search for appropriate coreferents. Also of interest was the language ability profile of those students making coreferential tie errors. This analysis reveals how the antecedent-anaphor relationship, specific discourse features, and baseline German language ability impact coreferential tie comprehension. In addition, an examination of these errors provides insights into why readers fail to achieve coreferential tie comprehension in German.

There were six different types of anaphoric references investigated in the coreferent selection task: personal pronouns, relative pronouns, indefinite pronouns, demonstrative pronouns, demonstrative noun phrases, and possessive adjectives. These different types of anaphoric expressions are analyzed in separate sections that follow.

**Personal Pronouns**

Of the 31 anaphoric expressions examined in the coreferent selection task, 12 of them were singular, gendered personal pronouns (Figures 10 and 11). Six of the singular personal pronouns were neuter in gender. The singular, neuter personal pronoun “es” [it] was used three separate times to refer to the noun “Europa” [Europe]. The first two instances of this pronoun proved to be relatively unproblematic for the readers (Items 4 and 5 in Figure 10). Only 5.556% of the readers misinterpreted these two anaphoric references and those readers, as a group, had below average baseline German language
Figure 10: Percentage of Incorrect Responses for Anaphoric Expressions 1-15 of "Europa Heute"
Figure 11: Percentage of Incorrect Responses for Anaphoric Expressions 16-31 of “Europa Heute”
ability scores. In other words, the mean score for this group on the AATG National German Examination for High School Students was less than 41.222 points out of 60, the average score for the entire sample of subjects.

One reason for the high success rate overall may be that students favored an antecedent that was the subject of its clause, just as the pronoun “es” [it] was the subject of its clause. In other words, it seems likely that many subjects correctly employed the parallel function strategy (Grober, et al., 1978; Sheldon, 1974). Another feasible explanation for such few errors is that the intended antecedent, “Europa” [Europe], was thematically prominent in the discourse and, thus, easier for the subjects to locate. This result corroborates previous research findings that claim that a pronoun is easier to comprehend if its antecedent is foregrounded (Frederikson, 1981; Kintsch & Vipond, 1978).

Further support for this contention comes from the results for the third instance of the pronoun “es” [it] used to refer to “Europa” [Europe] (Item 7 in Figure 10). This anaphoric reference was much more difficult for readers to interpret, as the 42.593% of incorrect responses attests. Many subjects chose the noun phrase “dieser wunderbar reiche, feingegliederte Kontinent” [this wonderfully rich, finely constructed continent] as the antecedent for “es,” [it], in spite of the fact that “der Kontinent” [the continent] is a masculine noun and would be referred to anaphorically with the masculine personal pronoun “er” [it]. This finding is perhaps not surprising given that this group of subjects had an average baseline German language ability score of 32.261, indicating that as a whole they had below average German language skills. This lack of skills may have made them less sensitive to important syntactic cues, like gender. The noun phrase “ein Überbleibsel” [a remnant] was also a popular response and
did meet the syntactic requirement of being a singular, neuter noun. This choice, however, made less sense semantically in the sentence.

It seems that the proximity of candidate antecedents was an important consideration for these below average students in their choice of a coreferent for “es” [it]. In both cases, the incorrectly selected antecedents were either in the same line or only one line before the occurrence of the pronoun “es” [it]. In contrast, the correct antecedent, “Europa” [Europe], was presented four lines before the pronoun was used. To make matters even more complicated for the readers, the noun phrases, “dieser ... Kontinent” [this ... continent] and “ein Überbleibsel” [a remnant], were themselves definite noun phrases used to describe “Europa” [Europe]. Apparently, the concept of “Europa” [Europe] was already backrounded for these readers and replaced in memory by the more recently mentioned descriptive noun phrases by the time they encountered the pronoun “es” [it] in the text. This backrounding presumably made it more difficult for these subjects to access the intended coreferent and, thus, diminished their coreferential tie comprehension.

The singular, neuter pronoun “es” [it] was also used to refer to the noun phrase “das berühmte Wort” [the famous saying], listed as Item 3 in Figure 10. This pronoun was misinterpreted by roughly one third of the subjects (31.481%). As a whole, this group of students was also found to have below average baseline German language ability (mean baseline German language ability score = 35.467). Of those who missed the item, four believed “es” [it] referred to “der Dritte Stand” [the Third Estate], even though all of these readers correctly identified the singular, masculine pronoun “er” [it] as also referring to “der Dritte Stand” [the Third Estate]. As below average students, this group may
not have paid close enough attention to the gender of the pronoun and its potential antecedents to notice this discrepancy or did not deem the gender difference important enough to affect their choice of coreferents. Both possibilities indicate a lack of sensitivity to syntactic constraints imposed by grammatical gender.

Several other subjects who interpreted "es" [it] incorrectly determined that the pronoun had a nongender specific coreferent, such as "etwas" [something] or "alles" [everything], presented earlier in the paragraph. The remaining readers believed that "es" [it] referred to the preceding "series of questions and answers," or "the whole quote," as they indicated in their coreferent selection tasks. These subjects had a sense of what the pronoun was used to refer to, but could not clearly identify its intended antecedent. As a result, their understanding of this anaphoric reference was less than precise.

Late in the final paragraph of the text, the singular, neuter pronoun "es" [it] was also used twice to refer to the neuter noun "Amerika" [America] (Items 30 and 31 in Figure 11). Of all the anaphoric references in the passage, this one proved to be the most problematic for readers. The first instance of the pronoun was misinterpreted by 61.111% of the subjects and the second instance was missed by 64.815%. Although the mean baseline German language ability score was below average for this group of subjects, both low and high ability students were well represented in the group (mean baseline German language ability score = 37.771). A common misconception among lower ability students was that "es" [it] referred to "Rußland" [Russia], which occurred in the preceding line and agreed with the pronoun in number and gender. Background knowledge or beliefs about Russia could have led these readers to this
interpretation. Given the context, it is quite possible that readers concluded that “Rußland” [Russia] was the one that “wirkt und waltet” [operates and governs] and “droht” [threatens].

Higher ability readers, on the other hand, tended to interpret “es” [it] as “das Gleichgewicht” [the balance], which also appeared in the preceding line and agreed with the pronoun in number and gender. Selecting an appropriate antecedent may have been made more complicated by the fact that “das Gleichgewicht” [the balance] was a definite noun phrase used to describe “Amerika” [America], so the subjects may have believed that they were interchangeable as coreferents. Semantically, this choice is not logical, however, because “das Gleichgewicht” [the balance] can not “wirkt und waltet” [operate and govern] or “droht” [threaten] whereas “Amerika” [America], the correct antecedent, can do these things.

In this case as before, the intended coreferent, “Amerika” [America], appeared four lines prior to the use of the anaphoric expression “es” [it]. The considerable distance between the pronoun and the antecedent seems to have caused the correct antecedent to have become backgrounded while pushing other candidate coreferents to the foreground. This finding also illustrates the tendency on the part of these readers to select just-mentioned noun phrases as antecedents and an unwillingness or inability to search beyond the immediate proximity when choosing coreferents.

The coreferent selection task also included the singular, neuter personal pronoun “ihm” [it], which was used to reflect a dative case reference to “das russische Herrschaftssystem” [the Russian system of sovereign rule]. This pronoun was difficult for roughly one third of the subjects to interpret correctly
(35.185%). These subjects had an average baseline German language ability score of 32.842, indicating that as a group they had below average German language skills. The most common misinterpretation among these readers was that “ihm” [it] referred to “Russland” [Russia] in general, instead of to its “Herrschaftssystem” [system of sovereign rule]. This antecedent choice did at least match the intended coreferent in gender and number. It also illustrates that the readers were able to identify a thematically prominent concept in the discourse, but as a coreferent, it was too vague. These subjects obviously got the gist of the author’s intention, but they failed to identify his specific meaning.

Another frequent misconception was that “ihm” [it] was actually “ihm” [he] referring to “Hitler” or “Stalin.” This interpretation was grammatically plausible, as both neuter and masculine nouns can be referred to anaphorically by the personal pronoun “ihm” in dative case. Such an interpretation, however, was not semantically feasible and seemed odd given that both of these proper names were mentioned much earlier in the text and would presumably have been already backgrounded. Apparently, subjects interpreted the pronoun “ihm” as strictly masculine and then searched for an antecedent whose natural gender complied with that criterion, even if that meant ignoring foregrounded concepts. This result suggests that native readers of English with low baseline German language ability are inclined to interpret gendered pronouns in German as referring to animate, gendered entities because this is the common practice in their first language.

A similar pattern of misinterpretations appeared for the four singular, masculine personal pronouns that were investigated in the coreferent selection task. The masculine personal pronoun “er” [it] was used twice in the first
paragraph to refer to the noun phrase "der Dritte Stand" [the Third Estate] (Items 1 and 2 in Figure 10). In both instances, only 9.259% of the subjects misunderstood the pronoun, indicating that it was relatively easy for subjects to resolve correctly. Among those subjects who missed this item, however, the most popular incorrect interpretation for "er" [it] was "Abbe Sieyès," the name of the Frenchman described in the first paragraph. In addition, those subjects misinterpreting the item in this way had a very low mean baseline German language ability score of only 16.5. This finding lends further credence to the premise that readers with low German language skills deliberately seek animate, gendered antecedents for gendered pronouns, even if those antecedents make little semantic sense in the context.

Even more support for this contention comes from subjects' interpretations of the singular, masculine personal pronoun "ihn" [it], used twice in the text to reflect accusative case references to "den zweiten Weltkrieg" [the second World War] (Items 23 and 25 in Figure 11). The first instance of this anaphor occurred two thirds of the way through the second paragraph and directly after its intended antecedent. Despite the nearness of the anaphor to its antecedent, a majority of readers (51.852%), nonetheless, misinterpreted this anaphoric reference. The second instance of the pronoun "ihn" [it] occurred shortly thereafter in the text and slightly further away from its intended coreferent. It was also misunderstood by a large percentage of readers (44.444%).

Those subjects who missed one or both of these two items had below average German language skills, as indicated by an average baseline German language ability score of 30.036. Of those students misconstruing these pronoun references, 46.429% believed that "Hitler" was the correct coreferent
for "ihn" [it]. In these two instances, readers specifically looked for an animate, masculine noun to be the coreferent for the gendered pronoun "ihn." The most obvious and accessible antecedent to meet this criterion was the masculine person "Hitler," which was mentioned earlier in the same sentence. These findings imply that second language readers, particularly those with low ability, are prone to resort to inappropriate first language reading strategies when attempting to identify the coreferents of second language anaphoric references.

Other common incorrect responses for these two anaphoric expressions included such things as: leaving the items blank or putting a question mark above them, and interpreting the singular, masculine anaphor "ihn" [it] as the singular, feminine noun "die Art" [the way] or as the plural noun "die Angelsachsen" [the Anglo-Saxons]. These errors indicate that some students experienced confusion and frustration with these items, while others failed to consider gender or number details in their search for a suitable coreferent. Overall, these subjects appeared unable to use key syntactic and semantic clues inherent in the anaphor or in the surrounding discourse to locate the intended antecedent successfully.

In addition to singular, masculine and neuter pronouns, the coreferent selection task also investigated subjects' interpretation of the singular, feminine personal pronoun "sie" [it], used to refer to the noun phrase "die neue Machtstellung Rußlands" [the new political power position of Russia]. The results for this anaphor are presented as Item 28 in Figure 11. Subjects in this group had below average baseline German language ability scores (mean baseline German language ability score = 30.833). This pronoun proved troublesome for readers, in that 44.444% of them were unable to identify the
correct coreferent. Some subjects responded to this item by leaving it blank or putting question marks above the anaphor. This was an especially common response for students with lower baseline German language ability scores.

