ADVANCED WRITING IN ENGLISH AS A FOREIGN LANGUAGE

A Corpus-Based Study of Processes and Products

Horváth József

Lingua Franca Csoport

ADVANCED WRITING IN ENGLISH AS A FOREIGN LANGUAGE

Published by Lingua Franca Csoport

The publication of this book has been jointly financed by a grant from the University of Pécs (Grant Number 23001) and by support from the English Applied Linguistics Department of the university.

Reviewed by Hollósy Béla

© 2001 Horváth József

ISBN 963 641 850 0

Printed by Agora Nyomda, Pécs

The electronic version of this book is available at
http://www.geocities.com/writing_site/thesis

CONTENTS

ACKNOWLEDGMENTS xi

INTRODUCTION 1

Chapter 1 ISSUES IN WRITING PEDAGOGY: A REVIEW OF THE LITERATURE

Introduction 5

1.1 SLA research and writing theory 6
- 1.1.1 Theory and practice in language education 6
- 1.1.2 The Input Hypothesis 7
- 1.1.3 Writing theories 8

1.2 On the approach dichotomy: Process vs. product 10
- 1.2.1 Research methodology 10
- 1.2.2 Empirical studies 12

1.3 Writing pedagogy: From theory to practice 16
- 1.3.1 Composing for communicating 16
- 1.3.2 Group work 18
- 1.3.3 The *Baseline Study* 20

1.4 Revision: Shaping text by writer and reader 21
- 1.4.1 Revision for grammatical accuracy 21
- 1.4.2 Revision for text creation 22
- 1.4.3 Empirical studies 24

1.5 Responding to writing 26
- 1.5.1 Main variables 26
- 1.5.2 Positive effect of feedback 28
- 1.5.3 Student agendas 29
- 1.5.4 Responding to feedback 30

1.6 Concluding remarks 31

Chapter 2 ISSUES IN CORPUS LINGUISTICS: A REVIEW OF THE LITERATURE

Introduction 33

2.1 Rationale for corpus linguistics 34
- 2.1.1 Data in language analysis 34
- 2.1.2 Competence versus performance 35
- 2.1.3 Lexicography and language education 36

2.2 Corpora: History and typology 38
- 2.2.1 Early corpus linguistics 38
- 2.2.2 The Brown Corpus 39
- 2.2.3 The LOB Corpus 40
- 2.2.4 The London-Lund Corpus 41
- 2.2.5 The Bank of English 41
- 2.2.6 The British National Corpus 42

2.2.7 The International Corpus of English 43
2.2.8 Typology 44
2.3 Current issues in design and technology 44
2.3.1 Corpus development 44
2.3.2 Concordancers: Functions and packages 48
2.3.3 Principles and techniques in corpus analysis 53
2.4 Data-driven learning: CALL with classroom concordancing 55
2.4.1 Computer assisted language learning 55
2.4.2 Discovery in data-driven learning 57
2.4.3 Applications of DDL 59
2.5 Learner corpora: Issues and implications 60
2.5.1 The International Corpus of Learner English 60
2.5.2 The composition of the ICLE 61
2.5.3 Other written learner corpora 62
2.6 Concluding remarks 63

Chapter 3 WRITING PEDAGOGY AT THE ENGLISH DEPARTMENT: PRODUCING PROCESSES

Introduction 65
3.1 Data and participants 66
3.2 Pedagogical concerns and writing in the JPU ED curriculum 67
3.2.1 Principles and their sources 67
3.2.2 Writing in the curriculum 68
3.2.2.1 The undergraduate core curriculum 69
3.2.2.2 The postgraduate curriculum 71
3.3 Syllabus development 71
3.3.1 Objectives 72
3.3.2 Tasks and techniques 75
3.3.2.1 Classroom techniques 76
3.3.2.2 Out-of-class activities 78
3.3.3. Text types 80
3.3.3.1 Personal writing 80
3.3.3.1.1 Reflective essay based on a quote 81
3.3.3.1.2 Descriptive essay on student's dictionary 81
3.3.3.1.3 Essays based on the theme selection table 82
3.3.3.1.4 An essay with two introductions and conclusions 82
3.3.3.1.5 The miniature essay 82
3.3.3.1.6 Completion of a task in a writing textbook 82
3.3.3.1.7 Essay on tape 83
3.3.3.2 Academic writing 83
3.3.3.2.1 Analysis of newspaper content 85
3.3.3.2.2 Analysis of peers' writing 85
3.3.3.2.3 Surveys among students and teachers 85

3.3.4 Readings 86
3.3.5 Feedback and evaluation 88
3.3.5.1 Feedback techniques 88
3.3.5.2 Evaluation 90
3.3.6 Students' views 91
3.4 Future directions 96

Chapter 4 THE JPU CORPUS: PROCESSING PRODUCTS

Introduction 99
4.1 The development of the corpus 100
4.1.1 Conditions of and rationale for data collection 100
4.1.2 Corpus design principles 101
4.1.3 Data input 102
4.1.4 Seeking permission 104
4.1.5 Clean text policy 104
4.1.6 Text types 106
4.1.7 Procedures applied 107
4.2 The JPU Corpus 109
4.2.1 The current composition of the corpus 109
4.2.2 The five subcorpora 117
4.2.2.1 The pre-service data 117
4.2.2.1.1 ES 117
4.2.2.1.2 LPS 117
4.2.2.1.3 WRSS 119
4.2.2.2 The in-service data 120
4.2.2.2.1 RRS 120
4.2.2.2.2 PGS 120
4.3 Analysis of the corpus 122
4.3.1 Hypothesis 1 122
4.3.2 Hypothesis 2 123
4.3.3 Hypothesis 3 124
4.3.4 Hypothesis 4 126
4.3.5 Hypothesis 5 127
4.3.6 Hypothesis 6 128
4.3.7 Hypothesis 7 129
4.3.8 Hypothesis 8 130
4.3.9 Hypothesis 9 131
4.3.10 Hypothesis 10 135
4.4 Pedagogical exploitation of the corpus 138
4.4.1 Learning driven by data from the learner 138
4.4.2 Exploiting for classroom work 139
4.4.3 Guiding individual study 141
4.4.4 Other applications 143
4.5 Future directions 143

CONCLUSION 145

REFERENCES 151
APPENDIXES 167
INDEX 191

LIST OF TABLES

1 Activities for grammatical accuracy 22
2 A typology of corpora 44
3 An example of normalized comparative analysis 54
4 The eight divisions in the core curriculum 69
5 The framework of the current Language Development division 70
6 Research paper options in the Fall 1998 course 84
7 The three types of reading material in the five semesters 87
8 The evaluation categories of the research paper in 1997 89
9 The evaluation categories of the research paper in 1998 89
10 Statistics of scripts in the five subcorpora 111
11 Gender representation in the JPU Corpus 112
12 The 20 most frequent words in the JPU Corpus 112
13 The 20 most frequent content words in the JPU Corpus 112
14 The ten most frequent words in the five subcorpora 113
15 Rank order of the five subcorpora: ratio of hapax legomena 114
16 Contrasting the rank orders by hapax legomena and tokens 114
17 Rank orders of most frequent words in three corpora and the JPUC 115
18 Word forms occurring 100 times or more in the ES 117
19 Word forms occurring 100 times or more in the LPSS 118
20 Word forms occurring 100 times or more in the WRSS 119
21 Word forms occurring 100 times or more in the RRS 120
22 Word forms occurring 100 times or more in the PGS 121
23 Frequencies of contrasting transitional phrases 124
24 Frequencies of "mentioned above"/ "above mentioned" 125
25 Frequencies of the phrase in the non-writing subcorpora 125
26 Thesis statements, topic sentences, and statements of method 127
27 Frequencies of "I think that" in the three corpora 128
28 Distribution of the frequency of *very* in the three subcorpora 129
29 Frequency of each of the five words in the three subcorpora 130
30 Frequency of the two phrases in the three subcorpora 131
31 Content and method types in the 107 research papers 132
32 Rank order of types of introductory sentences in the WRSS sample 134
33 Results of analysis of variance of length of first sentences 135
34 Descriptive statistics of the two mini corpora 136
35 Rank order of types of concluding sentences in the WRSS sample 137
36 Results of analysis of variance of length of last sentences 138

LIST OF FIGURES

1 The variables of response to writing 27
2 Biber's model of cyclical corpus design 45
3 The example text in *Conc's* main window 49
4 Part of the Sorting Parameters dialog window in *Conc* 49
5 Dialog window where words may be omitted from the concordance 50
6 Part of the Display dialog box on the Options menu in *Conc* 50
7 Part of the Concordance window of the program 51
8 The Concordance and the main windows 51
9 The Index window's scrolling list of the alphabetical index of the file 52
10 Part of the Statistics window on the Build menu 52
11 The File menu of *Conc* 53
12 Johns's model of data-driven learning 58
13 Respective weight of each of the eight divisions in the core curriculum 70
14 The relative weight of assessment categories across the five courses 90
15 Number of students selecting values for fairness of evaluation 92
16 Number of students selecting values for assistance from students 93
17 Number of students selecting values for assistance from the tutor 93
18 Number of students selecting values for usefulness of the course 94
19 STD values of participants' evaluations of the four criteria 95
20 Mean figures of the evaluation of the four criteria 95
21 Comparison of usefulness and averages of rest of criteria 96
22 The process of data input 103
23 A window of part of the corpus in the Macintosh file system 106
24 Curricular and course origin divisions of the scripts 107
25 The number of scripts contained in the five subcorpora 109
26 Distribution of texts in the subcorpora: number of scripts 110
27 Distribution of texts: number of tokens in the subcorpora 111

LIST OF APPENDIXES

A Essay titles and themes suggested by the ICLE developers 167
B Formal Writing syllabus, Fall 1996 169
C Writing and Research Skills syllabus, Spring 1997 171
D Writing and Research Skills syllabus, Fall 1997 173
E Writing and Research Skills syllabus, Spring 1998 176
F Writing and Research Skills syllabus, Fall 1998 178
G Metaphors used in the WRS courses 181
H An example of a co-authored essay 183
I The theme selection table used in 1999 184
J Copy of the Permission form 185
K The most frequent word forms in the JPU Corpus 186

ACKNOWLEDGMENTS

First and foremost, I would like to thank the following students for contributing the electronic copy of their scripts to the JPU Corpus:

Adonyi Nóra, Andorkó Réka

Babarci Bulcsú, Bacskay Katalin, Baditz Mihály, Bagi Ádám, Bajomi Lázár Péter, Bajusz Judit, Bakonyi Berta, Baksa Gábor, Balassa Gábor, Balázsi Anikó, Balázsik Eszter, Bálint Anita, Balonyi Beáta, Barabás Mariann, Bán Helga, Bárdi Piroska, Barkaszi Orsolya, Barna Eszter, Barsi Gizella, Bauer Katalin, Bátor Mónika, Belső Adrienn, Bencze Krisztina, Benkő Ágnes, Berecz Mónika, Bereczki Veronika, Berke Ildikó, Bernáth Orsolya, Berta Anikó, Béresné Timár Andrea, Bieber László, Billege Virág, Bizse Ferenc, Blasszauer János, Bleyer Júlia, Bodroginé Farkas Piroska, Bogdán László, Bokodi Judit, Bokori Mónika, Bordás Pálné, Borsányi Valéria, Bradák Anna, Brunner Mónika, Bucsányi Erika, Buzási Balázs

Czeglédi Edina

Csapó Angéla, Csákó Fruzsina, Cseke Ildikó, Cserna György, Csiba Eszter, Csikós Erzsébet, Csipkéné Kocsis Éva, Csongor Csilla, Csonka Gábor, Csontos Zsuzsanna, Csordás Dávid

Daróczi Helga, Demeter Andrea, Dékány Renáta, Dévai Dóra, Dittrich Melinda, Ditz Éva, Dombóváriné Csillag Mária, Dózsa Éva, Döbrösy Ildkó, Dunai Péter

Egri Szabolcs, Elek Attila, Engert Ágnes, Enyedi Szilvia, Erdélyi Dóra, Erős Anikó, Ébel Emese, Élő Veronika

Faragó Edit, Farkas Ágnes, Farkas Gabriella, Farkas Roland, Fehér Fatime, Fehér Györgyi, Fejes István, Fekete Bálint, Fekete Tünde, Fodor Márton, Fodorné Horváth Anett, Fónai Annamária, Földesi Virág, Füzér Balázs

Gaálné Vörös Judit, Gadó László, Gallai Ditta, Gallen Mónika, Gavlikné Robotka Orsolya, Gazdag Zsuzsa, Gál Beáta, Gelencsér Péter, Gellénné Káldi Szilvia, Geresdi Annamária, Gergely Anikó, Gerhát Zsófia, Gertner Brigitta, Gombás Éva, Gregóné Szili Katalin, Grosch Andrea, Gulyás Judit

Gyallai Krisztina, Gyenisné Oszbach Mónika, Györök Tímea, Gyulavics Ildikó

Haas Klára, Hadnagyné Kovács Judit, Hajdú Anett, Harmat Ágota, Harmath Virág, Hartmann Edina, Hatta Diána, Hazafy Krisztina, Hegedűs Anita, Hegedűs Éva, Hencsei Vilmos, Herkovits Margit, Hetesi Sándor, Hirth Márkus,