The majority of readers, however, misinterpreted the pronoun "sie" [it] as referring to the singular, neuter noun "Ruβlands" [(of) Russia], even though some of them also identified the neuter pronoun "es" [it] as referring to this noun. This group of students clearly ignored the anaphor's feminine gender marker or were unable to identify it as such. Not only did these readers fail to employ appropriate gender clues, they may also have misinterpreted proximity and foregrounded discourse features. The genitive noun form, "Ruβlands" [(of) Russia], followed the intended antecedent, "die neue Machtstellung" [the new political power], and, therefore, may have been placed in the foreground of the discourse for some students. Because it came after the intended coreferent, this noun was also located closer to the anaphor. Subjects strictly employing a proximity strategy would have identified "Ruβlands" [of Russia] as the most recently mentioned noun and, thus, selected it as a likely coreferent. These subjects, however, obviously failed to identify both the syntactic function and form of this noun. In so doing, they might have recognized that it served only to modify another noun phrase, the one to which the personal pronoun actually referred.

In addition to singular personal pronouns, there were four plural personal pronouns examined in the coreferent selection task. The plural personal pronoun "sie" [they] was used twice in the same sentence to refer to "die Mächte" [the powers] (Items 16 and 17 in Figure 11). This pronoun usage proved very troublesome for readers. Both instances of the pronoun were
misinterpreted by 44.444% of the subjects. This group of readers had an average baseline German language ability score of 31.208, indicating below average language abilities. A majority of the subjects who misinterpreted the pronoun claimed that the appropriate antecedent for “sie” [they] was “die Europäer” [the Europeans]. Although this noun phrase did match the anaphor in number and gender and was in close proximity to the anaphoric reference, it made no semantic sense given the context. With this interpretation, the sentence would read “...ohne uns Europäer wüßten sie (die Europäer) nicht..., wüßten sie (die Europäer) nicht einmal” [...without us Europeans, they (Europeans) would not know..., they (the Europeans) would not know at all].

The use of “uns” [us] by the author implies the sense of a we group and a they group, that can not be all inclusive. This implication makes it impossible for the pronoun “sie” [they] to refer back to the we group (“die Europäer” [the Europeans]). This below average group of students neglected to attend to this detail and searched instead for the nearest available antecedent, regardless of whether it fit into the context or not.

Other subjects who missed the item either left it blank or chose another noun located close to the pronoun as the coreferent, such as “die Freiheit” [the freedom] or “ein Echo” [an echo]. These antecedent choices syntactically did not meet the number restriction and semantically made little sense given the context. It seems that in these instances the proximity of candidate coreferents directly influenced students’ selections when other anaphor or discourse clues eluded them.

Yet another group of students who misinterpreted “sie” [they] strove to grasp the author’s intended meaning, but reached, perhaps, too far beyond the
text in their search for coreferents. This is illustrated in such interpretations of “sie” [they] as “die übrigen” [the remaining (countries)]; “die Welt” [the world], and “countries.” In each case, these readers were almost too willing to infer an antecedent based on their understanding of the passage and, as a result, failed to examine the text itself for a suitable coreferent. This strategy was used by both high and low ability students.

Another use of the plural personal pronoun “sie” [they] also proved problematic for readers. This pronoun was used twice in a sentence to refer to the noun “die Angelsachsen” [the Anglo-saxons] (Items 22 and 24 in Figure 11). The first instance of this pronoun was misinterpreted by 38.889% of the subjects and that percentage jumped to 48.148% for the second instance. The majority of readers who answered incorrectly chose to leave either one or both of the items blank, indicating that they could not even guess at an appropriate coreferent for the pronoun.

Other incorrect interpretations, such as “Hitler and Stalin,” “Lenin, Hitler, Stalin,” “Zusammenspiel, Pakt und Krieg” [cooperation, pact and war] and “Hitlers Krieg und Niederlage” [Hitler’s war and defeat] revealed that subjects could identify “sie” [they] as a plural anaphor, but simply combined different nouns in the preceding discourse to form a plural antecedent. As a whole, these readers scored below average on the baseline German language ability test (mean baseline German language ability score = 29.000) These poor German language skills presumably made it more difficult for them to interpret these anaphoric references effectively in discourse.
Relative Pronouns

There were four relative pronouns included in the coreferent selection task, of which two had singular antecedents and two had plural antecedents. The singular, feminine relative pronoun “der” [that] was used to reflect a dative case reference to “die Art” [the way] (Item 21 in Figure 11). This anaphoric reference was misinterpreted by 40.741% of the subjects, implying that it was a fairly problematic anaphor for them. The great majority of readers who missed the item gave no response at all. Most students simply left the space above the relative pronoun blank, which suggests that they did not have any idea how to interpret it. A few other students put a question mark over the anaphor to indicate their confusion or frustration with the task.

One reason for the difficulty experienced by the subjects may be that the relative pronoun “der” [that] was in dative case, functioning as the object of the preposition “in.” In dative case the singular, feminine relative pronoun takes the same form as the singular, masculine relative pronoun in nominative case. Studies with first language learners of German have shown that sentences where the relative pronoun was the subject of its clause were easier to understand than sentences where the relative pronoun was the object of the clause (Grimm, et al., 1975; Mills; 1977a; Mills, 1986). It is possible, therefore, that some subjects sought a singular, masculine antecedent for what they thought was the subject of the relative clause. Not finding one in the immediate vicinity, these students simply left the item blank.

Further evidence of this strategy comes from the fact that most of the other students who misinterpreted the anaphoric expression chose a singular, masculine noun as the coreferent, even though they had to ignore more
recently mentioned nouns to find such an antecedent. These readers selected either “Hitler” or “ihren Teil” [their part] as antecedents because they met the singular, masculine criterion. The data also revealed that this group generally had below average baseline German language abilities (mean baseline German language ability score = 30.864). This finding suggests that students with poor second language skills in German approach relative pronoun resolution in much the same way that children learning German as a first language do. It also seems that these subjects were more influenced by the gender and number of the anaphor (at least what they believed the gender and number to be) than they were by the proximity of possible antecedents.

Another singular, feminine relative pronoun “die” [that] was used to refer to the noun phrase “der führenden europäischen Macht” [the leading European power] in the latter part of the passage. This anaphor was particularly problematic for students. Just over half of the readers missed this item on the coreferent selection task (51.852%), as illustrated by Item 29 in Figure 11. Approximately two thirds of this group had below average baseline German language ability scores (mean baseline German language ability score = 25.051), but the remaining one third had well above average language ability scores (mean baseline German language ability score = 53.000).

Among both low and high ability students the most popular antecedent choice by far was “Amerika” [America]. In fact, 93% of the group chose “Amerika” [America] as the coreferent for the relative pronoun “die” [that]. Semantically, the subjects’ selection made logical sense and captured the essential meaning of the sentence. The noun phrase “der führenden europäischen Macht” [the leading European power] had, in fact, been used to
describe “Amerika” [America] and could semantically be interpreted as the subject of the clause. In addition, the concept of “Amerika” [America] was clearly the prominent topic of the sentence and may have remained foregrounded despite the introduction of new concepts.

Syntactically, however, the choice of “Amerika” [America] as an antecedent was not feasible, because the relative pronoun “die” [that] is feminine and could not be used to refer to a neuter noun like “Amerika” [America]. Even more surprising is that subjects did not recognize that in German, as in English, “a relative clause follows the person or thing it refers to as closely as possible to avoid ambiguity” (Longman, 1988, p. 24). Using this basic rule, in addition to the gender constraint, should have reduced the set of candidate antecedents down to only one choice for the subjects. For the majority of readers, however, the intended coreferent, “der führenden europäischen Macht” [the leading European power], could not compete with the more prevalent concept, “Amerika” [America], during the resolution process.

In addition to singular relative pronoun references, the coreferent selection task also measured subjects’ ability to resolve plural relative pronouns. The task included, for example, the plural relative pronoun “welche” [which], which was used to refer to the noun “die Mächte” [the powers]. This anaphoric reference caused problems for one fourth of the readers (24.074%), as Item 14 in Figure 10 illustrates. Subjects who misinterpreted this anaphor had a mean baseline German language ability score of 26.385, indicating below average German language skills.

All but one of the readers in this group chose “unser letzter melancholischer Triumph” [our last melancholy Triumph] as the coreferent for
"welche" [which]. It is difficult to imagine what motivated subjects to choose this noun as the antecedent for "welche" [which]. Syntactically, "unser...Triumph" [our...Triumph] is a singular, masculine noun and, thus, could not be referred to anaphorically by a plural, feminine relative pronoun like "welche" [which]. Because the relative pronoun "welche" [which] serves as the subject of its clause, the verb in the clause must agree with this anaphor. The plural verb in the clause, "rufen" [call], therefore, provided these readers with an important syntactic clue that could have helped them to narrow the field of candidate antecedents. In addition, the correct antecedent "die Mächte" [the powers] was actually closer to the relative pronoun than was the competing antecedent "unser...Triumph" [our...Triumph]. Again, subjects seemed to ignore or to be unfamiliar with the proximity rule for relative pronouns noted by Longman (1988) above.

Semantically, the choice of "unser...Triumph" [our...Triumph] as a coreferent also makes little sense. It seems much less plausible that the relative clause in which "welche" [which] appears would read "our triumph, which calls today for the 'End of Europe'" than "the powers, which call today for the 'End of Europe." It is possible, however, given the low baseline German language skills of the subjects in this group, that they chose "unser...Triumph" [our...Triumph], because it is an English cognate and, thus, readily understandable to them. In this instance, it may be the readers' lack of basic vocabulary knowledge instead of specific discourse clues that caused them to exclude the intended coreferent from their antecedent search.

The final relative pronoun investigated in the coreferential selection task was the plural form "denen" [which], which was a dative case reference to "die
Ideen" [the ideas]. Unlike the plural relative pronoun “welche” [which] that did prove troublesome for some students, “denen” [which] caused them little overall difficulty. As Item 18 in Figure 11 demonstrates, only 11.111% of the readers identified this anaphor incorrectly. Two factors may have contributed to the high success rate for this item. First of all, the relative pronoun was preceded by a very short clause which contained only one possible antecedent, the intended one. By simply favoring the coreferent in the most recent clause, a common selection strategy, the reader could locate the correct antecedent easily. In addition, attending to the gender and particularly the number of the anaphor allowed subjects to constrain the set of candidate antecedents, because few preceding nouns or noun phrases met the plural criterion.

Of those students who did misinterpret the item, most attempted to apply the second strategy by selecting the plural noun phrase “Asien und Afrika” [Asia and Africa] as the antecedent. This group failed to employ the first strategy, however, in that they looked beyond the directly preceding clause for a suitable plural coreferent. This oversight took the students too far away from the intended coreferent and caused them to misinterpret the anaphor. It is possible that below average German language skills made it more difficult for the readers to identify “die Ideen” [the ideas] as the closest plural noun and, thus, a logical antecedent candidate (mean baseline German language ability score = 29.167).

**Indefinite Pronouns**

There was only one indefinite pronoun investigated in the coreferent selection task. The plural form “anderen” [others] was used to reflect a dative
case reference to “ein europäischer Staat” [a European state]. This anaphoric expression proved extremely difficult for readers to resolve successfully. More than half the subjects misinterpreted this anaphor (53.704), as Item 19 in Figure 11 demonstrates. This group of subjects had below average baseline German language ability scores as a whole, but roughly one quarter of the group had very high language ability scores (mean baseline German language ability score = 35.034). At first, it seemed that the fact that the anaphor was plural but its intended antecedent was singular would have thrown students off course and caused them to search for a plural antecedent. This, however, was not the case. The three most popular competing coreferents were all singular. Apparently, these readers were undisturbed by a plural reference to a singular noun. The plural anaphor may have caused another problem, however, in that it masked the gender of the antecedent. The plural forms of “andere” [other] in dative case are identical for all three genders. There was no way, therefore, for the students to eliminate possible antecedents on the basis of gender, as nouns from all three genders could be represented by this plural anaphor. As a result, among those subjects who missed the item, both neuter and feminine nouns were popular coreferent choices.

In addition, these readers obviously failed to use proximity to the anaphor as a selection strategy, as the noun closest to the anaphoric reference was, in fact, the intended antecedent. Instead, students chose to go back to the preceding clause to look for an appropriate coreferent. This clause provided readers with three different antecedent possibilities: “Rußland” [Russia], “ein barbarisches Land” [a barbaric land], and “der Grenze” [the border]. The first one, “Rußland” [Russia], was the preferred choice among this group of students.
As the subject of the sentence, it is possible that this noun was considered to be the most thematically prominent and, therefore, was selected as the antecedent. The other two nouns, “ein barbarisches Land” [a barbaric land] and “der Grenze” [the border] were also common selections, although it is less clear why readers chose them as coreferents. Semantically, the noun phrase “ein barbarisches Land” [a barbaric land], fits into the context of the sentence fairly well, unlike the noun “der Grenze” [the border]. Syntactically, however, both choices are found wanting because it is hard to imagine that the author would place such distance between the anaphor and its intended antecedent.

**Demonstrative Pronouns**

Only one demonstrative pronoun was examined in the coreferent selection task. The singular, masculine demonstrative pronoun “dieser” [this] was used to refer to the noun phrase “der technische, wirtschaftliche Aufstieg” [the technical, economic rise] (Item 27 in Figure 11). This anaphoric reference was misinterpreted by nearly one third of the subjects (31.481%). In that group, over 75% of the students had below average baseline German language ability scores. The overall mean score for the entire group was 30.412. The most common error made by readers was to identify “das russische Herrschaftssystem” [the Russian rule system] as the coreferent for “dieser” [this]. Choosing this neuter antecedent forced these readers to ignore the gender of the anaphor, which was distinctly masculine, and to search beyond the most recent clause for an appropriate coreferent. In addition, these students failed to employ the parallel function strategy, which would have encouraged them to select “der ... Aufstieg” [the ... rise] as the correct coreferent, because it serves
as the subject of the preceding clause just as “dieser” [this] acts as the subject of its clause (Grober, et al., 1978; Sheldon, 1974).

Another popular misinterpretation was the selection of only “der technische” [the technical] as the antecedent, particularly among very low ability subjects. Apparently, these students neglected to notice that the word “technische” [technical] was not capitalized, as all German nouns are, and that it had the characteristic adjective suffix “-ische.” These clues would have helped them to identify “technische” [technical] as just an adjective and, therefore, as an inappropriate coreferent for “dieser” [this]. The general rule of usage for demonstrative pronouns in German, similar to relative pronouns, is that they act “als Stellvertreter eines Substantives (+ Artikel)” [as substitutes in the place of a noun (+ article)] (Drosdowski, 1988, p.324). Once these readers identified an article that showed the same gender and number as the anaphor they may have simply assumed that it was followed by the noun antecedent they were seeking. The fact that this kind of error occurred with other noun antecedents preceded by lengthy adjective phrases indicates that this discourse feature was particularly problematic for readers.

**Demonstrative Noun Phrases**

There was only one demonstrative noun phrase among the 31 anaphoric expressions tested in the coreferent selection task. The lengthy noun phrase, “diese gesegnete Halbinsel aus Halbinseln” [this blessed peninsula consisting of peninsulas], was used to refer to “Europa” [Europe], the passage's central theme (Item 6 in Figure 10). This anaphoric reference was misinterpreted by 29.630% of the subjects. Every reader in this group had below average
baseline German language abilities (mean baseline German language ability score = 27.500). There were essentially three types of incorrect responses made by subjects who missed the item. One third provided no antecedent for the noun phrase at all or indicated their confusion by putting only a question mark above the item. Clearly, this group had no idea where to even look for an appropriate coreferent.

Another set of students interpreted the noun phrase as referring to the string of adjectives “wirtschaftlich, machtpolitisch, moralisch, und geistig” [economically, politically, morally, and spiritually], that appeared in the text just prior to the anaphor. Although more prevalent in this instance, this type of error was also made with some of the other anaphoric references, especially if the intended coreferent was preceded by a long adjective phrase. These readers seemed to have ignored a number of important discourse features that could have helped them to select the intended antecedent. First and foremost, they did not attend to the overall form of their selected coreferent. This factor alone would have helped them to see that this entire phrase was presented in lower case and, therefore, did not include a noun, which would have been capitalized. Second, these students disregarded the grammatical function of their chosen coreferent. Attending to this feature would have allowed them to recognize that the selected words were adjectives that could serve to modify a possible antecedent but could not themselves be an antecedent. Third, the subjects failed to search for a thematically prominent coreferent. This exercise would have led them back to the preceding sentence, the first sentence of the discourse, in which the theme “Europa” was conspicuously presented. Instead of utilizing these general techniques, the readers simply favored the contents of
the most recent clause even though that clause contained no plausible antecedents. This approach was perhaps not surprising given the poor German language skills of these students.

The third type of error made by students was to misconstrue "diese gesegnete Halbinsel aus Halbinseln" [this blessed peninsula consisting of peninsulas] as an anaphoric reference for "das Zentrum der Erde" [the center of the earth] or just "der Erde" [(of) the earth]. In choosing a coreferent, these readers simply favored the most recently mentioned noun or noun phrase. In so doing, however, they neglected to look for a thematically prominent antecedent or to use their general knowledge of the antecedent and the anaphor to help them choose a more appropriate coreferent. In other words, had these readers noticed that "Europe" was the main theme of the discourse and that it would be unusual to refer to "the center of the world" or "the world" alone as a "blessed peninsula consisting of peninsulas," it might have been easier for them to locate the intended antecedent. It seems likely, however, that a deficit in basic German language skills thwarted the efforts of these students to use these more appropriate coreferent selection techniques.

**Possessive Adjectives**

There were eight possessive adjectives investigated in the coreferent selection task, of which six were singular and two were plural. Each of the six singular possessive adjectives was an inflected form of the neuter anaphor "sein" [its] and all were used to refer to the noun "Europa" [Europe]. These anaphoric references are listed as Items 8 to 13 in Figure 10. This cluster of possessive adjectives caused few problems for the subjects. Percentages of
incorrect responses ranged from only 9.259% to 12.963%, which were among the lowest percentages obtained for any item in the task. Apparently, most readers were able to identify the intended antecedents quite easily. A feasible explanation for this finding is that these anaphoric references referred to a coreferent, "Europa" [Europe], that was the topic of the discourse and intentionally foregrounded. These anaphoric expressions, therefore, adhered to the *Aboutness Principle of Anaphora* (APA) as described by Bosch (1983). This principle states that "an anaphorically used expression refers to an object which the discourse at the relevant moment is about" (Bosch, 1983, p. 20). As a result, the intended antecedent was made more accessible to the readers and allowed most of them to locate it correctly. The small minority of subjects who did misinterpret these items tended to have very low baseline German language ability scores (mean baseline German language ability score = 25.538). This finding suggests that not only conspicuous discourse features but also basic language skills can facilitate the processing of anaphoric references.

One of the two plural possessive adjectives, "unserer" [our], also referred to "Europa" [Europe] or more specifically "Europäer" [Europeans] (Item 15 in Figure 10). This anaphor was misinterpreted by considerably more subjects than the possessive adjective "sein" [its]. Roughly one third of the readers were unable to identify the intended antecedent for "unserer" [our] (31.481%). Of this group, the majority had very low baseline German language ability scores (mean baseline German language ability score = 30.941). There were no definitive patterns among the incorrect responses given by subjects. Common inappropriate antecedents included: "ein Echo" [an echo], which directly preceded the anaphor; "Schöpfungen" [creations], the noun which the
possessive adjective modified but not the noun to which it referred; and "die Mächte" [the powers], which occurred earlier in the sentence and to which Europe was being compared. Other students left the item blank or interpreted "unserer" [our] in a more literal sense as referring to "Americans," the group to which they specifically belonged. These subjects erroneously included themselves as part of the author's intended audience. This last interpretation, in particular, illustrates how readers' beliefs not just about the passage but also about themselves as readers can influence their interpretation of anaphoric references.

The readers' task was certainly complicated by the fact that locating the appropriate coreferent called for inferential reasoning and required that the subjects understand in a global sense what the passage was about. Although "Europäer" [Europeans] was nowhere directly stated in the text, the discourse topic was Europe and the author was himself a European. He, therefore, used "unserer" [our] to refer collectively to himself and other "Europeans." This usage demanded that the readers recognize that the target audience included only other Europeans and that they reach beyond the text itself to identify this implied antecedent. The majority of subjects were able to do this successfully. Those who could not either restricted their search for a coreferent to concepts specifically introduced in the text or looked beyond the text for a coreferent but incorrectly included themselves in the antecedent group.

The second plural possessive adjective investigated in the coreferent selection task was "ihren" [their], used to refer to "die Angelsachsen" [the Anglo-Saxons]. The percentage of incorrect responses for this anaphor are presented as Item 20 in Figure 11. This anaphoric expression proved to be rather difficult
for readers to interpret accurately, as 44.444% of them missed the item. This percentage closely matches those for the plural personal pronoun “sie” [they], also used to refer to “die Angelsachsen” [the Anglo-Saxons]. Apparently, this entire pronoun cluster caused subjects a lot of interpretation problems. Those students who misconstrued the anaphor had an average baseline German language ability score of 29.583, indicating that as a group they had below average German language skills.

Among the students who missed the item, the most popular interpretations by far were that “ihren” [their] referred to: “Lenin, Hitler, and Stalin;” “Hitler and Stalin;” “Hitler;” “Deutschland” [Germany]; or “Hitlers Krieg und Niederlage” [Hitler's war and defeat]. In contrast to the majority of other incorrect responses made by readers, these selections were obviously not made on the basis of proximity to the anaphor. The possessive adjective was, in fact, introduced immediately after the antecedent to which it referred. The concepts “Hitler,” “Stalin,” “Lenin,” and “Deutschland” [Germany], however, were mentioned one to three lines before the anaphor appeared and were essentially backgrounded by the author. Nonetheless, they were chosen by several subjects as an appropriate coreferent. Evidently, these low ability subjects relied on their general knowledge of the candidate antecedents in the referential situation to help them determine which one was the more likely coreferent. In other words, subjects' background or general knowledge of World War II may have predisposed them to interpret “ihren” [their] in the phrase “ihren Teil mit beisteuerten” [their part contributed (to the war)], as referring specifically to “Hitler,” “Stalin,” or “Germany,” because this is what is typically emphasized in American histories of World War II. In this instance, students were perhaps too
eager to reach beyond the text and to introduce their own beliefs or interpretations into the coreferent selection process. As a result, they were unable to recognize the author's intended meaning.
CHAPTER V
DISCUSSION, SUMMARY, IMPLICATIONS, RECOMMENDATIONS, AND LIMITATIONS

Overview of the Study

This study investigated the processing and comprehension of anaphoric expressions in German text by readers at various levels of baseline German ability. In order to carry out this investigation, three specific purposes for the study were established. The first purpose was to examine the effect of anaphora on the general cognitive processing behaviors and reading comprehension of readers of German at various levels of baseline German language ability. This purpose was addressed in Phase I of the study. The second purpose was to assess how readers process specific anaphoric references within German texts. Phase II of the study addressed this purpose. The third purpose was to determine if readers of German were able to identify correctly the coreferents of various anaphoric expressions in a German text and if this ability was related to their overall comprehension of the text and their baseline German language ability. This purpose was addressed in Phase III of the study. In addition, this study was designed to add to the relatively small amount of both second language research and eye-tracking research that has investigated learners’ abilities to process and understand anaphoric references.
Phase I of this study examined the effect of anaphora on readers' general processing and comprehension of German texts. Analysis of Covariance (ANCOVA) was used for statistical analysis. There were three independent variables investigated in Phase I of the study. The first independent variable, baseline German language ability, was treated as a continuous, quantitative variable and was included as the covariate in the ANCOVA tests. Type of referencing device, the second independent variable, was treated as a fixed, categorical variable with two levels: (a) authentic, representing an authentic text with anaphoric expressions; and (b) experimental, representing the same text except that the anaphoric expressions were replaced with their antecedent nouns and noun phrases. A third independent variable, text, was treated as a random, categorical variable with two levels: expository text and literary text.

There were four quantitative, dependent variables measured in Phase I of this study: fixation frequency, average fixation duration, total fixation duration, and reading comprehension. Readers' cognitive processing of the texts was assessed by the first three dependent variables, which are indices of eye-movement behavior. Reading comprehension, the fourth variable, was measured via written recall protocols. Separate 2 X 2 (Text X Type of Referencing Device) mixed ANCOVA tests were run for each of these four dependent variables.