Horony Bernadett, Horvai Krisztina, Horváth Eszter, Horváth Gergely, Horváth János, Horváth Judit, Horváth Miléna, Horváth Nóra, Horváth Sándor, Hrivnák Orsolya, Huber Vera, Hubert Andrea, Huszanagics Melinda, Huszár Zsuzsanna, Huszárné Szabó Mária

Illésné Gesztesi Piroska, Inczédy Lászlóné, Iván Lucia

Jaskó Veronika, Jávorszky Mária, Jónás Ildikó, Józsa Anita, Juhász Hedvig,

Kalmárné Máriás Ildikó, Kamper Gergely, Karádi Orsolya, Karagits Menyhért, Kaszás Henrietta, Károly Tünde, Kárpáti Ildikó, Kecse-Nagy Katalin, Kertai Hajnalka, Keszthelyi Erika, Kisgyörgy Péter, Kiss Ágnes, Kiss Gabriella, Kiss Judit, Kiss Eszter, Kohajda Edina, Kokas Marianna, Komáromi Csilla, Kondor Edith Mária, Kopjár Ivett, Kovács Attila, Kovács Zsolt, Kópis Andrea, Kövesdi Veronika, Kőberl Kinga, Krammerné Gehring Éva, Krigler Gábor, Kripner Mónika

Láng Katalin, Láng Rita, Lászlai Noémi, Lehmann Magdolna, Leksz András, Lengvári Ágnes, Lénárt Henriett, Luspai Gabriella

Madarász Mónika, Magné Turcsics Tímea, Magony Tímea, Magyar Zsuzsanna, Magyarfalvi Dóra, Makra Melinda, Marosváriné Deák Zsuzsa, Matheisz Éva, Matolcsi László, Mádai Gellért, Márk Mónika, Márki Miklós, Márkus Melinda, Meggyesi Gábor, Merényi Adrienn, Mészáros Krisztián, Mikola Erzsébet, Mojzes Kamilla, Molenaar P. Gábor, Molnár Krisztián, Mondok Ágnes, Muth Imre

Nagy Ágnes, Nagy Borbála, Nagy György, Nagy Elvira, Nagy-Mélykuti Luca, Nemes Dóra, Németh Erika, Németh Piroska, Nógrádi László

Oláh Márta

Öllős Anikó

Paczolay Éva, Palla Lívia, Pandur Zsanett, Papósné Horváth Judit, Papp Árpád, Papp Eszter, Pádár Dénes, Páli Noémi, Perjés Beatrix, Pete István, Pető Orsolya, Petz János, Pércsich Richárd, Pétersz Tamás, Pfeifer Bernadett, Pintér Katalin, Piskó Beáta, Poturak Laura, Povázsai Zita, Pónya Klára

Radnai Tamás, Rácz Emese, Rátkay Orsolya, Rekettye Gábor, Renkó Zsuzsanna, Ribiczei Zsuzsa, Rohonczi Krisztina, Rohonyi Borbála, Roll Eszter, Roza Szilvia, Rozgonyiné Inke Krisztina, Rósa Gézáné

Saskői Kinga, Sata Előd, Sárfy Mária, Schlögl Krisztián, Schönberger Tamás, Schram Ágnes, Schrottiné Tamási Monika, Schubert Gábor, Scsaurszki Rita, Simon Györgyné, Simó Boglárka, Sinka Hilda, Skultéty Judit, Som Katalin,

Somorjainé Orosz Andrea, Soós Zoltán, Srankó Nóra, Stemler Kata, Süle Brigitta

Szabados Péter, Szabó Gabriella, Szabó Ildikó, Szabó Imre, Szabó Soma, Szakály Zsófia, Szalontai Mária, Szarka Brigitta, Szántó Györgyi, Szathmári Timea, Szegedi Csilla Eszter, Szécsi Tünde, Sziklai Gábor, Szmolenszky Tünde, Szokol Nóra, Sztanics Lidia, Szűcs Mariann, Szűcs Rita

Takács Anita, Takács Zsolt, Tánczikné Varga Szilvia, Tánczos Balázs, Tarnai Dóra, Tarnai Gábor, Tihanyi László, Tímár Zsuzsa, Tokorcsi Éva, Torma Szilvia, Tóth Andrea, Tóth Ágnes, Tóth Boglárka, Tóth Mónika, Tóth Tamás, Tövisháti Lilla, Tőke Orsolya, Trapl Szilvia, Triznay Zsuzsa, Trócsányi Balázs, Turi Szabolcs, Turiák Tibor, Túri Andrea, Tüű Orsolya

Urbán Miklós, Úz Márk

Vadon Balázs, Vajda Violetta, Váciné Nyári Eszter, Vámosi Gyula, Varga Bernadett, Varga Tamás, Vargáné Uzsoki Márta, Végh Orsolya, Viszlai Eszter, Vízhányó Barbara, Vlaskovits Dóra

Wagner Zita, Wittinger Katalin, Wollent Ida

Záborszky Zoltánné, Zajzon Tamás, Zongor Judit

Zsiborás Eszter, Zsigovicsné Kovács Margit

I gratefully acknowledge the assistance I have received from the following colleagues and organizations:

The British Council in Budapest, the East European CALL Centre, Sylviane Granger, IATEFL Hungary, Tim Johns, the late Kohn János, Kevin Mark, Steve Starkey, Szabó Gábor, Szépe György, Macey Taylor, the TESOL CALL Interest Section, and Sarah Turner.

Special thanks go to Andor József, my Ph. D. consultant, for his insightful comments and orientation. I wish to gratefully acknowledge Nikolov Marianne's advice, criticism and encouragement. Kiszely Zoltán has shared with me valuable insights. I also appreciate comments on the Ph. D. study version of this book made by Bárdos Jenõ, Lendvai Endre, Magay Tamás, Stephanides Éva, and the two opponents, Budai László and Hollósy Béla.

I am indebted to the reviewer of this book, Hollósy Béla, for his suggestions.

The publication of this book has been made possible by a grant from the University of Pécs and by support from the English Applied Linguistics Department of the university.

And thank you, Bodza, Tamás, and Dávid.

INTRODUCTION

When clicking on the Print button of my word processor to produce the hard copy of a university course syllabus, I initialize a period of time that will hopefully engage students and me in the discovery of new aspects of meaning in the writer—reader relationship. As the ink-jet chugs on, I muse on how what is planned will be implemented in the classroom and in private consultations.

The syllabuses I designed and produced between 1996 and 1998 primarily targeted students who registered for mandatory pre-service undergraduate and optional in-service postgraduate courses at the English Department (English Applied Linguistics Department since September 1998) of Janus Pannonius University, Pécs. The first written product a student received from me had to be perfect in every respect: it had to address the reader so that she or he felt the course was designed with individual needs in mind. It had to provide all the necessary information to set the context of exploration and learning for what was to follow. And it had to arouse curiosity in the content of the sessions and the content of the written assignments to complete.

By 1996, when I first met such a group of students, I had been teaching at the department for seven years. Since 1992, I had also been collecting student scripts by those participants in Language Practice, Computer Assisted Language Learning, Methodology, and Introduction to Indian Literature in English courses who were willing to share with me the electronic copy of their essays and research papers.

Between 1992 and 1999, I collected such scripts from over 300 students—as of the end of January 1999, the corpus consisted of over 400,000 words. By sharing with me their ideas, findings, and opinions in print and on disk, these students have enabled me to gather information for the study of written learner English as a Foreign Language (EFL).

This book is concerned with the description and analysis of advanced writing in EFL. It provides a curricular and syllabus development focus as it takes account of writing pedagogy processes at Janus Pannonius University (University of Pécs since 2000). The course content of undergraduate and postgraduate English-major students was studied. Using authentic records, the study attempts to cover a wide spectrum of issues related to EFL students' writing skills in a variety of text types. The description and analysis of over 300 students' scripts, in the JPU Corpus, is presented to address the aspect of processing products.

This is a cross-disciplinary undertaking: it is informed by writing pedagogy via classroom observations made over the years of Writing and Research Skills courses. It is also motivated by current empirical interest in exploiting machine-readable collections of written and spoken texts for language description, lexicography, discourse analysis and corpus-based language education techniques such as data-driven learning. The fundamental question it attempts to explore and answer is how the description of scripts written by advanced Hungarian university students of EFL can contribute to an understanding of writing processes and products.

Why develop a learner corpus? The endeavor holds potential benefits in at least three areas, each of which will be explored in this book:

- to collect evidence of language use;
- to serve as a basis of research;
- to serve as a basis of innovative pedagogical application.

The cross-disciplinary framework of the study means that to present these subjects, I have drawn on recent writing pedagogy and corpus linguistics. On the writing pedagogy pane, a wide spectrum of relevant factors will be presented. To be able to provide a thorough investigation of EFL writing pedagogy, issues such as the following will be discussed and systematized: writing theories, curriculum development, writing instruction procedures, assignment and course goals, feedback types, revision strategies, and the role of peer revision.

On the corpus linguistics pane, the following areas will be considered: the theory of performance and competence, theoretical arguments for and against corpus evidence, the nature and empirical use of corpus matter, and data-driven learning that exploits both native and learner corpora.

Chapter One sets the context of the study by providing a description and analysis of theoretical issues and empirical research in the fields of contrastive rhetoric, writing pedagogy and materials development. It presents the outcomes of continued cooperation between the teaching profession and academia. After evaluating the claims product- and the process-oriented writing instruction has made, it concludes by setting the research agenda for integrating learner writing development procedures with the method of corpus linguistics.

Chapter Two then pursues how this can be done by presenting relevant corpus linguistic research. After an analysis of the underlying theoretical considerations and a historical overview of the development of the corpus linguistic method, it aims to provide a detailed explanation of variables in corpus planning, development and manipulation. The chapter ends with the discussion of the specific nature of learner corpora, the development of which represents an exciting new vista in both language pedagogy and corpus linguistics. By presenting the composition and application of the International Corpus of Learner English, the chapter concludes by explaining that the study of learner scripts can contribute to enhancing the authenticity of writing pedagogy.

Following the reviews, Chapter Three presents, employing a mix of qualitative and quantitative data, the writing pedagogy procedures at the English Department of Janus Pannonius University, focusing on the Writing and Research Skills courses I taught in the past three years.

Chapter Four presents compositional details of the JPU Corpus and the results of empirical research. It examines linguistic data drawn from the corpus and a set of examples of the pedagogical exploitation of that data for writing skills development. As will be seen, the largest Hungarian EFL learner corpus offers opportunities to describe the lexical and text organization patterns of written learner discourse. Another contribution of this chapter to the field is the collection of concordance-based descriptions and evaluations of learner scripts, which allow for the development of study guides for individual and group use.

The study's conclusions suggest that the JPU Corpus has the potential for further nationwide, and possibly international, cooperation between corpus linguists and writing professionals.

Chapter 1

ISSUES IN WRITING PEDAGOGY: A REVIEW OF THE LITERATURE

> *Writing, because it allows us to represent to ourselves our learning, our ways of making meaning, teaches us the most profound lesson about how we read, write, and use language, about what it means to know.* (Zamel, 1992, p. 481)

Introduction

Writing is among the most complex human activities. It involves the development of a design idea, the capture of mental representations of knowledge, and of experience with subjects. The interlocking processes of writing by novice and expert authors have been studied by such diverse disciplines as cognitive psychology, stylistics, rhetoric, text linguistics, critical literary theory, hypertext theory, second language acquisition, and writing pedagogy. From such a wealth of approaches and themes, this book will be concerned with what is immediately relevant to the teaching and learning of writing in EFL at advanced levels.

This chapter proposes to set the context of investigating written learner English at university level. A descriptive and analytical undertaking, such a project needs to be informed by general second language acquisition theory, research design considerations and specifically by the results of research in writing pedagogy. I will present the theoretical framework of my study and then review the literature that has shaped the present project.

The chapter is divided into six sections. In the first, a general introduction to second language acquisition (SLA) research and writing theory will set the context of the issues considered in this book (1.1). The notions and practice of product- and process-oriented writing instruction will feature in the next section (1.2). Narrowing down the scope of investigation, the following section aims to systematize what is known about the practice of writing pedagogy (assignments, course goals, and writing instruction procedures, in 1.3). Focusing on the interaction between teacher and learner, and among students, section 1.4 will elaborate on revision strategies, and the role of peer revision. The literature review will then present the theory and practice of feedback that students receive on their scripts (1.5). The concluding section (1.6) will synthesize the most important strands of the literature.

I hope that after this discussion, the present research agenda for integrating learner writing development with the method and findings of corpus linguistics will have been made explicit.

1.1 SLA research and writing theory

1.1.1 Theory and practice in language education

In reviewing and critiquing SLA research traditions and trends, Ellis (1998) pointed out that much of the effort was either theoretical or pedagogical. He argued for a model whereby the communication between researchers and teachers can take the form of one of three types: research informing pedagogy, research informed by pedagogy, and research and pedagogy interacting to address theoretical and practical concerns, and emphasized the importance of the last approach. He also argued that any SLA theory can only be applicable by language pedagogy if it is relevant to it (Ellis, 1995): the goals of the theory must be compatible with the aims of teaching.

A similar proposal was made by Brumfit (1995) in the discussion of teacher professionalism and research. Offering his views on British educational policy and on the needs for integrating global SLA research with local observations, he suggested that for classroom practice descriptions to be significant, one needs to consider the common variables in different language teaching contexts (p. 41). Specifically, Brumfit suggested that educational research needs a systematic program, rather than focusing on fragmented projects.

Three strands of investigation were suggested (Brumfit, 1995, pp. 39-40). The first ought to describe classroom practice so that events, attitudes and policies are spelled out. The second should take on to explain what was found in the first phase: drawing on the data gathered, theory needs to construct models to be able to adequately structure that knowledge. Third, studies directed at the pedagogical processes need to extract what ought to take place in education from what is happening there. Brumfit argued that these three approaches will enable empirical research to establish the program orientation.

For the field to arrive at valid conclusions on the acquirer of language, Larsen-Freeman (1991) pointed out the importance of studying and describing the learner. Reviewing research into the differential success of acquiring a second language (L2), she critiqued findings related to variables such as learner age, language learning aptitude, attitude and motivation, personality, cognitive issues and learning strategies. She concluded that future research and language education will need to corroborate findings and test such hypotheses as the following: learning is a gradual process; it is not linear; unless learners are ready to proceed to new phases of learning, no long-term acquisition takes place. In a discussion of instructed SLA research, Larsen-

Freeman and Long (1991) called attention to the need to study the ways in which instruction affects SLA. For this process to be studied, they suggested that linguistic input sequence and frequency should be operationalized, together with those tasks that learners are exposed to in the classroom. By studying these variables, SLA theory can integrate action research findings initiated by the teacher (Larsen-Freeman & Long 1991, p. 327), a proposal similar to that made by Ellis (1995; 1998) and Dörnyei (1997).

In many ways, the tasks the language educator faces in teaching and in initiating research and the tasks in which learners perform have common features. Both aim to integrate what is known with what is being learned about the situation or the language item being studied. Yet there are crucial differences, too. In a discussion of the interface between language learning theory and practice, Prabhu (1995) offered a four-component model to describe this relationship. These are the ideational (concepts and processes of language learning), operational (pedagogical practice), ideological (social variables), and managerial (pedagogical decision-making). As far as the operational module is concerned, Prabhu pointed out the contrast between teaching and learning, saying that while teaching can be planned and sequenced, learning follows a route based on mental processes that are difficult to observe.

1.1.2 The Input Hypothesis

However, there is a growing body of research evidence on the rate of acquisition and the optimal conditions for successful acquisition to occur. In this area, the work by Krashen has shown direction. The Input Hypothesis (Krashen, 1985) claims that to ensure long-term success in language acquisition, there must be comprehensible input. The theory comprises five hypotheses, of which the Monitor Hypothesis and the Affective Filter Hypothesis are particularly relevant for writing research.

The Monitor Hypothesis is concerned with language production—the ability to use language is a result of competence based on acquisition, while learning acts to enable speakers and writers to "change the output of the acquired system before [they] speak or write" (Krashen, 1985, p. 2). For this monitor (or editor) to operate, Krashen hypothesized, the user needs to be aware of the importance of accuracy, and the rule stating correct forms should be present. The Affective Filter Hypothesis states that for comprehensible input to become intake, a mental block should be lowered: this can occur when the speaker is self-confident, and when a potential failure to produce the necessary language is not seen as a risk. Krashen added that for the filter to be down, the speaker must focus on the message. This model of language acquisition was partly based on Krashen's survey and evaluation of theoretical work in applied linguistics, and on investigations of skill-specific empirical research, also motivating subsequent work on the implications of the hypothesis in language education. Particularly relevant of these studies is his summary of writing research (Krashen, 1984) and a recommendation for a reading-based program (Krashen, 1993).

In the writing study, Krashen (1984) hypothesized that his generic SLA hypothesis of comprehensible input held for the development of writing skills, suggesting that extended reading was necessary for organizational and grammatical improvement to occur. He analyzed a wealth of case studies that confirmed the hypothesis: the acts of planning, rereading, scanning, revising for clarification occurred significantly more often and with better results in good writers who also reported pleasure in reading. Also, while less able writers were shown to have much more difficulty in transferring what is known as writer-based prose to reader-based prose, more apt writers had less difficulty to consider readers' needs. Krashen concluded that although formal instruction of sentence-level rules can help improvement in writing, for significant and successful writing development to occur, this may only be a complement to receiving comprehensible input via reading.

In the reading-focused work, Krashen (1993) presented the framework and application of a program that allows the extensive use of what he called "free voluntary reading." Investigating the relationship between writing instruction and learning, he reported that because the rules of formal writing are far too complicated to learn, style does not result from more writing practice but from more reading. Opposing the view that "we learn to write by actually writing" (Krashen, 1993, p. 73) he claimed that improved writing quality, and the ensuing discovery of one's own style, is a result of frequent reading. Bárdos, in his encyclopedic survey of FL teaching (2000), also calls attention to the integration of reading and writing, as the processes involved in both (message identification, processing, comprehension, and expression) also become part of an individual's overall experience (p. 160).

1.1.3 Writing theories

For decades, the most influential paradigm of writing was contrastive rhetoric, proposed by Kaplan (re-assessed in 1983). The contrastive rhetoric tradition focused on the product of writing and established prescriptive approaches to the teaching of writing. Kaplan claimed that in English, writers tended to develop their thoughts in a linear fashion, advancing a thesis, forwarding supporting evidence in sequentially presented topic sentences, developed in unified paragraphs. The aim of writing pedagogy was to compare and contrast the text organizing patterns in the L1 and L2 and thus facilitate acknowledgment of differences. The primary technique in the classroom was imitating paragraphs so that the patterns were practiced. Raimes (1991) noted that this tradition was the dominant approach up to the mid-70s, when the focus shifted to the writer and the context of writing, and thus to a more process-oriented analysis of writing and writing pedagogy. The latter trend also coincided with greater emphasis on language as communication, focusing teachers' attention away from form as prescribed by controlled-traditional rhetoric to collaboration between teacher and student, and among the students themselves.

Particularly influential was the work of Hayes and Flower (1980) and Flower and Hayes (1981), who developed a cognitive theory of writing processes, eliciting information directly from writers via think-aloud protocols and observations (Grabe & Kaplan, 1996, p. 91). They proposed a model that was based on three tenets:

- processes of writing, such as planning, organizing, reviewing, and evaluating, often interact with each other;
- writing follows a goal the writer is aware of;
- processes are performed differently by experienced and inexperienced writers.

The theory identified a task environment (made up by a rhetorical problem and text produced), the three major components of the writing process (generating, translating and reviewing), each of which is controlled by a monitor. In proposing this model, Flower and Hayes also generated much needed empirical research.

One result of this research was that the use of protocols came under heavy criticism: it was argued that the validity of the model that relied on writers aiming to explain what they were doing while they were engaged in writing was limited. In response to the need for theory building and for validating theory in research, Bereiter and Scardamalia (1987) offered a new perspective: instead of bringing together factors characteristic of novice and expert writers, they suggested that different models can describe different levels and contexts. Basically, their two-process theory aimed to explain how and why differences occur in inexperienced and experienced writers' performance.

Two models make up the theory. The first is called "knowledge-telling," which involves the processes of inexperienced writers, and the second is "knowledge-transforming." In both, the writer considers three main factors: knowledge of content, knowledge of discourse, and ideas of a writing assignment. However, the first is primarily a step-by-step operation that is engaged as the writer collects material and lexis, whereas the second includes the writer's identification of a unique problem and goal so that the writing becomes essentially a process to solve the rhetorical problem. The first model describes the less experienced writer, whereas the second the expert writer. How one proceeds from one level to another, however, was not shown explicitly.

According to Silva (1990) the development and pedagogical application of these cognitive models meant a decreasing concern with error in English as a Second Language (ESL) and EFL. The emerging paradigm of the process approach called for a much more positive and encouraging setting, a workshop-like environment (p. 15). Still, as Leki noted (1991), contrastive rhetoric still has much to offer to language teachers: The information a contrastive analysis reveals of L1 and L2 text structures can contribute to what teachers and students regard as successful communication (p. 137).

In the nineties, one could witness a wide variety of writing pedagogy and research, applying and critiquing both major traditions. As noted by Raimes (1991), the field had come to acknowledge the complexity of the composing process, with individual research projects focusing on the central issues of form, the writer, content, and the reader (p. 421): an ethnography of writing was being produced (Silva, 1990; Atkinson & Ramanathan, 1995; and Leki, 1995 are among the recent examples of such endeavors). This recognition has a number of implications for theory and practice: the field had to gather more data on novice and expert student writers' performance, on the writing processes applied in various classroom settings, both L1 and L2, on the social contexts of pedagogy, and on how teachers themselves may initiate research into their practice.

1.2 On the approach dichotomy: Process vs. product

A central concept in recent FL and SL writing theory has been the binary nature of the process of writing and the product of writing. As has been noted in the previous section, much of what is known about the ethnography of student writing comes from the theory of L1 writing. As the models proposed by Hamp-Lyons (1986, 1989, 1990), Kaplan (1983), Leki (1995) and Silva (1993) attest, however, not all features of writing in the native language may be transferred to FL and SL writing. The process of producing various types of written discourse will be affected by such factors as involvement with the topic, awareness of the writer's individual rhetorical skills, interaction with a real audience, and how feedback on ideas presented in drafts is provided. While these matters will depend partly on the individual writer's own experience (or lack of it) in the first language, and the importance of writing (or lack of it) in the native culture, there are other variables that need explanation. In this section, then, influential studies will be reviewed with the aim of showing multiple approaches to the process—product dichotomy.

1.2.1 Research methodology

I will begin this discussion of process and product by a brief introduction to the recent history of writing research methodology, based on Krapels (1990). Focusing on L2 research conducted in the 1980s that aimed to corroborate the findings of L1 studies, she reviewed the multiple scholarly efforts that went into designing valid and reliable models and on this basis suggested fields for further investigation. The scope of models and participants is rich, and

Krapels' state-of-the-art review will continue to generate future studies. The repertoire of L2 research models includes

- case studies that involved a few participants in one writing task;
- studies that focus on females, advanced L2 students, undergraduates, native speakers of Spanish and Chinese, and students who were not chosen randomly (often the students enrolled in the researcher's classes);
- studies that report on tasks ranging from one to all required tasks in a course;
- studies that investigate the discourse of narrative and expository writing;
- studies that vary in type of topic and in time allowed for completion;
- studies analyzing data from product- and process-orientation (based on Krapels, 1990, pp. 48-49).

The repertoire of L2 composition findings includes claims such as the following:

- Limited competence in writing in English results from limited composing skills;
- Some composition processes of less skilled L2 writers share features of those of unskilled L1 writers;
- L1 writing processes transfer to L2;
- The processes of composition differ slightly in L1 and L2;
- In generating L2 writing, L1 is sometimes used;
- Some tasks and topics tend to trigger more L1 language use than do others (based on Krapels, 1990, pp. 49-50).

Based on this review, Krapels set the following research agenda for future studies: first and foremost, more ethnographic research could deepen the understanding of the processes as identified by the student writers themselves, even though in such studies comparability will be problematic. In terms of research questions, Krapels proposed that writing research investigate the relationship between rhetorical preferences in the first language and the writing processes in the L2. Another area for empirical research is the role writing has in the L1 culture and its impact on L2 writing processes. Perhaps most important, from a pedagogical point of view, will be the studies that look into how different types of L1 writing acquisition and learning affect development in L2 writing processes.

For an in-depth understanding and evaluation of writing pedagogy issues, Silva (1990) claimed that teachers and researchers in the field have to evaluate approaches based on a clear set of principles and that they need to conceptualize these approaches in a model that takes account of the following factors (p. 19):

- the theory that underlies the approach,
- empirical research that supports the theory, and
- the validation of the approach.

Silva (1990) proposed, on the basis of these three components, that an evaluation of any writing pedagogy approach or set of procedures in the field of ESL composition must consider the actors and the acts of writing instruction, including the writer, the reader, the text (read and produced), the context (pedagogical and cultural), and the interaction (among actors and acts). Besides, such an evaluation can result in a valid writing pedagogy theory and reliable research instruments for assessing how effective these approaches are. It is then, he argued, that research and practice may be able to establish and maintain high standards in the field (Silva, 1990, p. 21).