In Phase II of this study, readers' processing of specific anaphoric references in German texts was assessed. Readers' fixation behaviors in blocks of text containing either anaphors or repeated nouns or noun phrases were analyzed using Analysis of Covariance (ANCOVA). The covariate, baseline German language ability, was treated as a continuous, quantitative
variable. Type of referencing device, was treated as a fixed, categorical, independent variable with two levels: (a) authentic and (b) experimental.

Two quantitative, dependent variables were investigated in this phase of the study: (a) average fixation duration in blocks and (b) percentage of total fixation time in blocks. These variables, as indices of eye-movement behavior, provided information about readers' cognitive processing of particular anaphoric expressions in texts. Both variables were measured in specific blocks of a matrix mapped onto both versions of each text. Only those blocks containing anaphoric expressions in the authentic text version were compared with the corresponding blocks in the experimental version containing repeated nouns or noun phrases to test the effect of the referencing device. Separate one-way ANCOVA tests were run for each block of interest in each text.

In Phase III of this study, readers' comprehension of a longer German text was measured in addition to their ability to identify correctly the intended coreferents of specific anaphoric expressions within the text. The relationships between coreferential tie comprehension, overall text comprehension, and baseline German language ability were examined using a Pearson product-moment correlation procedure. In addition, a qualitative analysis of subjects' coreferential tie errors was performed to determine possible sources of influence on the resolution process.

Discussion

The following discussion of the results follows the same order as the presentation of the results reported in the preceding chapter. The findings for each phase of the study are discussed in separate sections and within each
section, the results of each dependent variable are presented individually.

Phase I of the Study

Fixation Frequency

✓ It was fully expected that the presence of anaphora in texts would increase the subjects’ fixation frequencies, because they would require additional fixations to interpret the coreferents of the anaphoric expressions in the texts.
✓ An increase in fixation frequencies was also expected to be the most apparent among subjects with low baseline German language abilities, because increased processing demands were expected to interact with deficient linguistic skills. ✓ The ANCOVA for fixation frequency, however, indicated no significant main effect due to the type of referencing device used in the text and no significant interaction between type of referencing device and baseline German language ability. In other words, the presence of anaphoric expressions in the texts did not significantly affect the fixation frequencies of readers at any level of baseline German language ability.
✓ Despite this nonsignificant finding, the means predicted by the ANCOVA model for fixation frequency lend support to the expectations stated above.
✓ These predicted means show that on the average subjects required slightly more fixations to process the authentic text versions than the experimental text versions. ✓ This result, although not statistically significant, suggests that the presence of anaphora in German texts does increase the cognitive processing demands placed on readers. Presumably, these additional fixations are needed to process the anaphoric references in the texts.
The ANCOVA for fixation frequency did result in a significant main effect for baseline German language ability. This finding shows that as baseline German language ability scores increased, readers' fixation frequencies decreased dramatically, irrespective of which text version they read. This result suggests that as readers of German become more proficient in their overall ability to use the language, the frequency with which they fixate texts decreases regardless of whether anaphoric references are present in the texts or not. In other words, as students progress in their knowledge of German, they become more efficient as readers and, therefore, need to fixate the text less frequently. This finding lends support to earlier research that indicates that as second language readers become more proficient in the second language, their reading behaviors tend to mirror those of highly skilled and native readers (Bernhardt, 1983; Everson, 1986; Oller & Tullius, 1972; Saito, 1989).

A significant main effect for text was also detected. This result indicates that subjects' fixation frequencies vary significantly from text to text. This variability is, however, not a function of the type of referencing device used in the text, as the Type of Referencing Device X Text interaction was nonsignificant. Apparently, among the population of texts, each text has unique discourse features that alone or in combination will cause fixation frequencies to vary.

**Average Fixation Duration**

It was anticipated that text versions containing anaphoric expression would result in longer average fixation durations, because subjects would require extra processing time to identify appropriate antecedents. It was also expected
that average fixation durations would be greater for students with lower baseline German language abilities, because they lack sufficient linguistic skills to process the anaphoric references as quickly as more advanced learners. These expectations, however, were not supported by the findings. The ANCOVA for average fixation duration failed to detect a significant main effect for type of referencing device or a significant interaction between type of referencing device and baseline German language ability. Apparently, the use of anaphoric expressions in the texts did not significantly affect the readers' average fixation durations, regardless of their level of baseline German language ability.

Contrary to the findings for fixation frequency, the means predicted by the ANCOVA model for average fixation duration reveal an unexpected trend. These predicted means demonstrate that subjects' average fixation durations were actually somewhat shorter for the text versions that contained anaphoric references than for text versions that contained repeated nouns and noun phrases. Even though this finding is not statistically significant, it implies that greater cognitive processing is not needed to handle anaphora in German texts. One possible reason for this finding is that nouns and noun phrases generally take more letters to express in writing than do anaphoric expressions and these additional letters may require extra processing time. Because the majority of subjects were processing words in a nonnative language, it is quite likely that they had not yet learned to recognize some of the words as entire units, but had to rely on letter by letter processing. First language research suggests that this kind of processing can significantly increase recognition times and alter fixation behaviors (Samuels & Kamil, 1984).
Although no main effect was found for type of referencing device, the ANCOVA for average fixation duration did indicate a significant main effect for baseline German language ability. This finding revealed that the average length of fixations decreased as readers became more proficient in their overall ability to use German, regardless of whether anaphoric references were present in the texts or not. It seems that as readers expand their knowledge of the language, they require fixations of shorter durations to process German texts.

The ANCOVA for average fixation duration also revealed a significant main effect for text. This result indicates that subjects' average fixation durations vary significantly from text to text. This variability is, however, not a function of the type of referencing device used in the text, as the Type of Referencing Device X Text interaction was nonsignificant.

**Total Fixation Duration**

It was expected that subjects would have longer total fixation durations for the authentic text versions than for the experimental text versions, because additional processing time would be needed to locate the intended coreferents. In particular, subjects with low baseline German language ability were expected to manifest greater total fixation durations as increased processing demands interacted with inadequate linguistic skills. The results of the ANCOVA for total fixation duration fully corroborate the findings for fixation frequency and average fixation duration. The ANCOVA for total fixation duration produced no significant main effect for type of referencing device and no significant interaction between type of referencing device and baseline German language ability. In other words, the presence of anaphoric expressions in the texts did
not significantly alter the total fixation durations of readers at any level of baseline German language ability.

The means predicted by the ANCOVA model for total fixation duration reveal a pattern of fixation behavior similar to that obtained for average fixation duration. The predicted means illustrate that on the average subjects’ total fixation durations were slightly shorter for the text versions that contained anaphoric references than for text versions that contained repeated nouns and noun phrases. Although this is a nonsignificant finding, it implies that fewer processing demands were placed on the reader when interacting with German texts that contain anaphoric expressions. As noted above, one reason for this finding is that nouns and noun phrases are typically longer linguistic units than anaphoric expressions. They may, therefore, require more time to process, especially if readers are relying on a letter by letter processing approach.

The ANCOVA for total fixation duration did reveal a significant main effect for baseline German language ability. This finding indicated that as readers of German became more proficient in their overall ability to use the language, the length of time spent fixing the texts decreased significantly, regardless of whether anaphoric references were present in the texts or not. In other words, as readers increase their knowledge of German, they become more efficient and effective readers and, therefore, need to allocate less total fixation time to the reading process.

A significant main effect for text was also detected. This result indicates that subjects’ total fixation durations vary significantly from text to text. This variability is, however, not a function of the type of referencing device used in the text, as the Type of Referencing Device X Text interaction was
nonsignificant.

Reading Comprehension

Previous research indicates that comprehension scores should increase in texts where anaphoric expressions have been replaced with more explicit nouns and noun phrases, because there is no opportunity to misinterpret the potentially ambiguous anaphors. According to first language research, this increase in comprehension scores is expected to be the most dramatic for subjects with low baseline language abilities (Richer, 1977). The results of the ANCOVA for reading comprehension, however, revealed no significant main effect for type of referencing device and no significant interaction between type of referencing device and baseline German language ability. Evidently, the presence or absence of anaphora in a text is not a significant indicator of how well students are able to recall that text.

Despite these nonsignificant findings, the predicted cell means produced by the ANCOVA model for reading comprehension indicate a trend in the expected direction. The statistical analysis revealed that subjects' predicted recall scores were actually higher for the experimental text versions than for the authentic text versions. Although this increase in comprehension scores is not statistically significant, it may be of practical importance. The inclusion of nouns and noun phrases in the place of anaphoric expressions seems to assist readers in getting more information from the text than they otherwise would. In contrast, the presence of anaphora in texts appears to place greater comprehension demands on readers and can cause them to recall less of the text.
A significant main effect was found for baseline German language ability. This finding revealed that readers’ comprehension of German texts improved as they became more proficient in their overall ability to use German, regardless of whether anaphoric references were present in the texts or not. It seems that as readers expand their knowledge of the German language, they are able to recall more information from the German texts that they read.

The ANCOVA for reading comprehension also revealed a significant main effect for text. This result indicates that subjects’ comprehension scores vary significantly from text to text. This variability is, however, not a function of the type of referencing device used in the text, as the Type of Referencing Device X Text interaction was nonsignificant.

**Phase II of the Study**

The findings for Phase II of the study corroborate many of the research results obtained in Phase I. With very few exceptions, no significant main effect was found for type of referencing device in the blocks of text investigated. Unlike Phase I, however, there was no significant main effect for baseline German language ability in the majority of blocks examined. In addition, only a few blocks revealed a significant interaction between type of referencing device and baseline German language ability. The following discussion focuses on those blocks of text in which a significant interaction effect or main effect was obtained.
Average Fixation Duration in Blocks

Expository Text: "Der Mechanisierte Hof"

The majority of ANCOVAs for average fixation duration in blocks of "Der mechanisierte Hof" revealed no significant interaction effects or main effects. The results, however, did indicate a significant interaction between the independent variables for Block 1 ("sie/die Maschinen" [they/the machines]) and for Block 2 ("wir/die Bauern" [we/the farmers]) of "Der mechanisierte Hof." Both blocks measured readers' ability to process plural pronoun references in German. In both instances, average fixation durations for the authentic text version, containing anaphoric references, increased slightly as subjects' baseline German language ability increased. In contrast, for the experimental text version, containing repeated nouns and noun phrases, average fixation durations decreased as baseline German language ability increased. The repeated noun phrases in the experimental text version were processed much more quickly by the the high ability students than by those with low abilities. Low ability students, however, were slightly faster at processing the anaphoric references, although the difference between the two groups was considerably smaller. Highly proficient readers had no trouble digesting the longer noun phrases, yet appeared to be slightly more attentive to the anaphoric references. It is likely that lower ability students were more disturbed by the greater length of the noun phrases and presumably had to allocate extra time to the processing of not only the noun but its accompanying article, which carried important syntactic information.

The fact that other plural pronoun references in the text ("wir" [we] and "uns" [us]) were not found to be significant for average fixation duration
suggests that German plural personal pronouns in general were not particularly problematic for the readers. The only significant interactions were found for the first two blocks of interest located at the beginning of the text. This finding implies that both high and low ability subjects became more adept at processing the anaphors and the repeated nouns and noun phrases as they progressed through the text, possibly due to increased familiarity and practice.

In addition to interaction effects, a significant main effect for type of referencing device was detected for Block 1 ("sie/die Maschinen" [they/the machines]) and for Block 5 ("sie/die Mechanisierung" [they/the mechanization]) of "Der mechanisierte Hof." The significant main effect for Block 1 must be interpreted very cautiously in light of the significant interaction found for this block. This main effect simply suggests that fixations of greater duration were needed to process the repeated noun phrase than to process the anaphoric reference. Average fixation durations for Block 1 of the text also varied as a function of the readers’ baseline German language ability. As a result, it is not possible to conclude that subjects’ fixation patterns for this block varied only on the basis of the type of referencing device used in the text version they read.