1.2.2 Empirical studies

Zamel (1992) set out to dissect how the complementary processes of reading and writing can be integrated. Holding the view that one cannot even begin to understand what goes on in the writing mind without reflecting on how writers interact with texts as readers, she proposed, following Krashen's (1984) and others' framework (Raimes, 1992; Hansen, 1987), that a full integration of reading and writing skills development was necessary to enable L2 writers to experience how readers interact with texts. She also aimed to recommend practical applications for the classroom. Among the factors analyzed were the processes of making meaning in reading, interacting with text, and raising awareness of reader's goals. Through these processes, she argued, students can make the process of discovering the importance of goal and audience in writing more valid. The activities suggested were logs, reactions, and sharing with other students. She pointed out that

> because these activities allow students to actively engage and grapple with texts, to explore how and why texts affect them, [they] can make discoveries about what other readers do with texts *they* compose. They come to realize that if reading involves reconstruction, they must help guide readers of their own texts in that reconstruction.... (Zamel, 1992, p. 481)

How this realization may take place with the help of writing pedagogy can, of course, be impacted by what views teachers hold of the processes involved in making meaning. For this purpose, a study aimed to elicit answers from the teachers themselves. Caudrey (1996) conducted an electronic mail survey among ESL teachers to investigate how they define and apply processes and products in their own teaching. He found that many came to adopt an approach that combines the two elements—one that stresses that "the writing process is a means to an end" (p. 13). While this was a positive finding that one could predict, the other major revelation was that a number of respondents applied the process approach rather rigidly, sometimes with whole classes of students "moved through the writing process...*in* step with each other" (Caudrey, 1996, p. 13). In other words, there were a number of instances, according to the answers, of a singular process being applied as opposed to multiple processes encouraged to engage a more cyclical application of writing processes. As the sample of the teachers involved in the survey was small, however, this finding may need to be substantiated in a follow-up study. Obviously, the practice of integrating various types of process approaches, the classroom sequences and syllabuses of these courses need further investigation, factors that the survey did not address.

Such concerns were emphasized by Bloor and St. John (1988), White (1988), Tsui (1996) and Davies (1988). The authors described task types and processes initiated by writing teachers that provide insight into the intricacies of process versus product. Using an English for Specific Purposes (ESP) project writing task, Bloor and St. John (1988) argued that this type of activity addresses the distinct needs of the students involved and engages them in learning language. In their classrooms at two British universities, EFL students were assigned to write field-specific project reports and to prepare oral presentations. An advantage of the project was the integration of writing and speaking by incorporating an oral task. Besides, the sheltered nature of project writing was a factor that students welcomed, according to the authors (Bloor & St. John, 1988, p. 90). The task set involved the following elements: a preparatory reading to set the context and genres for the writing task; a specific purpose for reading specialized literature; and a procedural methodology that ensured that students were focusing on meaning. As for the teachers, they focused students' attention on being readers and writers at the same time, so that during each phase of producing the project they could reflect on task achievement.

In an exceptional case study, Tsui (1996) introduced a writing ESL teacher (Li) who claimed to be dissatisfied with the method and techniques she had applied. A Chinese national, she had some experience in teaching writing but was frustrated in her efforts. She was also aware of the frustration many of the Hong Kong students she taught had. The source: the time consuming and often exhausting activities that were applied in the writing classes. Tsui gathered multiple types of data (the teacher's reports, scripts by students, observations of classroom and conference interactions, and student interviews and evaluations) to track down the process and product of how this teacher implemented a process approach to tackle the frustration and to learn how to better teach ESL writing. The most relevant finding of the project was that Li first introduced process-writing types of activities in her classes, then reverted to more traditional product-type tasks, and finally she began to adopt modified versions of process-type tasks, showing a development in her teaching skills and in her understanding of different student needs and skills. Especially revealing is how she reasoned for the changes that occurred in her teaching:

> I found myself in situations [in] which I had to abandon what I planned and react to the needs of students. I need to be not only more sensitive to needs but also more flexible. (Tsui, 1996, p. 116)

As flexibility in teaching can sometimes prove taxing for a non-native teacher (Medgyes, 1992, 1994), this intervention in one's own teaching orientation for the benefit of the learning outcome is well worth further investigation. Nevertheless, there is already research evidence of the need for flexibility in the development of a writing course syllabus itself. As shown by Davies's theoretical framework (1988), the process of working out a genre-based syllabus in which ESL students' needs in terms of the discourse requirements of their respective fields were accounted for is an area that can benefit from collaboration between students and teachers. Davies presented the duality of process and product by calling attention to a crucial factor of process for L2 writers: for them, writing is partly a mode of capturing meaning about the world, and partly an experience with which to learn "about a language through writing" (1988, p. 131). That is, while doing research, taking notes, formulating theses and gathering supporting evidence, the L2 writer will gather information about the subject *per se* and the language with which to express knowledge about this subject.

For both ESP and English for Academic Purposes (EAP) students, Davies argued, this necessitates a teaching approach that integrates reading and writing, focusing on the text types or genres that these students are exposed to and are required to produce themselves. In the actual development of the syllabus, then, the teacher's role is to engage in what Davies called an "open-ended collaborative analysis" (1988, p. 133) that will provide the necessary experience in the target types of texts. She also added that for these aims to be met, a writing syllabus needs to stimulate confidence.

While this framework emphasizes collaboration, there is little evidence to support the claim that the approach did stimulate more confidence in students than other syllabuses. A different perspective, and a different type of collaboration, was adopted by Boughey (1997), who investigated how activities designed for large groups of students enabled them to integrate reading and writing. In her study, 30 tertiary multilingual ESP students participated in one writing task activity, complemented by reading collections and studying handbooks. Boughey reported that because the writing task was set up for groups of students, the teacher was able to afford to give more detailed feedback. Besides, for the students in the groups the notion of audience was much less abstract. They also reported they were less shy than otherwise, and that the amount of research that the participants carried out would have been much smaller if the task had called for individual effort. The conclusion seems well founded: such experiences of writing as part of a group can become additional vehicles of generating context and dealing with the inherent problems of a large class. The drawback that some students reported reluctance to participate as members of a group can be minimized if students have the option of choosing writing tasks in which they would prefer to work individually or as members.

For participating in a writing program that adopts the process approach, a model was proposed by Singh (1992), who suggested that the three main steps are not rigid but can overlap or come in a different order depending on the nature of the writing task or individual needs. The steps are as follows:

Stage 1: planning
Stage 2: drafting
Stage 3: revising

At each step, a different set of functions and activities is emphasized. While planning, the writer generates ideas, surveys possibilities, decides on how to tackle the task and on how to order units, and chooses suitable information. While drafting, the student reviews any notes produced in the first phase and identifies problems. It is clear that an overlap has already occurred here: planning does not seem to involve any writing, yet in the model the second phase refers to text generated, and it already includes a revising element in the problem identification activity. In the last phase, the writer revises by checking text, eliminating errors found, and by rewriting to incorporate elements that enhance purpose and readership awareness. What is less elaborate in the model is how the stages are performed by individual students and what the role of the teacher is.

The foregoing review of the products of processes has focused on studies conducted in traditional off-line classrooms. In such environments, the participants meet in a regular classroom, discuss and negotiate face-to-face, produce drafts, reflect on readings and on feedback. Often, there is an opportunity for student-teacher writing conferences, either in a time-tabled office-hour slot or as part of the services of a writing center. But times are changing, and now there is an ever-widening pool of students served by non-traditional on-line classes dedicated to writing skills development. The processes of writing are affected by the technology that these classes make available, and so is the repertoire of teaching.

The environment that a course where learning is facilitated by computer-mediated collaboration was studied by Warschauer (1997). He identified seven features that are specific to online communication. Of these, the fact that such interaction can take place between multiple users, that it is independent of time and place, and that it can be accessed across a distance appear to be most significant in the long run. As the author noted, much as such collaboration may be potentially useful for participants, empirical research was necessary to establish how or whether traditional "transmission" approaches (Warschauer, 1997, p. 478) were being modified.

In Sullivan (1998), this type of environment was introduced and studied empirically. Using classroom transcripts as her data, Sullivan found that the ethnically mixed class of university students engaged in more interaction, as the computer-assisted setting fostered collaborative learning and social interaction. This did not result in improvement in language accuracy, but it did contribute to an increase in the quantity of language performance. It was also claimed that by interacting part of the time by computer, the minority students had more valuable opportunities for self-exploration and expression. As the study did not intend to add triangulation to the data, some of the claims call for further validation; however, the application of such technologies will probably continue to affect both writing instruction and research.

1.3 Writing pedagogy: From theory to practice

We have seen the development of the theory of writing, and the pedagogical decisions that aim to apply the results in ESL and EFL language education. In the following, a transition to the pedagogical practice will aim to highlight how such views have penetrated the classrooms of writing pedagogy, first by reflecting on syllabus and materials development, and then by describing and evaluating classroom procedures.

1.3.1 Composing for communicating

Leki and Carson (1997) were concerned with English for Academic Purposes and specifically with the writing experiences of ESL students in university courses in the U.S. Zinsser (1988a, 1988b; 1998) formulated a professional's view, whose major contribution was to draw attention to the individual reader's and writer's need for simple, uncluttered text. Research by Bello (1997), Cook (1996), Dickson (1995), Hoppert (1997), Kail (1988), Kerka (1996), Kirschenbaum (1998), Meyers (1997), and Ronesi (1996), among others, highlighted such diverse issues in writing pedagogy as general writing skills development, the ways in which reading and writing can be applied integratively for novice writers, the application of journal writing with adult learners, the setup and running of writing centers, and copyright matters. In these papers, a personal voice of aiming to improve was distinct, as was the recognition that even more research and innovation was necessary.

Raimes (1983a) posited that writing is a cognitive and learning experience that helps us to "find out what we want to say" (p. 261). Reflecting on how the grammar- and drill-focused tradition of writing instruction failed to elicit real communication between real writers and readers, she called attention to the composing element of the tasks labeled as "controlled composition." She suggested that in many of these activities, control was paramount, and little composition was being facilitated. To tackle the frustration that ESL students in the U.S. had (in her teaching experience involving tertiary-level students of academic English) with sentence- and paragraph-level problems, Raimes offered three recommendations, each addressing a distinct part of the process of writing instruction. Much of what she stated in this study seems to have been adopted, and so it is useful to review the principal recommendations.

First, the assignment for writing should not be reduced to some concrete or abstract theme or topic—the act of assigning must contain suggestions and guidance to complete it. Arguing that the processes of writing are not rigid entities, Raimes encouraged a cyclical, rather than linear, application of the processes of prewriting, writing and revising. Second, marking papers should involve not only mere corrections of grammatical errors, but also the process of conferencing with students, explanation and praise. Third, a combination of writing and reading tasks enables students to predict, such as in a specially designed cloze-test task, and in activities that aimed to develop a sense for tone of writing and word choice, thus letting students "see that they really know a lot about tone and textual and thematic development" (Raimes, 1983a, p. 269). Other techniques that also aimed to turn the writing class into a composing and thinking class are described in Raimes (1983b).

The complementary processes of composing and thinking were approached from a science writing perspective by Andersen (1988), who was concerned with how ESP students of English as a SL working in a specialized field were able to attain success during their university years and later in their chosen careers. Placing this writing pedagogy issue into the British social context, he proposed that overwriting was a distinct feature of much scholarly writing. Reviewing research that analyzed the acceptability of writing styles, complexity, content, and affiliation of scientific writing in English, and drawing on his own experience, he found that "clear and simple writing is produced by only a small minority of authors" (Andersen, 1988, p. 152). (For the professional writer's views on simplicity, see Zinsser, 1988; 1998). Although Andersen did not give a definition of the technique, "overwriting" appears to be a process whereby the scientist writer prefers the more complex phrase to the simpler, the longer sentence to the shorter, a frequent use of the passive, and long nominal compounds. Although clarity and transparency of writing is largely a subjective notion, as well as a field-specific one, Andersen (1988) suggested that instructors working in these specialized fields need to assist their non-native students in understanding the social and institutional contexts in which this register is used—the aim being to avoid using it "for the display of status" instead of revealing knowledge (p. 157).

Andersen's paper addressed social and stylistic factors in ESP/ESL writing—the practical issues related to success in writing were taken up in more detail by Kroll (1991), who investigated and described the chief components of an ESL course. Her observations included insights into the general concerns of curriculum development, the syllabus design of a writing class, the role of reading, writing assignments and theoretical issues in feedback types, covering the full spectrum of relevant factors. In each of these areas, she looked at what may result in success, and potential pitfalls, for the participants. She concluded that although writing is viewed as a process, it does generate a product whose success is not easy to predict. It hinges, among others, on how skillful a student is in controlling linguistic knowledge and systems, and in addressing a specific audience. Her main recommendation took the long-term view of: what ESL students will be able to achieve in the future.

> Our real goal is to gradually wean our students away from us, providing them with strategies and tools for their continued growth as writers and for the successful fulfillment of future writing tasks they might face once they have completed their writing course with us. (Kroll, 1991, p. 261)

This goal can be achieved with the continued formal and informal development of the training of writing teachers, Kroll added. A source of such training is manifold: it includes gathering reliable information on one's own teaching, observing classes, keeping abreast of research in the field, as well as developing innovations that build group dynamics within a writing course so that the community established there may be transferred to the professional communities where these students will seek audience recognition and response.

1.3.2 Group work

Applying generic group dynamics techniques in a research-component university writing course can take a number of forms. In the L1 environment, Zirinsky (1995) was concerned with how to assist U.S. students in planning, time-tabling, and conducting research that was to be presented in an extended piece of research paper. He reported that fostering collaboration among the students improved the ensuing scripts. Presenting a process syllabus to the students, Zirinsky facilitated this by involving groups of students in each of the main phases: the development of a research question, as opposed to an overall topic; the personalization of the research effort (meaning that students may need to understand how an expert, such as the writing teacher, goes about making a match between an editor's call for papers and the writer's own interest in a related question); the statement of the central thesis of the project; the use of sources of information; and the planning and writing of the report, after which students read each others' papers and critiqued them as well. Zirinsky also made the claim, following Kroll's (1991) view of future writing experience, that such an approach to writing can enable students to develop sustainable skills.

These steps can be taken in both traditional and technology-enhanced programs. For the former, Young-Scholten (1994) and Blue (1988) argued that one potential classroom management innovation was to adopt a writing center context by turning part of the contact hours into individualized reading and writing skills development. This was done by Young-Scholten (1994) in her U.K. and U.S. classes, in each of which upwards of 40 students between the ages of 18 and 60 were enrolled. Blue (1988) found that U.K. ESP student reactions were generally positive when they had an opportunity to participate in frequent one-to-one tuition and that this factor seemed to result in more willingness to rewrite.

For the technology-enhanced application, Sullivan and Pratt (1996) compared a traditional oral and a computer-assisted classroom in which the *Daedalus* software package was used, coming with modules for word processing, topic exploration, messaging, and *InterChange*, a real-time discussion program (for a review on studies in which the same package was used, see Horváth, 1999e). In their analysis of Puerto-Rican intermediate-level ESL students' attitudes, transcripts and tapes of classroom acts, they found that although environment did not affect attitudes to writing in general, there was a significant difference between the two classrooms: there was much less teacher-initiated and controlled discussion, and all students in the computer classroom participated, as opposed to a 50% rate in the oral class. Students involved in peer response groups in the computer class tended to give more specific suggestions (Sullivan & Pratt, 1996, p. 500). Whether less domination by the teacher and more specific comments by student writers resulted in improved writing, however, was not studied.

Caudrey's (1998) and Farrell's (1989) classroom observation projects offered different perspectives on computer-assisted writing programs. Farrell was concerned with the procedures used in a high-school writing center, whereas Caudrey reflected on how the teacher's early interventions in the composition tasks of EFL university students shaped their views on revision. Farrell reported (1989, p. 110) that one advantage of the project was that tutors had the time and experience to observe how student writers were developing their scripts and what types of problems they had. Also, the technological tools appeared to be an additional motivational factor.

Working with a small group of Danish university students in advanced writing courses, Caudrey (1998) introduced the technique of monitoring each student's progress during draft sessions. In the computer lab sessions, he had access to each developing script and so he could intervene when he recognized an organizational issue that needed prompt action. He hypothesized that the time teachers spend on providing written feedback could be minimized if they could observe how a script was being developed. Although no concrete qualitative or comparative analysis was done, Caudrey reported that some students were satisfied with the teacher's on-line assistance. One participant reported that the technique was "very good" as it allowed for revision during the composing process. Yet there were also problems. A student would have preferred to have dictionaries while writing, with another one complaining that the lab was too noisy and thus distracting. There was no information available on whether students could voluntarily sign up for this course. A drawback of the approach may be that students can experience even more serious writer's block if they know that someone is watching their work at the keyboard. Caudrey provided a brief statistical overview of the effect of the approach by comparing the marks five raters gave on three types of script:

- produced as a single draft;
- written in a draft—traditional feedback—revised version system;
- developed in the lab environment.

Caudrey reported a small increase in the marks for scripts produced in the lab, the mean grade for scripts written by the eleven students being 8.32 on a 13-point scale, as opposed to 7.54 in the single draft and 7.96 in the traditional revision class. Further research is certainly needed to validate, on a larger population of students, the efficacy and potential drawbacks of the approach.

1.3.3 The *Baseline Study*

So far, we have seen a number of approaches to writing pedagogy in the classroom. Empirical research has studied the factors that contribute to success in writing in ESP, ESL, and EFL. Now I will turn to a recent Hungarian study that reported on task and text types currently used in secondary EFL.

The cross-sectional baseline study (Fekete, Major & Nikolov, 1999) was conducted primarily to assess the language teaching and testing situation in the country's secondary schools. As far as writing instruction issues are concerned, a classroom observation project by Nikolov (1999) investigated the current practices of EFL teachers in incorporating writing tasks. Although the study established that there were a few schools that were good examples of effective teaching, the overall results are far from reassuring. The situation was not positive in the writing related section of the "Classroom Observation Project," either. The most frequent writing tasks observed in the 118 classes in years 10, 11, and 12 were based on Hungarian school-leaving exam test techniques, such as translation and gap-filling. This finding lends some support to the claim (Nikolov, 1999, p. 233) that examination techniques exercise a washback effect on what is going on in the classroom: if exams incorporate translation and gap-filling, teachers will tend to favor these types of tasks in their classes, too.

When looking closely at the table that listed the writing tasks observed across the three years (Nikolov, 1999, p. 235) one can find another somewhat worrying trend: most non-translation task types applied were meant to elicit students' manipulation of texts given. These tasks included copying, filling in data, arranging words into sentences and sentences into paragraphs (with paragraphs arranged into larger passages observed once). While such tasks can complement and sometimes improve grammar and organization skills, on their own they can hardly result in the development of a writing attitude needed for improving mastery over the language in the written mode of expression.

The study did not aim to gather information on how written tasks were developed as part of a syllabus, what the role of groups of students was in the various stages of the writing process, or how students received various types of feedback on their writing. However, the task and text type distribution information, coupled with the results on classroom management, on language use, and on the other skill areas, indicated that writing pedagogy was not a high priority in these classes and that the traditional grammar-translation method impacted this skill's treatment in the classes observed. As the sample of schools was not representative, however, we need further studies that could aim to investigate, on a representative national or regional sample, the procedures, performances, and syllabuses as related to each of the four skills.

1.4 Revision: Shaping text by writer and reader

Much interest in the 90s was directed to the empirical analysis of what goes on in the intricate interplay between how the student writer construes of theme, organization, and audience and how the teacher reader reassesses these notions. This field represents an exciting area of classroom practice and research, one that will probably continue to shape the way new generations of writers and readers approach the tasks of writing. In the process orientation tradition, as we have seen, revision may appear as an add-on after a sequence of clearly defined (and often, pre-defined) steps. Much as that approach may prove useful for a variety of student and teacher styles and preferences, such an isolationist approach to the need and nature of revising has its limitations. In this section, an introduction of a series of studies and of a recent text that focuses on revision will further clarify the concepts that are at play.

1.4.1 Revision for grammatical accuracy

In an early study, Frodesen (1991) reflected on the different views process- and product-oriented writing instruction had on the role of grammatical accuracy in ESL composition. Calling attention to variable learner attitudes and contexts, he suggested that for a writing program to be optimally successful, teachers need to help students develop such accuracy with only minimal terminology, and then went on to present four main groups of activities that aim to assist learners in building revision skills for grammatical accuracy. The system of these groups is summarized in Table 1.

Table 1: Activities for grammatical accuracy (based on Frodesen, 1991, pp. 266-275)

Main group of activities	Main goal	Related tasks and techniques
Text analysis	Develop writing and reading	Distinguishing between clause types, Selecting texts for studying article use, Summarizing findings
Guided writing practice	Solve problems diagnosed in individual learners	Dictation, Text elicitation, Text conversion, Text completion
Editing	Develop awareness of errors	Error detection, Correction, Read-aloud technique, Algorithms
Teacher correction and feedback on errors	Identify patterns of errors	Keeping error logs, Conferences

In this model, revision is seen primarily as a means that arises from a need to eliminate error: the main goal is to assist the learner and groups of learners to polish text so that their awareness may later be used in the pre-composing stages of writing. Frodesen concludes by stating that in "selecting and developing grammar-oriented activities for the classroom, the teacher should always bear in mind the students' needs and background as well as the demands of the writing tasks" (1991, p. 275). The task, in this interpretation, is obviously the end product: the error-free composition. However, task can be interpreted such that students focus directly on revising, with or without attention to grammatical accuracy. For such a definitely more innovative approach, we now turn to Lane's (1993) text.

1.4.2 Revision for text creation

After the End (Lane, 1993) took the concept of communicating with the developing writer to a different pane. The key word is *creative*. Lane aimed to inculcate in his readers (U.S. native speakers of English) the daily experience that they are creating when they are writing, and that they are doing so especially in revising what they are writing. The primary goal the collection of revision techniques communicates is simple and relevant: to share with the reader the discovery that when we revise, we can see better and that this realization is the source of much personal and collective benefit.

As a writer, Lane approached the theme with a revolutionary idea: when we write the imaginary "The End" of any text, it really is just the beginning. Implicitly arguing against the lock-step fashion which sees revision as one of seven rigid steps in a rigid process, the author demystified the act of revising and turned it into a flexible route to achieving goals. In particular, he offered the following suggestion, contrasting tradition and innovation (Lane, 1993, p. 3):

Traditional stages in writing	Lane's suggestion
1. Brainstorm	1. Revise
2. Map	2. Revise
3. Freewrite	3. Revise
4. Draft	4. Revise
5. Revise	5. Revise
6. Clarify	6. Revise
7. Edit	7. Revise

Clearly, at each major theoretical juncture of writing, revising takes place. In brainstorming, the monitor may already revise what gets elicited. When a theme is mapped out, we may cross branches out and insert new ones. This, of course, may well lead to brainstorming new ideas that may not need mapping, leading directly to editing, and so forth.

In the main text of this self-help resource, Lane then structured the techniques around key processes in becoming flexible writers. Operating with a set of no-nonsense and concrete terms, he defined and exemplified revision micro-strategies that language teachers can (and some do) use in their classrooms. Among the most innovative such terms and techniques are the following:

Snapshots and thoughtshots: In explaining how an activity may be based on this idea, Lane shared this tip with the teacher:

> Begin by explaining to students that writers have a magic camera that they can point at the world and create snapshots that contain smells and sounds as well as colors and light. (1993, p. 35)

This metaphor of capturing specific detail as if by camera is then used for an activity that puts the learner behind the camera as well as enabling them to revise so that they include specific, rather than generic, information in their description of a person.

The revision Lane argued for does not give priority to grammar; it is much rather an attitudinal shift that the teacher can foster in becoming part of the revising effort, not just in the assigning and correcting stages. It is no wonder, then, that several of the activities are non-directive and developmental in that the steps described do not get prescribed.

As for the specific language and behavior outcomes of such an attitude to revision, the recurring theme is this: a reader, who happens to be a teacher, has to have a voice, a distinct characteristic. Writers and students are not different. Throughout the hundreds of techniques, Lane works on this quality to surface in the writing class and eventually in the text. In addressing the teacher reader, he explains:

> That's what makes me smile—seeing a kid's voice leap off the page, speaking to you directly like some hotline to the soul. It was also a quality in writing that was hard to break down and teach. If it was there, great. There's a writer. (Lane, 1993, p. 158)

Lane nurtures this voice, this individuality in descriptive personal essay writing by fostering students' choice of theme and approach, by bringing them in close contact with their own audiences, and by exposing them to what he calls "boring, voiceless" (p. 164) research papers that students can revitalize.

This resource collection goes a long way toward enabling creative revision in the language classroom by helping students and teachers experiment. A contrast to the discrete grammar-focused revision approach, this latter aims to be holistic and thematic. In between these two extremes, there have been a number of other directions summarized in Grabe and Kaplan (1996). In the following, I will present the findings of their own research into revising and show what evidence empirical research has gathered on the benefits of different types of revision in different communities. As the specific details of a related issue, responding to writing, will feature in the next section, here I will focus on what these authors noted about revision processes as observed in peer-response and peer-revision settings.

1.4.3 Empirical studies

Realizing the impact that the nature, quality and quantity of response has on students' writing attitudes, Grabe and Kaplan (1996) proposed that the positive motivation that this process carries is a significant factor in shaping learner behavior. However, the research evidence and the various guidelines worked out in individual projects do not allow for generalizations. What seems to hold true, though, is that response from peers not only complements other forms and manners of revision strategies, but can determine, on its own right, their success if conditions are optimal. Studies showed that by promoting collaboration, students "develop a sense of community" and they benefit from being exposed to "a variety of writing styles" (Grabe & Kaplan, 1996, p. 386). Seen in this context, classroom writing, although in some phases by definition a private, intimate undertaking, will approximate authentic settings whereby audiences and writers interact through the medium of publications, genres, text types, and editorial preferences.

Summarizing recent research in the field of peer response to learner writing, Grabe and Kaplan extracted four factors that seem to contribute to the effectiveness of the approach (1996, p. 387). The first is the individual's conviction that response from one's peers will be beneficial. This seems to be an area where the teacher's role is paramount: helping to create the conditions for a group to act as a group is a pedagogical responsibility (for group dynamics, see also Dörnyei & Malderez, 1997). The second factor influencing effectiveness is the formal training students receive in peer response and revision. Although some teachers were shown to oppose structured and formalized guidelines for their students in such programs, students preferred when the writing teacher helped them define the rules. Listed as the third one (but probably coming first for most L1 and L2 writers chronologically) is the awareness of goals students have in asking for and providing a response. The fourth factor refers to the requirements that once such practice is begun, participants are held responsible for their involvement.