The significant main effect for type of referencing device for block 5 ("sie/die Mechanisierung" [it/the mechanization]) demonstrates that subjects’ average fixation durations were significantly longer for the experimental text version than for the authentic text version, regardless of their baseline German language ability. In other words, readers needed longer fixations to process the unambiguous, but lengthy repeated noun phrase, "die Mechanisierung" [the mechanization], than they did to process the less explicit, but shorter anaphor, "sie" [it]. Because of its great length, it is possible that readers required fixations
of longer duration just to process the additional letters in the repeated noun phrase. As the longest noun phrase in the experimental text version, this explanation seems likely.

A significant main effect for baseline German language ability was also found for two blocks of text: Block 1 ("sie/die Maschinen" [they/the machines]) and Block 2 ("wir/die Bauern" [we/the farmers]) of "Der mechanisierte Hof." In both cases, the presence of a significant interaction precludes any conclusive judgments about the effect due to baseline German language ability alone. In general, it appears that for both blocks of interest, subjects' average fixation durations changed dramatically as their language proficiency increased. The extent and direction of this change, however, was dependent on the text version they read.

**Literary Text: "Im Volksgarten"**

There were no significant interaction effects for average fixation duration in any of the blocks of interest for the text "Im Volksgarten." The majority of ANCOVAs for average fixation duration in blocks of "Im Volksgarten" also revealed no significant main effects for the two independent variables. The findings, however, did indicate a significant main effect for type of referencing device for Block 2 ("sie/Anna" [she/Anna]) and Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." In both instances, subjects required longer fixations to process the anaphoric reference than to process the more explicit repeated noun phrase. In the case of Block 2, it seems that the proper name "Anna" was processed more quickly because it was a more informative designation of the concept than was the pronoun "sie" [she], according to the
Lakoff-Cole scale (Cole, 1974). Presumably, the reader had to work less hard to access the author’s intended meaning when the more direct concept designator “Anna” was used.

For Block 9 of the text, the designation of the concept via a descriptive noun phrase, “dem Mädchen” [the girl], also required shorter fixations to process than the use of the pronoun form “ihr” [her]. Definite descriptions are listed as more informative than pronouns on the Lakoff-Cole scale and, thus, more apparent to the reader (Cole, 1974). Obviously, the repeated noun phrase proved to be more accessible to readers than the less explicit anaphoric reference. The gender disagreement between the anaphor and its coreferent in Block 9 may have also proved problematic for students. In this case, the syntactically neuter noun “dem Mädchen” [the girl] was referred to with an anaphor that reflects the natural gender of the noun, i.e. with a feminine anaphor, “ihr” [her]. The subjects, therefore, probably needed longer fixations to identify the neuter noun “dem Mädchen” [the girl] as the antecedent of the feminine pronoun “ihr” [her].

The results of the ANCOVAs also detected a significant main effect for baseline German language ability for the same two blocks, Block 2 (“sie/Anna” [she/Anna]) and Block 9 (“ihr/dem Mädchen” [her/the girl]) of “Im Volksgarten.” In both cases, subjects with high baseline German language ability required fixations of shorter duration during processing than subjects with low baseline German language ability, irrespective of which text version they read. In other words, as subjects’ overall proficiency in German increased, their average fixation durations for these blocks of text decreased significantly, regardless of whether the blocks contained pronoun references or repeated nouns or noun
phrases.

Percentage of Total Fixation Time in Blocks

Expository Text: "Der Mechanisierte Hof"

Similar to the results for average fixation duration, the majority of ANCOVAs for percentage of total fixation time in blocks of "Der mechanisierte Hof" revealed no significant interaction effects or main effects. The findings, however, did indicate a significant interaction between the two independent variables for Block 1 of the text ("sie/die Maschinen" [they/the machines]). This result corresponds to the ANCOVA finding for average fixation duration in Block 1 of the text. For Block 1 of the authentic text version, containing the anaphoric reference, subjects' percentages of total fixation time increased dramatically as baseline German language ability increased. In contrast, percentages of total fixation time decreased dramatically as baseline German language ability increased for Block 1 of the experimental text version, containing the repeated noun phrase. The repeated noun phrase in the experimental text version was processed much more quickly by the the high ability subjects than by those with low abilities. Low ability students, however, were much faster at processing the anaphor than were students with high language abilities. Apparently, highly proficient readers had few problems processing the longer noun phrase, yet were more meticulous in their processing of the anaphoric reference. It is also likely that lower ability students were more disturbed by the greater length of the noun phrase "die Maschinen" [the machines] and presumably had to allocate extra time to the processing of not only the noun but its accompanying article.
In addition to an interaction effect, the ANCOVAs detected a significant main effect for baseline German language ability for Block 4 ("uns/die Bauern" [us/the farmers]) of "Der mechanisierte Hof." In this case, subjects with high baseline German language ability required a greater percentage of their total fixation time to process the information in this block than did subjects with low baseline German language ability, irrespective of which text version they read. In other words, as subjects' overall proficiency in German increased, their percentages of total fixation time for this block of interest also increased significantly. Significant main effects for baseline German language ability in other blocks of interest for both "Im Volksgarten" and "Der mechanisierte Hof" have typically revealed that low ability subjects required more time to process the blocks than did high ability subjects. It is difficult to determine what characteristics of this block made it more problematic for high ability students than for low ability students.

**Literary Text: “Im Volksgarten”**

Although no significant interactions were found for average fixation duration in blocks, the analysis for percentage of total fixation time in blocks of "Im Volksgarten" did reveal a significant interaction between the independent variables for Block 9 of the text ("ihr/dem Mädchen" [her/the girl]). The percentage of total fixation time needed to process the block in the experimental text version was consistently small for both high and low ability subjects. This finding suggests that students at all levels of language ability had relatively little difficulty processing the unambiguous noun phrase "dem Mädchen" [the girl]. In contrast, readers with low baseline German language ability allocated a much
greater percentage of their total fixation time to the processing of the gendered pronoun “ihr” [her] than did students with high language abilities. Apparently, the repeated noun phrase was equally accessible to both low and high ability readers, but the pronoun reference demanded more processing time from the low ability subjects. One reason for this disparity may have been that the pronoun “ihr” [her] was considerably less informative as a concept designator than “dem Mädchen” [the girl] and, thus, required more time to process, especially for the low ability students. The fact that the grammatically neuter noun “dem Mädchen” [the girl] was referred to by a feminine pronoun, that reflected only its natural gender, may have also been particularly troublesome for the low ability subjects. This trouble may have, consequently, forced them to spend more time in this block of the text as they attempted to identify the correct coreferent for the gendered pronoun.

The ANCOVAs for percentage of total fixation time in blocks also show a significant main effect for type of referencing device in Block 2 ("sie/Anna" [she/Anna]) and Block 9 ("ihr/dem Mädchen" [her/the girl]) of “Im Volksgarten.” These findings corroborate the results for average fixation duration in blocks. In both blocks, subjects devoted a greater percentage of their total fixation time to the processing of the anaphoric reference than to the processing of the more explicit repeated noun phrase. In the case of Block 2, it seems likely that “Anna” was processed more quickly because, as a proper name, it was a more informative designation of the concept than was the pronoun “sie” [she]. Obviously, less processing time was needed to identify the intended antecedent when a proper name was used to refer directly to that antecedent than when only a pronoun was used.
For Block 9 of the text, the significant main effect for type of referencing device must be interpreted very cautiously, because a significant interaction between the independent variables was also found for this block. The main effect simply suggests that a larger percentage of readers’ total fixation time was allotted to the processing of the anaphoric reference than to the processing of the repeated noun phrase. The percentage of total fixation time in Block 9 of the text also varied, however, according to the subjects’ baseline German language ability. As a result, it is not possible to conclude that subjects’ fixation times for this block are strictly a function of which text version they read.

A significant main effect for baseline German language ability was also detected for Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." The presence of a significant interaction in this block precludes any definitive claims about the effect due to baseline German language ability alone. In general, it appears that for Block 9 of the text, "Im Volksgarten," the percentage of total fixation time needed to process the block tended to decrease as subjects' baseline German language ability increased. The extent of that decrease, however, depended on which text version the students read.

**Phase III of the Study**

**Correlational Analysis**

The data analysis for Phase III of the study revealed significant relationships among all three of the variables investigated. A significant correlation coefficient of 0.6161 was obtained between coreferential tie comprehension and overall text comprehension for readers of German at various levels of baseline German language ability. This result indicated a
significant relationship between these two variables. This high positive correlation demonstrated that the ability to interpret anaphoric references accurately within the text accounted for approximately 40% of the variance in subjects' overall comprehension scores for that text. In general terms, this finding suggests that certain linguistic processing skills, such as the ability to resolve anaphoric expressions, do contribute significantly to readers' abilities to comprehend written discourse.

Additional data analysis produced a significant correlation coefficient of 0.7833 between baseline German language ability and coreferential tie comprehension. The high positive correlation between these two variables indicated that baseline German language ability accounted for roughly 60% of the variance in subjects' coreferential tie comprehension scores. In other words, the subjects' baseline German language ability scores were good indicators of their ability to comprehend coreferential ties within the text. This result is not surprising considering that the test used to measure subjects' baseline German language ability, the AATG National German Examination for High School Students, is primarily a discrete-point grammar and vocabulary test and the coreferential selection task is essentially a grammar-like exercise. The two activities tap similar linguistic skills and, thus, are significantly correlated to each other.

Baseline German language ability and overall text comprehension were also found to be significantly related to each other. A correlation coefficient of 0.4998 was obtained between these two variables. This positive correlation indicated that baseline German language ability accounted for approximately 25% of the variance in subjects' overall text comprehension scores. This
finding suggests that baseline language ability explains a portion, but only a portion, of readers’ comprehension abilities. Research by Bernhardt (1985) corroborates this result. She concluded that linguistic competence alone was an inadequate predictor of the ability to comprehend discourse. Moreover, Morgan and Green (1980) claim that linguistic competence actually plays a smaller role in discourse comprehension than researchers once believed. It appears that overall language skills are related to text comprehension and can explain a significant amount (25%) of the variance in comprehension scores. Other factors, however, such as background knowledge, reader interest, and affective variables may, in fact, prove to be better predictors of readers’ comprehension abilities.

**Qualitative Analysis**

A qualitative analysis of readers’ coreferential tie errors in Phase III of the study revealed how specific discourse features, selection strategies, and baseline German language ability impacted coreferential tie comprehension. There were 31 anaphoric references investigated in the coreferent selection task and every item was missed by some percentage of the students. The data analysis indicated that percentages of incorrect responses ranged from a low of only 5% to a high of 65%. Further data analysis revealed that for every item on the coreferent selection task, those subjects who misinterpreted the item had mean baseline German language ability scores that were below average for the sample of subjects. This finding supports the results of the correlational analysis reported above, which demonstrated that baseline German language ability and coreferential tie comprehension are highly correlated ($r = 0.7833$).
This lack of baseline German language skills made many subjects insensitive to important syntactic and semantic features. Several low ability subjects, for example, ignored capitalization, word form, and word function clues to select adjectives as coreferents, even though those adjectives could not possibly serve as antecedents given the context. This type of error was most prevalent when the intended coreferent was preceded by a long adjective phrase and indicated that low ability students, in particular, have difficulty identifying important elements in longer, more complex discourse.

Other low ability subjects disregarded semantic constraints to interpret "sie" [they], which referred to "die Mächte" [the powers] as a reference for "die Europäer" [the Europeans]. In this case, the author used "uns" early in the sentence to imply the sense of a we group (Europeans) and a they group (non-Europeans). The use of the pronoun "sie" [they] later in the sentence semantically precluded the pronoun from referring to the we group, i.e., "die Europäer" [the Europeans]. These kind of sophisticated semantic relationships within the sentence, however, were not apparent to low ability students and caused them to rely on more basic antecedent selection strategies, such as proximity to the anaphor. Overall, these findings suggest that the less proficient readers are in their overall knowledge of the language, the more inept they are at interpreting anaphoric relationships within texts.

Among those subjects who misinterpreted items, two general categories of anaphoric references proved the most problematic for them: personal pronouns and relative pronouns. Personal pronouns often proved difficult for the readers to interpret correctly, because they were gendered forms that were often located far from their antecedents. One of the most interesting findings was how
subjects with extremely low baseline German language abilities chose to handle gendered personal pronouns. These subjects often selected a noun with the same natural gender as the pronoun to serve as the antecedent. They favored this strategy even when the chosen coreferent made little sense semantically or had already been foregrounded in the discourse. The singular, masculine pronouns “ihn” and “ihm,” for example, were both misinterpreted as referring to “Hitler” or “Stalin,” even though subjects had to disregard important semantic constraints and appropriate foregrounded nouns in order to choose these nouns as antecedents.