The second of these four factors, training, was shown as a significant variable in the study conducted by Villamil and de Guerro (1998). In the first project that investigated how the variable of rhetoric mode influenced peer revision, they studied the revision activities of fourteen Puerto Rican university ESL students. By systematically gathering audiotaped sessions, script samples of first and finished drafts, the researchers found that after receiving explicit training on the terminology and principles of narrative and expository writing, the majority (74 percent) of the revisions suggested by peers were incorporated in the writing process. They reported that narrative scripts were longer than the expository ones, and this trend continued to hold for each revision, further evidence of the hypothesis that it is more difficult cognitively to process a persuasive writing task (Villamil & de Guerro, 1998, p. 509).

In the analysis of the scripts and their revisions, five criteria were applied by two external raters for both types of writing: content, organization, grammar, vocabulary and mechanics, qualities that are most common in assessment. Most revisions in both types of script were grammar-based, followed by content in the narrative and vocabulary in the persuasive scripts—however, the difference in the ranking or the weight of these revisions was statistically insignificant. The authors, one of whom was the teacher of the students participating in the study, suggested that the revision experience will be beneficial for students later when they need strategic competence for text revision. Although this claim was not validated by follow-up studies or by interviews with students, the study succeeded in focusing attention away from error analysis to revision analysis based on what students discussed and what changes they incorporated in their drafts.

Incorporating major structural changes in a text was found least likely to occur: for both types of script, this was the least frequent revision change. This is not surprising: organizing ideas, arguments and topics within the development of these arguments is among the most demanding processes for professional and amateur writers. However, future research is needed to investigate how the writing classroom can address these issues at various levels of development, in L1s and L2s.

Already, however, evidence suggests that not all students are willing to act on the suggestions by their peers. For example, a study conducted by Sengupta (1998) revealed that among a class of Cantonese EFL students there was a marked reluctance to carrying out peer evaluation. Students saw the job of commenting on their scripts to belong primarily to the teacher, and for these participants the reader who counted was the expert instructor. The finding in Huang (1995) corroborates this result: in the pilot study, 22 Chinese university students of writing were assigned to English and Chinese discussion groups and reported little enthusiasm about providing feedback to motivate revision in a two-draft writing task. Huang hypothesized that for such group involvement to promote peer revision a longer experience may be necessary.

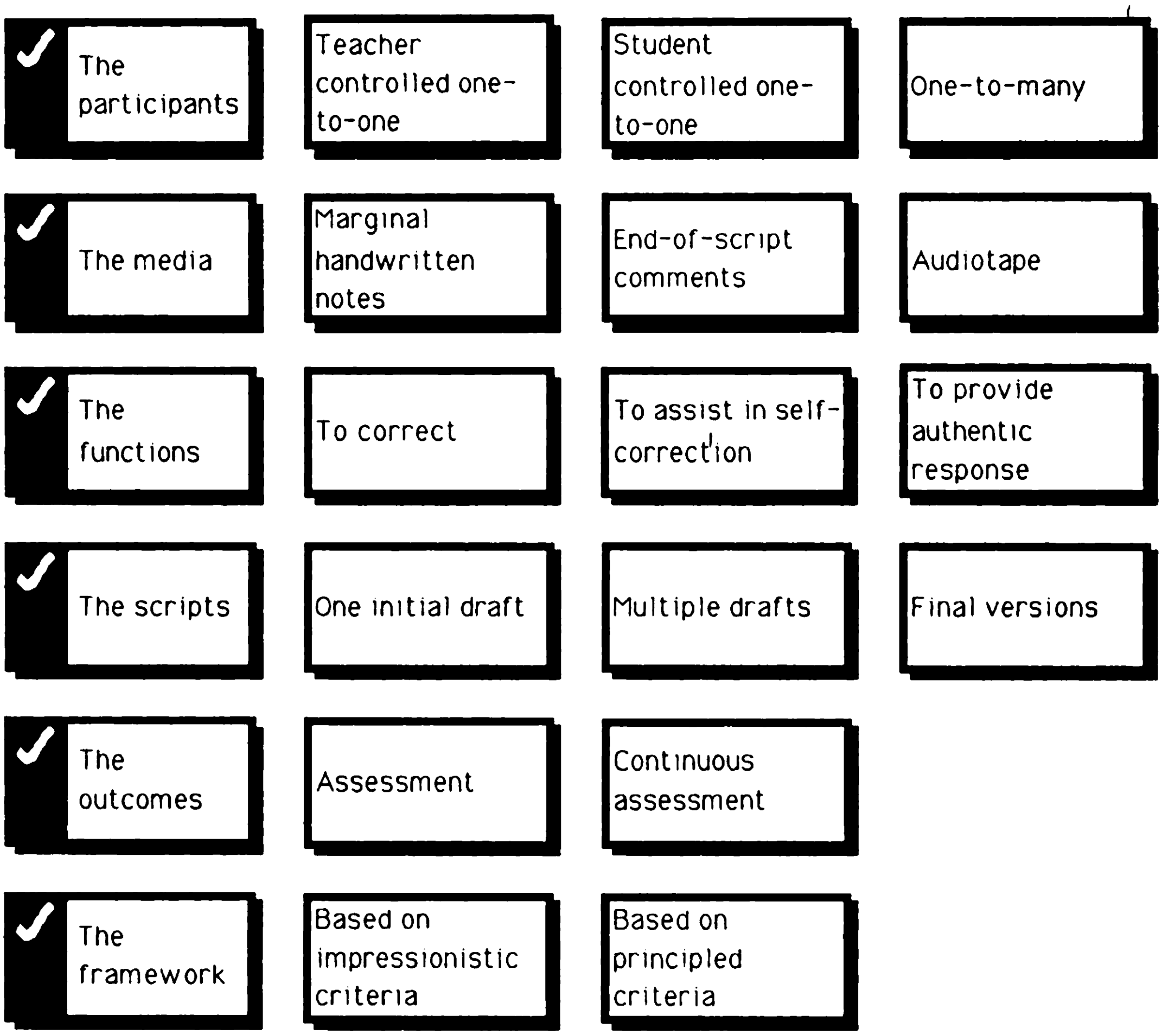
In the ESL context, a slightly different result was obtained in Mangelsdorf's study (1992). Among a culturally heterogeneous mix of university students in Arizona, it revealed that often peers were unable to provide the type of feedback that would be helpful for them to draft a script. However, a positive element of the process, according to the interviewees, was that "peer reviews led [the students] to consider different ideas about their topics and helped them to develop and clarify these ideas" (Mangelsdorf, 1992, p. 278). Once the improvement in writing quality became obvious, participating students were more willing to share and act on suggestions in their revisions.

As for the EFL view, an Asian study aimed to establish correlation between holistic rating of EFL college writing quality and quantity of revision (Sato, 1990). It investigated Japanese students' success in a picture description task. Of the ninety participants, three levels of writers were identified. The study reported that although no significant differences could be established in various syntactic levels, the two top groups made significantly more successful revisions in their final drafts. The paper suggested (Sato, 1990, p. 157) that further research was needed to study the relationship between different tasks and levels of achievement, and that including variables of proficiency in the target language and of writing expertise would enhance the validity of findings.

1.5 Responding to writing

1.5.1 Main variables

With so much effort going into developing writing courses, materials and procedures, one may be tempted to suppose that responding to an early draft or a final version of a script should pose no problem for the teacher. Giving feedback on writing, however, is not a trouble-free spot in writing pedagogy: numerous studies, and several attempts to grasp the underlying theory, have only come up with more questions. Apparently, the amount and type of feedback, the timing, the mode, the provider, and the subsequent application of it continues to pose research design and pedagogical problems. This section aims to review what is known today about these factors, beginning with the interpretation of the overall purpose of response and the problems that have been reported, tracking down its many suggested forms and contents, pursuing the issue to how feedback by the writing teacher is interpreted and incorporated into subsequent writing. Figure 1 charts the main variables.



✓ The participants	Teacher controlled one-to-one	Student controlled one-to-one	One-to-many
✓ The media	Marginal handwritten notes	End-of-script comments	Audiotape
✓ The functions	To correct	To assist in self-correction	To provide authentic response
✓ The scripts	One initial draft	Multiple drafts	Final versions
✓ The outcomes	Assessment	Continuous assessment	
✓ The framework	Based on impressionistic criteria	Based on principled criteria	

Figure 1: The variables of response to writing

Feedback is an integral part of any pedagogy. It aims to engage participants in authentic communication about the subject of tuition, and about its goals by signaling transitions in the process of learning. As such, feedback also forms part of assessment and evaluation, both continuous and task-specific. Some educationalists view feedback as having the function to correct: to put things right when they go wrong. Another view of feedback is that it should inform the learner of the progress made, and thus correcting syntax and organization errors is valid. Bartram and Walton (1991) proposed that although written production is primarily an individual activity, much can be achieved in applying tasks involving pairs and groups of students. In terms of providing teacher feedback on student writing errors, they suggested that the "red-pen syndrome should be avoided" (Bartram & Walton, 1991, p. 78) and instead listed a number of areas and techniques with which to facilitate accuracy and composition improvement. These included the need to react to content, to restrict correction to specific morpho-syntactic units, to involve students in correction, and to reformulate. They emphasized, however, the importance of communication between teachers and students not only after a script is written but also before and during that stage.

1.5.2 Positive effect of feedback

Fathman and Whalley (1990) conducted qualitative research among 72 U.S. students of college ESL into the effect of teacher feedback on grammar and composition. They found that such feedback resulted in improvement: it helped students identify and correct their own errors. Another result questioned the general validity of Zamel's (1985) claim that teachers' comments were often too vague to act upon: the "general comments giving encouragement and suggesting revisions" (Fathman & Whalley, 1990. p. 186) were reported as factors that contributed to the development of rewritten versions of students' scripts. While this appears to have been true of writing improvement in the short run, Zamel's (1985) observation that there is still little evidence that such improvement is long lasting was not refuted. Specifically, she claimed that teachers' comments often lacked consistency and relevance from the point of view of subsequent revision: they tended to highlight each and every grammatical error, favoring correct yet non-communicative prose while almost totally ignoring the content of the scripts.

To collect information on student attitudes to and preferences for receiving feedback, Dheram (1995) conducted a case study among five EFL students in Britain. She investigated whether students preferred comments on grammar or content, how they responded to feedback, and what the preferences meant for future writing instruction. Besides analyzing questionnaire and interview data, Dheram reviewed pre-feedback and post-feedback drafts and found that students became aware of the importance of revision as part of discovering meaning. Perhaps the most relevant finding was that content should enjoy top priority in teachers' response.

When a process approach is adopted, it is crucial that students are helped in the development of their scripts at every stage. To add further assistance, Frankenberg-Garcia (1999) put forth the innovative suggestion that feedback could be given even before a text is produced: at the initial stages of the development of ideas for a composition. This view reverberates the procedure whereby a classroom is seen as a workshop, with part of the time turned into intervening in the writing process. Frankenberg-Garcia pointed out that text-based feedback had serious limitations because the type of feedback students need most cannot be adequately given without having hard evidence of the types of decisions (good and bad) that students typically make when composing. The written text may be polished with little need for grammatical or compositional change, yet it may not reflect writer intention if the student had major difficulty with an idea, grammatical unit or vocabulary item and decided to apply an avoidance strategy, thus fossilizing a problem. To deal with the actual composing process, then, she argued, we need to gather information on the specific needs students have and incorporate that information in the verbal or written feedback that is given on the processes, rather than a draft. She emphasized that this approach was not intended to replace text-based writing feedback—rather, to complement it.

The form and content that teachers' feedback may take continue to challenge practitioners in the field. The ideas suggested by Cohen and Cavalcanti (1990), Chen (1997), Grundy and Li (1998), Leki (1990), Mosher (1998), Myers (1997), Allwright (1988) and Schultz (1994) represent some of the potentially most valuable recommendations. Here I will briefly introduce claims about what should be favored and avoided in feedback.

Myers (1997) gave a detailed account of her writing course for ESL students in which she adopted the technique of sentence reformulation. Using simple codes, she returned papers that students were requested to amend by incorporating the revisions she had made. By doing this, students prepared a clean copy with no grammatical inaccuracies so that they could focus in class on the content of their peers' writing, thus participating in a program that relied heavily on teacher direction in terms of language correction, but which eventually enabled students to exercise the role of peer editors of ideas in the sessions.

1.5.3 Student agendas

Work by Grundy and Li (1998) also pointed in the direction of allowing students to take more responsibility for their own writing, but their approach was more radical. Viewing correction in writing pedagogy as a function that has little validity, they proposed that we are witnessing a "you write—I correct" syndrome. Identifying the problem as a logical result of product-orientation, they aimed to attack this unsatisfactory situation by alternative techniques of response. These include Post-It notes by teachers that students respond to before revising, conferencing, checklists for revising that are complemented by the students' own concerns, learner logs, portfolios, and audio-taped responses. This last technique involved the teacher recording a corrected version of a student's essay. Grundy and Li (1998) claimed that not only does this technique facilitate quick response, but it also involves students in an authentic listening activity. When we consider Zinsser's (1998) comment that professional writers write for the ear, not only the eye and mind, we may find this technique truly authentic: it could result, in the long run, in raising awareness of what is commonly termed as "what sounds good."