Relative pronouns were often problematic for students, because they failed to recognize that relative clauses closely follow the person or thing to which they refer in order to reduce ambiguity. Instead, these subjects used what they thought was the gender and number of the anaphor to help them locate appropriate antecedents, or they searched for thematically prominent concepts to serve as coreferents. These approaches generally forced the readers to look too far away from the intended antecedent in their search for a coreferent and, therefore, caused them to misinterpret the relative pronoun.

The fact that both personal and relative pronouns proved to be among the most difficult items for subjects to interpret correctly was not unexpected given that pronouns are among the least informative concept designators (Cole, 1974). As such, pronoun antecedents are less obvious and place a greater processing burden on the reader. In contrast, definite noun phrases are more informative concept designators and place fewer processing demands on readers. It was not surprising, therefore, that the definite noun phrase investigated in the coreferent selection task was misinterpreted by only 30% of
the subjects, whereas personal and relative pronoun references were missed by as many as 60% and 50% of the subjects respectively.

Those subjects who failed to interpret the items correctly used several different selection strategies in their attempt to restrict the set of candidate antecedents. One of the most popular strategies was to select an antecedent on the basis of its proximity to the anaphor. In other words, subjects favored the most recently mentioned nouns or noun phrases as coreferents. This type of strategy accounted for such errors as misinterpreting “es” [it], a reference to “Amerika” [America], as referring to “Rußland” [Russia] or “das Gleichgewicht” [the balance]. More than 60% of the subjects misunderstood this anaphor, making it the most problematic item in the coreferent selection task. Another difficult item was the pronoun “sie” [they], which referred to “die Mächte” [the powers], but which subjects misidentified as “die Europäer” [the Europeans] in spite of serious semantic constraints. In both cases, the high error rate was attributed to the fact that the nouns most often chosen as antecedents were just mentioned in the preceding clause whereas the intended antecedent was presented much earlier in the discourse.

Choosing antecedents on the basis of proximity was often reinforced by what appeared to be a foregrounding strategy. In other words, readers selected certain coreferents because they thought those coreferents represented foregrounded information or the discourse topic. This strategy, for example, led students to interpret the singular, feminine pronoun “sie” [it], which referred to “die neue Machtstellung Rußlands” [the new political power position of Russia], as simply “Rußlands” [(of) Russia]. In this instance, the genitive noun phrase “Rußlands” [(of) Russia] followed the intended antecedent and, therefore, may
have been foregrounded for some students. This noun phrase was also located closer to the anaphor, so it may have encouraged subjects to use proximity as a strategy. In so doing, however, students failed to recognize the syntactic form and function of this noun phrase, which indicated that “Rußlands” [(of) Russia] was simply a genitive noun phrase used to modify the intended antecedent “die neue Machtstellung” [the new political power position].

On several occasions, a foregrounding strategy would have actually facilitated readers’ coreferential tie comprehension, but they chose to ignore thematically prominent information. In these instances, students preferred to rely on nontextual information, such as background knowledge or personal beliefs and perceptions when selecting an antecedent. Misinterpreting “ihren” [their] as referring to “Hitler and Stalin” or “Lenin, Hitler, and Stalin” in the phrase “ihren Teil mit beisteuerten” [their part contributed (to the war)] illustrates how students’ background knowledge or perceptions of World War II may have impacted their coreferent selections. In such cases, readers were perhaps too eager to reach beyond the text and to introduce their own beliefs into the resolution process.

Another common response was simply to leave the item blank, i.e. to skip the item and move on to another one. This type of response, for example, accounted for the majority of errors made on the frequently missed pronoun “sie,” used to refer to “die Anglosachsen” [the Anglosaxons]. This strategy was often accompanied by metacognitive comments such as putting a question above the item or in the margin. These comments indicated that students were themselves aware of their inability to answer some items correctly, perhaps because they were too frustrated or confused by those items.
Another noteworthy finding was that the percentage of incorrect responses rose dramatically in the second half of the coreferent selection task. For the first half of the task, i.e., the first 15 items, the average percentage of incorrect responses was only 16.914%. For the second half of the task, i.e., the last 16 items, the average percentage of incorrect responses rose dramatically to 44.444%. One reason for this large increase may have been that subjects' fatigue, frustration, confusion, and carelessness intensified as the task wore on. In other words, they were less willing and able to perform the task as they proceeded through it.

A more likely explanation, however, is that over half of the personal pronouns and relative pronouns occurred in the second half of the coreferent selection task. These two types of anaphoric expressions were the most problematic items for students. In contrast, the first half of the coreferent selection task was dominated by possessive adjectives, which were the easiest items for the subjects to identify correctly. It is likely, therefore, that the overall composition of the coreferent selection task caused the second half of it to be more difficult than the first half, and this difficulty translated into a substantially higher error rate for the second half of the task.

Summary of the Major Findings

The major findings of the present study are summarized below according to the research questions they attempted to answer.
Phase I of the Study

Question 1
What effect does the inclusion of anaphoric expressions in German text have on the general cognitive processing and comprehension of that text by readers of German at various levels of baseline German language ability as measured by the indices of fixation frequency, fixation duration, and written recall?

It was expected that the presence of anaphora in German texts would increase the subjects' fixation frequencies, their average fixation durations, and their total fixation durations, because they would require additional fixations and processing time to identify the appropriate antecedents for the anaphoric expressions in the texts. This increase was also expected to be the greatest for students with low baseline German language abilities, because they presumably lack sufficient linguistic skills to process anaphoric references as quickly as more advanced learners. The ANCOVAs for fixation frequency, average fixation duration, and total fixation duration, however, indicated no significant main effect due to the type of referencing device used in the text. In addition, no significant interaction between type of referencing device and baseline German language ability was detected for these three dependent variables. In other words, the inclusion of anaphoric expressions in texts does not appear to significantly affect the general cognitive processing of readers at any level of baseline German language ability as measured by fixation frequency, average fixation duration, and total fixation duration.
It was also anticipated that readers' comprehension scores would increase for texts in which anaphoric expressions had been replaced with more explicit nouns and noun phrases, because there would be no opportunity for the readers to misinterpret potentially ambiguous anaphors. This increase in recall scores was expected to be the most dramatic for subjects with low baseline language abilities, according to first language research findings (Richel, 1977). The results of the ANCOVA for reading comprehension, however, revealed no significant main effect for type of referencing device and no significant interaction between type of referencing device and baseline German language ability. Evidently, the presence of anaphora in a text does not significantly affect how well students are able to recall the text.

**Question 2**

How does the baseline German language ability of readers affect their general cognitive processing and comprehension of German text containing anaphoric expressions as measured by the indices of fixation frequency, fixation duration, and written recall?

The ANCOVAs for fixation frequency, average fixation duration, and total fixation duration all revealed a significant main effect for baseline German language ability. These results indicated that as baseline German language ability scores increased, readers' fixation frequencies, their average fixation durations, and their total fixation durations decreased significantly, regardless of whether anaphoric references were present in the texts or not. As students progress in their knowledge of German, it seems that they become more
efficient and effective at processing German texts and, therefore, need to allocate fewer fixations and less processing time to reading tasks.

The ANCOVA for reading comprehension also indicated a significant main effect for baseline German language ability. This finding revealed that readers' comprehension of German texts improved as they became more proficient in their overall ability to use German, regardless of whether anaphoric references were present in the texts or not. Apparently, as readers' knowledge of German increases, they are able to recall more information from the German texts that they read.

**Phase II of the Study**

**Question 3**

What effect does the inclusion of anaphoric expressions in German text have on the specific cognitive processing of those anaphoric references by readers of German at various levels of baseline German language ability as measured by the indices of fixation duration and percentage of total fixation time in blocks?

The findings for Phase II of the study corroborate many of the research results obtained in Phase I. With very few exceptions, no significant main effect was found for type of referencing device for the dependent variables of average fixation duration in blocks and percentage of total fixation time in blocks. In addition, only a few blocks revealed a significant interaction between type of referencing device and baseline German language ability. The following section summarizes the findings for those few blocks of text in which a
significant interaction effect or main effect for type of referencing device was obtained according to the dependent variables of average fixation duration and percentage of total fixation time.

The ANCOVA results for average fixation duration in blocks of “Der mechanisierte Hof” revealed a significant interaction between the independent variables for the plural pronouns in Block 1 ("sie/die Maschinen" [they/the machines]) and in Block 2 ("wir/die Bauern" [we/the farmers]). In both instances, average fixation durations for the authentic text version, containing anaphoric references, increased slightly as subjects’ baseline German language ability increased. In contrast, average fixation durations decreased as baseline German language ability increased for the experimental text version, containing repeated noun phrases.

The ANCOVAs for average fixation duration in blocks of “Der mechanisierte Hof” detected a significant main effect for type of referencing device for Block 1. It is not possible, however, to conclude that subjects’ fixation durations for this block are strictly a function of which text version they read, because a significant interaction between type of referencing device and baseline German language ability was also found for this block. In addition, a significant main effect for type of referencing device was found for block 5 ("sie/die Mechanisierung" [it/the mechanization]) of “Der mechanisierte Hof.” This finding demonstrates that subjects’ average fixation durations were significantly longer for the experimental text version than for the authentic text version, regardless of their baseline German language ability.

The ANCOVA results for average fixation duration in blocks of “Im Volksgarten” revealed no significant interaction effects between the two
independent variables. The findings, however, did indicate a significant main effect for type of referencing device for Block 2 ("sie/Anna" [she/Anna]) and for Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." In both instances, subjects required longer fixations to process the anaphoric reference than to process the more explicit repeated noun or noun phrase.

The ANCOVAs for percentage of total fixation time in blocks of "Der mechanisierte Hof" detected a significant interaction between the two independent variables in Block 1 of the text ("sie/die Maschinen" [they/the machines]). This result generally corroborates the ANCOVA for average fixation duration in Block 1 of the text. For Block 1 of the authentic text version, containing the anaphoric expression, subjects' percentages of total fixation time increased dramatically as baseline German language ability increased. In contrast, percentages of total fixation time decreased dramatically as baseline German language ability increased for Block 1 of the experimental text version, containing the repeated noun phrase. The ANCOVA results for percentage of total fixation time in blocks of "Der mechanisierte Hof" detected no significant main effects for type of referencing device.

The ANCOVA results for percentage of total fixation time in blocks of "Im Volksgarten" revealed a significant interaction for Block 9 ("ihr/dem Mädchen" [her/the girl]). The percentage of total fixation time needed to process the block in the experimental text version was consistently small for both high and low ability subjects. In contrast, readers with low baseline German language ability allocated a significantly greater percentage of their total fixation time to the processing of the gendered pronoun "ihr" [her] in the authentic text version than did students with high language ability.
The ANCOVAs for percentage of total fixation time in blocks detected a significant main effect for type of referencing device in Block 2 ("sie/Anna" [she/Anna]) and Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." These findings corroborate the results for average fixation duration in blocks. In both blocks, subjects devoted a greater percentage of their total fixation time to the processing of the anaphoric reference than to the processing of the more explicit repeated noun or noun phrase. For Block 9, however, it is not possible to conclude that subjects’ fixation times for this block are strictly a function of which text version they read, because a significant interaction between the two independent variables was also found.

**Question 4**

How does the baseline German language ability of readers affect their specific cognitive processing of anaphoric references in German text as measured by the indices of fixation duration and percentage of total fixation time in blocks?

Unlike the results for Phase I of the study, the majority of blocks revealed no significant main effect for baseline German language ability as measured by the dependent variables of average fixation duration in blocks and percentage of total fixation time in blocks. The following section summarizes the findings for those few blocks of text in which a significant main effect for baseline German language ability was obtained according to the dependent variables of average fixation duration and percentage of total fixation time.
The ANCOVA results for average fixation duration in blocks revealed a significant main effect for baseline German language ability in Block 2 ("wir/die Bauern" [we/the farmers]) of "Der mechanisierte Hof." The presence of a significant interaction in this block, however, precludes any definitive claims about the effect due to baseline German language ability alone.

The ANCOVAs for average fixation duration in blocks detected a significant main effect for baseline German language ability in Block 2 ("sie/Anna" [she/Anna]) and in Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." In both cases, subjects with high baseline German language ability required fixations of shorter duration during processing than subjects with low baseline German language ability, irrespective of which text version they read.

The ANCOVA results for percentage of total fixation time in blocks revealed a significant main effect for baseline German language ability in Block 4 ("uns/die Bauern" [us/the farmers]) of "Der mechanisierte Hof." The findings indicated that subjects with high baseline German language ability required a greater percentage of total fixation time to process the information in this block than did subjects with low baseline German language ability, regardless of which text version they read.

The ANCOVAs for percentage of total fixation time in blocks detected a significant main effect for baseline German language ability in Block 9 ("ihr/dem Mädchen" [her/the girl]) of "Im Volksgarten." The presence of a significant interaction in this block, however, precludes any definitive claims about the effect due to baseline German language ability alone.
Phase III of the Study

Question 5

Are readers of German able to identify correctly the coreferents of various anaphoric expressions in a German text and is this ability related to their overall comprehension of the text and to their baseline German language ability?

The data analysis for Phase III of the study revealed significant relationships between all three of the variables investigated. A significant correlation coefficient of 0.6161 was found between coreferential tie comprehension and overall text comprehension for readers of German at various levels of baseline German language ability. This high positive correlation demonstrated that the ability to interpret anaphoric references accurately within the text accounted for approximately 40% of the variance in subjects’ overall comprehension scores for that text. In general terms, this finding suggests that certain linguistic processing skills, such as the ability to resolve anaphoric expressions, do contribute significantly to readers’ abilities to comprehend written discourse.

A significant correlation coefficient of 0.7833 between baseline German language ability and coreferential tie comprehension was detected. The high positive correlation between these two variables indicated that baseline German language ability accounted for roughly 60% of the variance in subjects’ coreferential tie comprehension scores. In other words, baseline language ability scores are good indicators of readers’ abilities to comprehend coreferential ties within texts.
Baseline German language ability and overall text comprehension were also found to be significantly related to each other. A significant correlation coefficient of 0.4998 was obtained between these two variables. This positive correlation indicated that baseline German language ability accounted for approximately 25% of the variance in subjects' overall text comprehension scores. This finding reveals that baseline language ability is related to discourse comprehension but can explain only a portion of readers' overall comprehension of written texts.

Implications for Pedagogy

The fact that readers processed and comprehended text versions containing anaphoric expressions as readily as text versions containing more explicit nouns and noun phrases provides strong support for the use of authentic and unaltered texts in second language instruction. The results of this study expressly indicate that neither processing nor comprehension was significantly facilitated when texts were manipulated through the removal of anaphoric expressions and the inclusion of repeated nouns and noun phrases. With very few exceptions, readers at all levels of baseline German language ability were capable of handling a variety of anaphoric references in a discourse context. Contrary to some first language research evidence, there is little indication from the results of this study that low ability students would benefit from simplified reading materials (Richer, 1977). In fact, there is some indication that in a small percentage of instances the simplification of anaphoric references actually impeded readers' comprehension and processing, perhaps by removing some of the authenticity from the text.
A second important implication is that reading comprehension scores cannot be predicted solely on the basis of grammar-based activities, such as the baseline German language ability test or the coreferential selection task. The findings from this study indicate that basic German language skills do account for a portion (roughly one quarter) of readers’ discourse comprehension, but that there is also a lot of variance in readers’ comprehension scores that cannot be explained by subjects’ baseline language abilities. Likewise, the results reveal that the comprehension of a specific language feature, such as anaphora, accounts for only a portion (roughly one third) of readers’ discourse comprehension. Such findings imply that teachers cannot determine students’ reading comprehension abilities on the basis of strictly grammar-based activities. The ability to perform linguistic activities is a significant component of the comprehension process, but is not equivalent to that process. Reading comprehension measures, therefore, must involve more than merely an assessment of learners’ ability to manipulate grammatical features. Such measures must, instead, allow instructors to form a more complete picture of how readers process and comprehend written information to form their own interpretation of the discourse.

These findings do not imply that grammar instruction, such as exercises in the formulation and use of pronouns, is unnecessary or undesirable. On the contrary, this study has illustrated that linguistic competence can facilitate discourse comprehension. Grammar instruction, therefore, can be a useful aid for the processing and comprehension of texts. The results suggest, however, that such instruction will be more effective if taught within a discourse context. Teachers may spend much class time telling students that the masculine
pronoun “ihn” [it/he] can be used to refer to the inanimate, masculine noun “der Krieg” [the war]. It is, nevertheless, very unrealistic on the basis of such instruction to then expect students to interpret “ihn” [it/he] correctly in a discourse context in which animate, masculine nouns like “Hitler” compete with the intended coreferent during processing. Contextualizing grammar instruction may help readers to use their linguistic knowledge more efficiently and effectively during the comprehension process.

Recommendations for Further Research

This study is the first of its kind to investigate the effect of anaphora on the processing and comprehension of readers of German at various levels of baseline German language ability. Replications and expansions of the present study, therefore, are clearly needed. This study has focused primarily on how nonnative (American) readers of German process and understand anaphoric expressions in texts. Studies involving second language learners of German from other first language groups are also necessary. In particular, these investigations should include learners with first languages whose anaphoric systems are similar, as well as dissimilar to German. In addition, research should explore how anaphora is processed and understood in languages that have gendered anaphors similar to German, such as French and Spanish. These studies should investigate the processing and comprehension of both first and second language learners of the language.

There is also a need for more research using native German readers. Existing studies involving native readers of German have investigated primarily children's acquisition and comprehension of anaphoric forms. In addition, this
research has examined only a small number of anaphoric structures, most notably relative pronouns. Future studies, therefore, should examine how native readers of German at different stages of language development process and comprehend a variety of anaphoric forms. Further research in this area will reveal how linguistic sensitivity to anaphora develops among first language readers of German. This knowledge will provide a framework for comparing and contrasting how such linguistic sensitivity develops in second language readers.

This study attempted to investigate several different types of anaphoric structures. Particular emphasis was placed, however, on gendered pronoun references. Additional investigations need to broaden the range of anaphoric categories to include such anaphoric forms as epithets and ellipses. Moreover, Demel (1987) found that some uses of noun phrases in an English text were culturally based and, therefore, difficult for ESL readers to interpret correctly. Research is needed to determine if certain anaphoric references in German carry culturally based information that may be inaccessible to second language readers.

Future research directions should also encompass investigations of the effect of cataphoric references on the processing and comprehension of second language readers. Because cataphoric forms depend on succeeding discourse for their interpretation, they may be even more problematic for second language readers to interpret correctly.

The focus of this study was to determine how the presence of anaphora affected subjects’ comprehension of written texts. There is also a need, however, to investigate how anaphoric references in spoken language impact
learners' comprehension abilities. Interpreting anaphors in spoken discourse may prove particularly challenging for students because they must not only identify but also remember candidate antecedents presented in preceding discourse. These activities may require specialized and highly sophisticated processing skills, the precise nature of which only further research can clarify.

This study only briefly addressed the impact of grammatical case on the processing and comprehension of anaphoric expressions. Because German has distinct markers for each of its four cases, anaphor-antecedent relationships can often be obscured in discourse and especially difficult for second language readers to interpret. Studies, therefore, need to investigate how grammatical case may affect readers' interpretation of both noun phrases and anaphoric forms within texts. This research may then indicate under what syntactic conditions the processing and comprehension of anaphora is either facilitated or impeded.

A major finding of this study was that anaphoric references in German text had no significant effect on readers' processing or comprehension. Readers' collective baseline German language ability, however, was a significant factor affecting their processing and understanding of text. This finding suggests that anaphora in combination with other discourse features may have a greater impact on readers' processing and comprehension of German text. Future studies need to explore some of these possible combinations.

Limitations of the Study

1. It was not possible to select a random sample of subjects for this study.

   The subjects who participated in this study were all paid volunteers.
The results, therefore, can not be generalized to other populations.

2. Another serious limitation of the study was the small sample size. Even with a monetary incentive, it was difficult to obtain a large number of students who read German and were willing to participate in the study. If sample sizes had been larger, the statistical power of the analyses would have been greater. With greater statistical power, it is possible that significant differences between types of referencing device and baseline German language ability scores could have been detected for some of the dependent variables.

3. The location of anaphoric references or repeated noun phrases within the text may have influenced readers' eye movements. Research has indicated that short words, such as pronouns or articles, that occur at either the beginning or the end of a line are more likely to be skipped during reading. If these words happened to be anaphoric expressions, short repeated nouns, or articles that were part of repeated noun phrases, the findings of the study may have been affected.

4. In order to collect eye-tracking data, subjects are required to sit in an unnatural position. The head must be stabilized by a metal head bar and chin rest designed to minimize any head movement that could invalidate the eye-tracking data. In addition, subjects had to read the text from a computer screen located 21 inches away from their heads, while surrounded by video camera equipment and an infrared light source. The study is limited, therefore, by the extent to which reading under these necessary, yet artificial conditions resembles normal reading behavior.
5. The study is also limited by the fact that reading words displayed on a computer screen may be considerably different from reading words printed on paper. Images presented on a computer monitor may vary greatly in quality and are affected by such factors as resolution, color, hardware, and software design. A computer screen that afforded reasonably good resolution and color was used in this study, but even this screen may not have been comparable to the printed page.

6. The length of the passages to be displayed on the computer screen was limited by the size of the screen. The equipment used in this study allowed a maximum of 12 double-spaced lines, containing 80 characters each, to be displayed on the monitor at one time. This limitation prevented a longer German text from being used in Phases I and II of the study, in which eye-tracking data were collected.

7. In Phase III of the study, subjects were instructed to write the coreferents for specific anaphoric expressions above the anaphors. This task is an artificial, grammar-like activity and, therefore, may not provide a completely accurate assessment of the subjects' coreferential tie comprehension as it occurs during reading.

8. The subjects who participated in this study may not have had the same ideas about how to write recall protocols or may not have had previous experience in writing them. Johnston (1983) states that difficulties may arise if students misinterpret the level of detail that they should include in their recalls and the degree to which they should maintain the surface structure of the original passage in their recalls. Individual subjects may also have differed in the perspective from which they
chose to present their recalls. Subject differences and lack of experience may have affected the results of the study.
APPENDIX A

INFORMATION FLIER ABOUT THE STUDY
ATTENTION STUDENTS STUDYING GERMAN
EARN $10.00 QUICKLY AND EASILY!!!!

HOW: Participate in a research project in German.

WHO: All levels of German students are needed (Native speakers, high, intermediate, and low ability students).

WHEN: At your convenience. Call to schedule a time. Evenings and weekends available.

WHERE: Room 260, Arps Hall. (Upstairs in the Foreign Language Research Lab).

WHAT: The data collection session lasts approximately 2 hours. The session consists of a reading comprehension test reading and recalling 2 short German texts and a longer German text, and completing a short grammar-like exercise. Subjects will receive $10.00 upon completion of the session.

ALL PARTICIPATION IS VOLUNTARY AND DOES NOT AFFECT YOUR GERMAN COURSE GRADE IN ANY WAY.

INTERESTED STUDENTS SHOULD CONTACT:
VICKY BERKEMEYER, 771-5571

This project has been approved by Professor Vitt, Chairman of the German Department, and Dr. Corl, Coordinator of Undergraduate Instruction.
APPENDIX B

LETTER OF REQUEST TO DISTRIBUTE INFORMATION FLIERS
DEAR INSTRUCTOR,

Attached you will find a handout describing a research study being conducted in the Department of Educational Studies. Your cooperation in announcing this study to your classes and in passing out this information is urgently needed. It will be difficult to recruit the number of subjects needed for the study without your help.

Subjects at all levels of German study (native, high, average, and low ability students), who can read German text are needed as subjects. As noted in the handout all subjects will receive $10.00 for their participation.