1.5.4 Responding to feedback

Obviously, the practicality of any feedback type will depend on a number of variables: educational context, type of syllabus, length of assignment, number of students, and, maybe most importantly, what Leki (1990) called the "persona" of the writing teacher (p. 59). Leki conceptualized the teacher as having a set of three divergent functions in responding: the real reader self, the teacher as the coach, and the evaluating teacher. As these functions may conflict, and because the writing teacher will eventually need to evaluate how content is presented in a number (and often, a high number) of scripts, Leki claimed that the writing teacher may become schizophrenic, juggling these roles. To help maintain a pedagogically sound balance, she recommended the following directions for feedback.

First of all, applying a multiple-draft composition syllabus ensures that assignments are integrated so that feedback on each draft may be usefully incorporated by students. This also has the advantage that the teacher may intervene in the writing process when it is most needed. Second, when assignments form a well planned project, the writing course will facilitate long-term development, with teacher comments applied in subsequent tasks as well. Third, students can be given a set of questions that elicit information on what they, the primary stakeholders, consider the best features of their writing. This may not appear to be a teacher's feedback at first sight: after all, the teacher provides the questions, and the students reflect on them. However, by identifying what is valuable for them, these students enable the teacher to better focus on those elements of writing, thus bridging the gap between writer's intention and reader's interpretation, a significant benefit considering that student writers do not always have the skills to communicate their goals fully.

As the final issue in teacher response to student writing, we need to consider the effect it has on students' perceptions and its implications: how students respond to response. Primarily interested in the meta-cognitive processes activated by expert feedback, Hedgcock and Lefkowitz (1996) collected and analyzed data from ESL and EFL students. They established that for feedback to be used effectively, students must be engaged in the process. They obtained evidence that FL learners were in favor of feedback that helped them formulate the content and structure of their scripts. Rather surprisingly, the majority (82%) of the students preferred "red-pen" corrections, apparently because this resulted in most short-term improvement in surface-level features, with FL students being of the opinion that "form should precede, and have priority over, expression of meaning, concepts, or original ideas" (Hedgcock & Lefkowitz, 1996, p. 297). What they did not intend to measure, however, was how teacher feedback was attended to in revised texts. Nevertheless, the study can be regarded as a welcome example of applying quantitative and qualitative research methods.

To highlight an additional implication of feedback practice, I showed that teachers' comments themselves may serve as resources for teaching and exploration for students (Horváth, 1997b, 1998d). English major FL students at Janus Pannonius University were given samples of teachers' comments on timed essay tests and asked to read, review, and reflect on them. This was done so that they could familiarize with the discourse the raters of the essays produced and it broadened students' understanding of the areas that the comments elaborated on, especially noting what the teachers marked as positive features of the scripts.

1.6 Concluding remarks

This chapter has addressed issues in writing theory and pedagogy as proposed by leading practitioners of the field. The interest in writing pedagogy has continued to challenge empirical research, with concerns about course goals, task types, classroom procedures and revision techniques receiving a fair amount of attention. There seems to be a tendency to consolidate the results by turning to the development of an ethnography of writing that can explain how instructed writing development takes place as writers interact with themes, expert teachers and peers.

The theory of L2 writing has been informed by L1 writing theory in the contrastive rhetorical tradition, establishing the need for verifiable research into language varieties. The communicative approach to language teaching coincided with the move away from the sentence-level concern with grammar and the focus on product, in the process orientation. The development of L2 writing theory and pedagogy has been motivated by the practice of task-based learning, in both traditional and online contexts.

As the chapter has shown, the majority of L2 writing development studies were conducted by native speakers of the target language, raising questions about the validity of some of the claims made about innovation in the writing curriculum and syllabus when such endeavors do not tackle educational and cultural differences.

However, writing research has become a major component of recent applied linguistics studies. For continued progress, the field could benefit from cross-institutional and cross-cultural projects, as well as from combining insights gained by writing pedagogy with corpus linguistic data, so that the ethnography of writing can be supplemented by reliable data on student performance. To address the theoretical and practical implications of this endeavor, we will now turn to surveying the literature of the discipline of corpus linguistics, a sub-field of which is the development and exploitation of corpora of learners' written performance.

Chapter 2

ISSUES IN CORPUS LINGUISTICS: A REVIEW OF THE LITERATURE

> *Language corpora are becoming available cheaply, sometimes free. The likely impact on language teaching will be profound—indeed the whole shape of linguistics may alter at speed.* (Sinclair, 1997, p. 38)

Introduction

The previous chapter has reviewed the current theoretical and practical concerns of writing pedagogy. I have made the claim that besides an ethnographic description of processes and products of writing and writing pedagogy, we also need evidence from a larger set of language sample that FL students produce. That claim will be refined in this chapter, which aims to present the case for the need of corpus analytic methods in descriptive and applied linguistics.

One of the leading figures in corpus linguistics applying machine-readable collections, Leech (1997a), defined a corpus as "a body of language material which exists in electronic form, and which may be processed by computer for various purposes such as linguistic research and language engineering" (p. 1). The theoretical underpinnings, the technical development, and the study of such corpora have gained considerable ground in the past decades, signaling a trend away from decontextualized linguistics toward a study of language that takes account of context based on what is often referred to as "real" language. This chapter will review the growing literature that has given evidence of this enterprise.

The chapter is divided into six sections. The first will offer a discussion of the theoretical issue of performance versus competence, focusing on the contrasting views of Chomskyan generative linguistics and corpus-based linguistic analysis (2.1). Section 2.2 will be based on a brief historical overview of major corpora as it clarifies the types that have been established recently. The next section (2.3) identifies the issues of representative design, and some technical details of corpus development. Section 2.4 further narrows the scope by identifying a link between computer-assisted language learning and data-driven learning. Section 2.5 reviews work in the field of learner corpus linguistics, centering on the International Corpus of Learner English project.

Finally, I will identify the benefits of applying corpora in language studies in section 2.6.

The concepts, definitions, and processes reviewed in this chapter will be central to the presentation of writing pedagogy at Janus Pannonius University and to the description and analysis of the JPU Corpus.

2.1 Rationale for corpus linguistics

2.1.1 Data in language analysis

A crucial issue in any analysis of language is the role of data. Evidence sought to support a theory of structure or language use provides the basis on which to evaluate the feasibility and applicability of a hypothesis. The role of linguistic evidence also has practical implications in language education, as it impacts on the manner in which a syllabus is presented (Seliger& Shohamy, 1989). As the examples may be either intuitive (coming from the linguist's own repertoire) or observed (recorded in some psycholinguistic elicitation or field work), the issues of competence and performance present the framework in which this question has been studied.

In the field, two competing traditions have emerged: 'asocial' linguistics that incorporates intuition to capture generic features and universals of language and of particular languages, and 'social' linguistics that investigates generic and language-specific notions based on observations of utterances (Wardhaugh, 1995, pp. 10-12). In the former paradigm, linguistic inquiry springs from a need to establish sufficient elements that can adequately describe the grammar of language; the latter engages the actual language community (or population) and extracts from it a corpus that is then used to test hypotheses. This section will offer a brief evaluation of these two traditions.

The most influential theoretical linguist of the twentieth century is Noam Chomsky, whose generative grammar is embedded in the asocial tradition. In defining linguistics as the study of grammar, he developed a set of strict principles and operators that language employs in generating all possible and grammatical utterances (Chomsky, 1957; 1965). The main focus, then, is on what is possible. This represents one of the main differences between the asocial and the social paradigms. In socially embedded linguistics, it is not only what is possible that is studied, but also what is probable. According to Kennedy (1998) this does not mean that linguistic theory does not benefit or is not "compatible with" (p. 8) the study of a corpus. On the contrary; as science requires evidence with which to refute or support a hypothesis, corpus linguistics provides a rich set of such evidence that allows for generalization.

2.1.2 Competence vs. performance

Traditional generative linguistics is concerned with the competence of an idealized native speaker whose sociolinguistic status, age, and gender are viewed as immaterial to the study of the generation of grammatical utterances. By contrast, empirical linguistics, of which corpus linguistics is a representative, sets itself the agenda of investigating the variables that lead to differential performances across these spectra. It interprets competence as "tacit, internalized knowledge of a language" (McEnery & Wilson, 1996). The generative linguist, who is concerned with capturing linguistic competence, applies a corpus of internal, closed sets of examples derived through introspection (a process that, according to Labov, 1996, might introduce error into the description of linguistic phenomena). The corpus linguist's data set derives from an external, open body of actual language performance, or the actual, social and contextualized application of competence. These performances are recorded following strict rules, with the necessary and available biographical and sociolinguistic information tagged to it (Stubbs, 1996). As corpus linguistics opens up the database upon which description and analysis is based, the evidence becomes available for further verification, too, representing another advantage (McEnery & Wilson, 1996, p. 13).

As Fillmore (1992) noted, the two types of linguist should ideally "exist in the same body" (p. 35). Contrasting the images and concerns of whom he called an "armchair linguist" and a corpus linguist, Fillmore pointed out that no corpora will ever offer all the evidence linguistics needs, but also that corpora have allowed linguistic scholarship to establish new facts about language, facts that one "couldn't imagine finding out about in any other way" (Fillmore, 1992, p. 35). But he also called attention to the importance of introspection and analysis by a native-speaker linguist. Biber (1996) also suggested that both generative linguistics and variation studies looking at linguistic performance derived from corresponding aspects of linguistic competence represent valid positions.

The call for a combination of the two approaches is based on the assumption that native speakers are competent decision-makers on issues of syntax. While the claim may be a perfectly valid one, I would like to raise an issue related to the theoretical limitations of the basis of linguistic inquiry. As no corpora can ever fully represent the language performance of a community (see, for example, Partington, 1996, p. 146), so, too, are introspective linguists limited in their competence (Labov, 1996). This adds further support to the claim that theoretical linguistics and corpus linguistics can and should co-exist.

Such co-existence occurs in a social context. The notion of context (or setting) in which language competences materialize (Hymes, 1974) as well as its central importance, was further highlighted by Sinclair (1991), who claimed that as introspective linguists do not, as a rule, require a discourse context for their own examples, the naturalness of the evidence suffers. Defining this feature of an utterance as a choice of language that is appropriate to the context, Sinclair observed that because of the difficulty of simulating context, examples are often unlikely "ever to occur in speech or writing" (1991, p. 6). This is why, he went on to argue, linguistics should be careful not to misrepresent what it aims to describe. In other words, what may be authentic (in that system, possible) to the individual linguist in a particular context for supporting a particular claim may not be authentic (in that system, probable) to the language community.

2.1.3 Lexicography and language education

So far we have seen contrasting views on the primacy of theory and of evidence, the nature of evidence, and the issue of authentic context. Moving on to the rationale of corpus linguistics in the field of lexicography and language education, we need to address the interface between a linguistic enterprise and its pedagogical application. Traditionally, dictionaries were compiled mostly via introspective techniques, with individual lexicographers aiming to compile sets of data that described a limited array of items and meanings. By contrast, corpus linguistics views the generation of meaning as a process in which syntax and semantics are not isolated but interfaced. By relying on a growing body of evidence (Bullon, 1997; Sinclair, 1991; Stubbs, 1995; Summers 1998), lexicography driven by corpus linguistics establishes this relationship and provides useful help for distinguishing between discrete meanings. However, even corpus linguistics does not, normally, need to rule out intuition. As Summers (1996) pointed out, lexicographical studies and dictionary entry frames need corpora to determine, for example, the frequency of individual units in a large general corpus, but linguistic intuition is necessary in the ordering.

In terms of language education, ~~corpus~~ linguistics has helped direct attention to what constitutes authenticity of material, learning experience and classroom language, key factors determining the relevance of learning especially in the communicative language teaching tradition. A direct result of the approach is what data-driven learning and the development of learner corpora have achieved (discussed in detail in 2.4 and 2.5). One of the proponents of this approach, Johns (1991a), posited that learning, especially on advanced levels, can greatly benefit from assisted and direct manipulation of corpus data. He argued against the stance held by such figures of applied linguistics as Widdowson (1979; 1991) who placed the emphasis not on authenticity of material but of learning experience, arguing for the use of simplified texts to help ensure authenticity and comprehensibility at the same time for the learner. As a consequence, he cast doubt on the relevance of corpus findings to the process of teaching and learning foreign languages (Widdowson, 1991). Calling attention to the principle of pedagogic relevance, Widdowson made the following point:

> Language prescription for the inducement of learning cannot be based on a database. They cannot be modelled on the description of externalised language, the frequency profiles of text analysis. Such analysis provides us with facts...but they do not of themselves carry any guarantee of pedagogic relevance. (1991, pp. 20-21)

As opposed to Widdowson, Johns (1991a) argued that authentic and unmodified language samples were essential in language learning. Widdowson (1979, 1991) focused on the learners' need to exploit materials that represent authenticity of purpose *and* were within their grasp. In Johns's argument, the requirement of no modification is central. For learning material to represent full authenticity, the original purpose and audience should not be altered. Schmied (1996) took a stance whereby the corpus can be instrumental with pedagogical relevance still maintained. In his view, examples and materials derived, and, as need made this necessary, modified from a corpus still had applicability: Adaptation is possible to various learner development levels, but the example used to illustrate a language pattern may be valid if it comes from a corpus (Schmied, 1996, p. 193).