Let me assure you that I have cleared this project with Professor Vitt, Chairman of the German Department; and Dr. Corl, Coordinator of Undergraduate Instruction, and have gotten their full approval to request your help and to distribute these handouts. Your cooperation in this matter is purely voluntary. You are in no way obligated to pass out this information.

I encourage you, however, to help another graduate student complete her dissertation work. I also urge you to participate in the study yourselves. Subjects who are highly proficient or native speakers of German are also greatly needed. Data collection sessions are very flexible and instructors will also receive $10.00 for their participation.

I appreciate your help and cooperation very much. Please contact me if you have any questions or concerns at 771-5571.

Thank you for your time and assistance.

Sincerely,

Vicky Berkemeyer

Vicky Berkemeyer
APPENDIX C

AUTHENTIC AND EXPERIMENTAL VERSIONS OF THE
EXPOSITORY TEXT: "DER MECHANISIERT HOF"
"Das Schwierige," sagt der Bauer, der sich jetzt auf das alte Sofa niedersetzt, "das Schwierige ist gar nicht der Preis der Maschinen, sondern die wenigen Tage, in denen sie arbeiten. In der Spitzenzeit der Ernte, da brauchen wir alles. Der Traktor geht schneller durch das Feld. Er hilft uns über das ganze Jahr hindurch. Die Mechanisierung verändert nicht nur das Dorf, sie verändert uns innerlich. Wir sind nicht die gleichen wie früher, das können Sie wohl glauben."

"Das Schwierige," sagt der Bauer, der sich jetzt auf das alte Sofa niedersetzt, "das Schwierige ist gar nicht der Preis der Maschinen, sondern die wenigen Tage, wenn die Maschinen arbeiten. In der Spitzenzeit der Ernte, da brauchen die Bauern alles. Der Traktor geht schneller durch das Feld. Der Traktor hilft den Bauern über das ganze Jahr hindurch. Die Mechanisierung verändert nicht nur das Dorf; die Mechanisierung verändert die Bauern innerlich. Die Bauern sind nicht die gleichen wie früher, das können alle wohl glauben."
APPENDIX D

AUTHENTIC AND EXPERIMENTAL VERSION OF THE LITERARY TEXT: "IM VOLKSGARTEN"
Im Volksgarten  
(Authentic Version)

"Ich möchte einen blauen Ballon haben!" sagte Anna.
"Da hast du einen blauen Ballon, Anna!" sagte die Mutter.
"Ich möchte ihm die Luft auslassen," sagte sie einfach.
"Willst du ihn nicht diesem armen Mädchen schenken, Anna?!
"Nein, ich will ihn fliegen lassen!" Sie lässt den Ballon aus, sieht ihm nach, bis er verschwindet in den blauen Himmel.
"Mutter, ich hätte ihn lieber dem armen Mädchen geschenkt!"
"Da hast du einen anderen blauen Ballon, Anna, schenke ihm diesen!" sagte die Mutter.

Im Volksgarten  
(Experimental Version)

"Ich möchte einen blauen Ballon haben!" sagte Anna.
"Da hast du einen blauen Ballon, Anna!" sagte die Mutter.
"Ich möchte dem Ballon die Luft auslassen," sagte Anna einfach.
"Willst du den Ballon nicht diesem armen Mädchen schenken, Anna?!
"Nein, ich will den Ballon fliegen lassen!" Anna lässt den Ballon aus, sieht dem Ballon nach, bis der Ballon verschwindet.
"Mutter, ich hätte den Ballon lieber dem armen Mädchen geschenkt!"
"Da hast du einen anderen blauen Ballon, Anna, schenke dem Mädchen diesen Ballon!" sagte die Mutter.
APPENDIX E
EXPOSITORY TEXT: "EUROPA HEUTE"
The following is an essay from a German reader. Please read the essay with the intent of understanding as much of it as you can. Read the essay as many times as you wish and take as much time as you need. After you have finished reading it, give it to the researcher. You will then be asked to write down everything you can remember about the essay.

EUROPA HEUTE

von

Golo Mann

1958


Europa ist nicht mehr das Zentrum der Erde; nicht wirtschaftlich, nicht machtpolitisch, nicht moralisch und geistig. Diese gesegnete Halbinsel aus Halbinseln, dieser wunderbar reiche, feingliederte Kontinent ist nur noch ein Überbleibsel dessen, was es war, nicht wissend, was mit seiner großen Vergangenheit anfangen, wie sich benennen, wie seine Zukunft gestalten. Das, was Europas Größe war und die Quelle seiner weltgestaltenden Kräfte, hat es schließlich um seine Erstgeburt betrogen: seine Geteiltheit in sich, der rasende Wettbewerb seiner Staaten und Nationen. Unser letzter melancholischer Triumph ist es, daß die Mächte, welche heute "Finis Europa" rufen, "Asien den Asiaten,"
APPENDIX F

COREFERENT SELECTION TASK FOR THE
EXPOSITORY TEXT: "EUROPA HEUTE"

Europa ist nicht mehr das Zentrum der Erde; nicht wirtschaftlich, nicht machtpolitisch, nicht moralisch und geistig. Diese gesegnete Halbinsel aus Halbinseln, dieser wunderbar reiche, feingegliederte Kontinent ist nur noch ein Überbleibsel dessen, was es war, nicht wissend, was mit seiner großen Vergangenheit anfangen, wie sich benennen, wie seine Zukunft gestalten. Das, was Europas Größe war und die Quelle seiner weltgestaltenden Kräfte, hat es schließlich um
"Afrika den Afrikanern," doch immer nur ein Echo unserer eigenen Schöpfungen geben; ohne uns Europäer wüßten sie nicht, was "Freiheit" ist und Demokratie und Nation und Sozialismus und Kommunismus, wüßten sie nicht einmal, daß es Asien und Afrika gibt. Noch die Ideen, mit denen man Europa zum Rückzug zwingt, sind europäischen Ursprungs. Rußland, zuerst ein barbarisches Land jenseits der Grenze, dann Jahrhunderte lang ein europäischer Staat unter anderen, ist zur Weltmacht geworden und hat gleichzeitig tief sich in Europa hineingefressen. Fragen wir nicht, wie das geschehen konnte; wie Deutschlands Zusammenspiel mit Lenin, 1918, diese Entwicklung begünstigte, und Hitlers Pakt mit Stalin, und wieder Hitlers Krieg und Niederlage; und wie auch die Angelsachsen ihren Teil mit beisteuerten durch die Art, in der sie den zweiten Weltkrieg führten und ihn verloren, indem sie ihn gewannen. Sagen wir nur, das ist geschehen und das läßt sich jetzt nicht mehr rückgängig machen. Welche
APPENDIX G

STUDENT INFORMATION FORM
SUBJECT NUMBER: ____________ Date: ____________

Name: ______________________, ______________________
   last name          first name

Age: ____________

Telephone Number: ________________

Address: ________________________________

City: ______________ State: __________ Zip: __________

Graduate Student [ ] Undergraduate Student [ ] Other [ ] __________

Major or Proposed Major: ________________________________

Interests: ______________________________________________

How long have you studied German? ________________

Are you a native speaker of German? ________________

What other languages have you studied and for how long? ________________
   (e.g. French, 2 years)
   ________________
   __________________
APPENDIX H
SCREEN DISPLAY OF A NUMBER PATTERN
APPENDIX I

PRACTICE TEXT
Practice Text

Der Rhein

APPENDIX J

RECALL INSTRUCTIONS AND RESPONSE FORM
SUBJECT NUMBER: ___________

RECALL 1

Please write down everything you can remember from what you read in your native language.
APPENDIX K

COREFERENT SELECTION TASK: EXPLANATION, EXAMPLES, AND PRACTICE SENTENCES
Pronouns and some noun phrases frequently refer to items, called "coreferents," found elsewhere in the text. In the sentence pair (a) below, for example, "sie" refers to "Karin." In the sentence pair (b), "ihr" also refers to "Karin."

(a) Karin ist Studentin. **Sie** wohnt in Heidelberg.

(b) Karin studiert Chemie. **Ihr** Bruder studiert Musik.

The following sentences contain more examples of pronouns or noun phrases and their coreferents. The coreferent has been written above the pronoun or noun phrase.

(c) Frau Kraft hat einen neuen Wagen. Hast du **ihn** gesehen?

(d) Es gibt keine Bücher, in **diesen** ich die Information finden kann.

(e) Was mir an dem Haus gefällt ist **seine** Farbe.

(f) Fritz hat das Spiel verloren. **Das** tut Monika leid.

(g) Das Essen im Restaurant ist gut, aber **es** ist teuer.

(h) Herr Krone sitzt im Restaurant. **Der alte Mann** beobachtet gerne die anderen Gäste.

Practice example:
Find the coreferent for the underlined pronoun in the sentence pair below. Write the pronoun's coreferent above the pronoun.

(i) Rita hat die Fenster geputzt. **Sie** sind aber noch schmutzig.
APPENDIX L

SCREEN DISPLAY OF THE BLOCK MATRIX FOR
THE EXPOSITORY TEXT: "DER MECHANISIERTE HOF"
AUTHENTIC AND EXPERIMENTAL VERSIONS
Der mechanisierte Hof

'Das Schwierige,' sagt der Bauer, der sich jetzt auf das alte Sofa niedersetzt, 'das Schwierige ist gar nicht der Preis der Maschinen, sondern die wenigen Tage, in denen sie arbeiten.

In der Spitzenzeit der Ernte, da brauchen wir alles. Der Traktor geht schneller durch das Feld. Er hilft uns über das ganze Jahr hindurch. Die Mechanisierung verändert nicht nur das Dorf, sie verändert uns innerlich. Wir sind nicht die gleichen wie früher, das können Sie wohl glauben.'
Der mechanisierte Hof

'Das Schwierige,' sagt der Bauer, der sich jetzt auf das alte Sofa niedersetzt, 'das Schwierige ist gar nicht der Preis der Maschinen, sondern die wenigen Tage, wenn die Maschinen arbeiten. In der Spitzenzeit der Ernte, da brauchen die Bauern alles.

Der Traktor geht schneller durch das Feld. Der Traktor hilft den Bauern über das ganze Jahr hindurch. Die Mechanisierung verändert nicht nur das Dorf; die Mechanisierung verändert die Bauern innerlich. Die Bauern sind nicht die gleichen wie früher, das können alle wohl glauben.'
APPENDIX M

SCREEN DISPLAY OF THE BLOCK MATRIX FOR
THE LITERARY TEXT: "IM VOLKSGARTEN"
AUTHENTIC AND EXPERIMENTAL VERSIONS
Im Volksgarten

'Ich mochte einen blauen Ballon haben!' sagte Anna.

'Da hast du einen blauen Ballon, Anna!' sagte die Mutter.

'Ich mochte ihm die Luft auslassen,' sagte sie einfach.

'Willst du ihn nicht diesem armen Mädchen schenken, Anna?'

'Nein, ich will ihn fliegen lassen!' Sie lasst den Ballon aus,

sieht ihm nach, bis er verschwindet in den blauen Himmel.

'Mutter, ich hatte ihn lieber dem armen Mädchen geschenkt!' 

'Da hast du einen anderen blauen Ballon, Anna, schenke ihr diesen!' sagte die Mutter.
Im Volksgarten

'Ich möchte einen blauen Ballon haben!' sagte Anna.

'Da hast du einen blauen Ballon, Anna!' sagte die Mutter.

'Ich möchte dem Ballon die Luft auslassen,' sagte Anna einfach.

'Willst du den Ballon nicht diesem armen Mädchen schenken, Anna?!

'Mein, ich will den Ballon fliegen lassen!' Anna lasst den Ballon aus, sieht dem Ballon nach, bis der Ballon verschwindet.

'Mutter, ich hatte den Ballon lieber dem armen Mädchen geschenkt!'

'Da hast du einen anderen blauen Ballon, Anna, schenke dem Mädchen diesen Ballon!' sagte die Mutter.
LIST OF REFERENCES


601-609.


Morgan, J. L., & Green, G. M. (1980). Pragmatics and reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence, and education*. (pp. 113-140). Hillsdale,