Taking a position similar to that expressed by Widdowson (1991), Owen (1996) criticized the application of corpus evidence in language education when it negated the appropriateness of intuition. Describing the problem of an advanced FL student who was primarily interested in receiving prescription, rather than description, Owen argued that teachers' experience with language and roles as standard-setters should not be ignored. He went on to claim that teachers can hardly clarify usage problems for their students based entirely on consulting a corpus. In fact, he suggested,

> the tension between description and prescription is not automatically relieved by reference to a corpus. Intuitive prescription is fundamental to the psychology of language teaching and learning....Even if teachers had the time to check every prescription they want to make, the corpus would not relieve them of the burden of using their intuition. (Owen, 1996, p. 224)

This evaluation of a practical concern is in line with what other experts, such as Fillmore (1992) and Summers (1996), claimed. Biber (1996) summed up the advantages of text-based linguistic study. He identified four features that make the corpus linguistic endeavor particularly relevant. These were the following:

- ➤ their empirical nature allows the analysis of naturally occurring texts;
- ➤ the texts are assembled on a principled basis;

- automatic and interactive computer techniques can be applied;
- they can inform both quantitative and qualitative research.

The major proposition of corpus linguistics is that real examples can better support hypotheses about language than invented ones. A number of experts have made the claim (Aston, 1995, 1997; Berry, 1991; Bullon, 1988; Hoey, 1998; Sinclair, 1987a). McEnery and Wilson (1996) also underscored the importance of the synthesis of qualitative and quantitative language study. In fact, according to them, the recent increase in the study of corpora, a process they call a revival (p. 16), has been due to the realization that one needs to "redress the balance between the use of artificial data and the use of naturally occurring data" (p. 16). How this revival has been made possible by the development of influential corpora will be the subject of the next section.

2.2 Corpora: History and typology

The rationale of corpus linguistics is to directly access, derive, and manipulate evidence from a collection of texts. Such collections may be static or dynamic, depending on the media in which they are stored. The distribution of static and dynamic corpora can also be viewed from the point of view of content and representativeness. In this section, I will provide an overview of these two types, charting the development, function, and applications of pre-electronic and electronic corpora, and providing a typology of these based on Kennedy (1998) McEnery and Wilson (1996), Sinclair (1991), and Greenbaum (1996a, 1996b) as the main sources.

2.2.1 Early corpus linguistics

A static corpus is any naturally occurring and recorded sample of language use: the language has a non-metalanguage purpose to achieve. That is to say, the text's primary aim is to communicate. From this definition it follows that a corpus does not necessarily have to be stored in a digital format. In fact, for centuries, and especially in biblical studies, corpora were exclusively analog.

According to Kennedy (1998), there were five main applications of these pre-electronic corpora:

- biblical and literary studies from the 18th century, based on manually produced concordances of content words;
- lexicographic investigations to provide literary examples for dictionaries such as the *Dictionary of the English Language* and the *Oxford English Dictionary*;

- dialect studies in the 19th century to describe lexical variation;
- foreign language education innovations such as the work of Thorndike in the 1920s;
- grammatical inquiries, such as the one by Fries in the U.S., and more recently Quirk's Survey of English Usage (SEU) Corpus.

The size and the systematic composition of the SEU Corpus already pointed in the direction of electronic corpora, and in fact part of it was later digitized to allow for technologically and linguistically more advanced searches and applications. The spoken samples of the SEU Corpus were to be transferred to electronic media in the 70s, forming the basis of what became known as the London-Lund Corpus (LLC, discussed in more detail later), an initiative of Svartvik.

The development of dynamic digital corpora had its theoretical and experiential foundations in the pre-electronic projects, together with a growing awareness of the need to accumulate larger collections that can be captured and stored on computer to facilitate faster access, more refined analyses, and thus more reliable and valid information drawn from these studies. With the simultaneous advance that information technology made, this was a time of convergence of linguistic interest and technological potential.

2.2.2 The Brown Corpus

In 1961, the first electronic (machine-readable) corpus was being planned by Francis and Kucera. It was to comprise one million words of English text, arranged in two major subcorpora: informative (non-fiction) and imaginative (fiction) texts. The former set contained the majority of the 500 samples: 374 texts, with the latter accounting for the rest (126). Taken together, they were to form the Brown Corpus, the major breakthrough enterprise in corpus linguistics developed and finished by 1964 in what Kennedy (1998, p. 23) called a hostile linguistic environment dominated by the theoretical and practical implications of the anti-corpus stance of Chomskyan generative grammar.

The Brown Corpus was developed to represent as wide a variety of written American English as was possible at the time. With the enormous task of transferring analog data into an electronic format done manually, the achievement is still considered a major one. The Brown Corpus contains such additional information as origin of each sample and line numbering.

2.2.3 The LOB Corpus

One rationale for the development and publication of the Brown Corpus was to provide an impetus for similar projects elsewhere. This was later answered in the late 1970s in the next major first-generation corpus project, the Lancaster—Oslo/Bergen (LOB) Corpus by Johansson, Leech and Goodluck: the British equivalent of the Brown Corpus. It was a cross-institutional effort, with the Universities of Lancaster and Oslo, and the Bergen-based center for Norwegian Humanities Computing participating. With minor differences, both the sampling and the length of the LOB followed the standards of the Brown Corpus. A more crucial difference, however, lay, interestingly, in LOB's similarity to the Brown Corpus: it, too, contained written texts produced in 1961. But as it was compiled later, the development benefited from the new technology that had become available by then. Most importantly, the advances made the use of a coding system possible, with storage in a variety of media, including three different computing platforms (DOS, Macintosh and Unix). The corpus and its manual are available through ICAME, the International Computer Archive of Modern English (Johansson, Leech, & Goodluck, 1978).

With these two language analysis resources, linguists had the opportunity to compare and contrast written U.S. and U.K. English texts, exploiting frequency and co-text information (for a comparison of frequency, see Kennedy, 1998, p. 98). Besides, the careful study of *hapax legomena*, word forms that occur once in a corpus, which typically represent the majority of types of words in most large corpora, was now possible, with implications for lexicography, collocation studies and language education.

The influence of these two first-generation corpora proved long-lasting: not only did they set standards for representation and structuring in sampling, but they also gave rise to other corpus projects of regional varieties. These included the Indian English Corpus published in the late 1970s and the New Zealand and Australia Corpora of English, each of which aimed to be modeled on the first two corpora. For the first time in linguistics, a large collection of objective data was available. But this was relative: they also contributed to the realization that the upper word limit of one million words was a restriction that had to be re-assessed and abandoned: for analysis to be based on more representative samples, linguists needed larger sets, especially for studying lexis that occurred less frequently in earlier corpora, and for contrastive analyses across the subcorpora.

2.2.4 The London-Lund Corpus

As noted earlier, the LLC, developed in Sweden, was formed on the basis of a previously statically stored corpus, the SEU Corpus. It was the first collection of spoken evidence, incorporating such descriptive codes besides the texts as tone units, onsets, pause and stress information. Although in terms of representativeness the LLC was not entirely satisfactory, it was a major step toward the integration of spoken texts in corpora.

Work on corpus development sped up in the eighties, fueled partly by the recognition that studies incorporating objective evidence made investigations more valid and reliable, and partly by the increasing facility with which to store and manipulate data. Innovations such as optical readers and software opened up the new vista of exploiting more spoken language. These developments gave rise to second-generation corpora, each based on earlier work but with different purposes and corresponding sampling principles. Another major difference between first- and second-generation corpora lies in the acceleration with which the results of linguistic analysis were incorporated in applied linguistics and language pedagogy. Of these new efforts, three projects stand out as most influential: the Bank of English, the British National Corpus, and the International Corpus of English. In each project, the activity of a national or international team, the funding of major academic and government organizations, and the economic viability of the results in the publication market continued to be operational factors.

2.2.5 The Bank of English

Originating in the seven million words of the Main COBUILD Corpus, the Bank of English is the largest collection of written and spoken English text stored on computer. Called a megacorpus (Kennedy, 1998, p. 45), its initial function was to "help learners with real English" by enabling applied linguists to do research into the contemporary language primarily for language education. The revolutionary contribution the corpus project has made to the development of learner dictionaries (*Collins COBUILD English Language Dictionary*, the original 1987 edition and the 1995 revision) has been the most influential result. A joint venture of Collins Publishers and the English Department of Birmingham University, it has provided new approaches (see, for example, Sinclair, 1987b) to lexicography. This can be seen in a number of innovations: First, in the concrete analysis of features of traditional and innovative learner dictionaries (Bullon, 1988). Second, in the research endeavor that has sprung from a need to amass more reliable data about the language. Third, in the publication business that has helped fund and maintain the scholarly interest, at least for some time (Clear, Fox, Francis, Krishnamurthy & Moon, 1996). It resulted in sampling a large database of evidence and extracting such information from it as was regarded as necessary for language learners (Fox, 1987; Renouf, 1987a): It incorporated the results in a lexical approach to language teaching that combined form and meaning, and it has been instrumental in setting high standards in corpus design and encoding (Renouf, 1987b).

Directed by Sinclair, the corpus was renamed in 1991 the Bank of English, and by now has reached a state whereby every month, some 2 million new words (tokens) are added. The team repeatedly made "the bigger the better" claim, meaning that for truly reliable accounts of lexis and grammar, large collections are necessary. The current size is 500 million words of written and spoken text, with storage on high-tech media, including the internet. To serve the growing body of researchers and teachers, a sample of 50 million words, together with concordance and collocation search engines, is available via the COBUILD Direct service of the web site at <http://titania.cobuild.collins.ac.uk> (reviewed by Horváth, 1999a).

As Sinclair noted (1991), data collection, corpus planning, annotation, updating and application continued to challenge the team. Seeking permission of copyright holders has always been among the hurdles, but there are signs of a changing publishing policy that may allow for automatic insertion of a copyrighted text for corpus research purposes.

The Bank of English has continued to innovate in all the related work: in the way corpus evidence is incorporated in learner dictionaries, in study guides and recently in a special series of concordance samplers, in the application of a lexical approach to grammar (Sinclair, 1991), and in the theoretical and technical field of marking up the corpus. Analyzing discrete meanings of words, collocations, phraseological patterning, significant lexical collocates and distributional anomalies makes available a set of new results that shape our understanding of language in use. As the reference materials produced are based on a constantly updated corpus, new revisions of these materials sustain and generate a market, making the venture economically viable, too.

2.2.6 The British National Corpus

The BNC came to be formed at the initiative of such academic, commercial and public entities as the British Library, Chambers Harrap, Lancaster University's Unit for Computer Research in the English Language, Longman, Oxford University Computer Services and Oxford University Press. The majority of its content, 90 percent, is written, with 10 percent made up of spoken samples, running to a total of 100 million words in over 6 million sentences. Any of its constituent texts is limited to 40,000 words (Burnard, 1996).

The BNC was among the first megacorpora to adopt the standards of the Standard Generalized Markup Language (SGML; more about annotation in Section 2.4) as well as the guidelines of the Text Encoding Initiative, which aims to standardize tagging and encoding across corpora. By so doing, not only has the BNC become a representative of a large corpus that has made use of earlier attempts to allow for comparability, but it also has sought to become a benchmark for other projects (Kennedy, 1998, p. 53; "Composition of the BNC," 1997). A sample of the corpus and its dedicated search engine, SARA (Burnard, 1996), have been made available at the web site <http://info.ox.ac.uk/bnc>.

The pedagogical use of the BNC has already received much attention, with Aston (1996, 1998) describing and evaluating the benefit advanced FL students in Italy gain in how they conduct linguistic inquiries. Aston reported that by accessing and studying this large corpus, students were highly motivated, primarily because of their critical attitude to published reference works that they can contrast with the results of their own conclusions.

2.2.7 The International Corpus of English

With so much cross-institutional interest and work devoted to individual projects, it was not long before researchers began pursuing the possibilities of identifying a research agenda for even more ambitious aims: to collect a corpus that would represent national and regional varieties of English. The International Corpus of English (ICE) is such an undertaking, which allows for checking evidence for comparative phonetic, phonological, syntactic, morphological, lexical and discourse analysis. Sociolinguists and language educators are also seen as beneficiaries of this corpus development drive. With Meyer coordinating the project based on Greenbaum's set of sampling procedures, the ICE represents the written and spoken language varieties of twenty countries and regions: Australia, Cameroon, Canada, the Caribbean, Fiji, Ghana, Great Britain, Hong Kong, India, Ireland, Kenya, Malawi, New Zealand, Nigeria, the Philippines, Sierra Leone, Singapore, South Africa, Tanzania, and the USA. When complete, each subcorpus will be modeled on the Brown Corpus initiative: each of the 5,000 samples in a subcorpus containing 2,000 words. (Updates on the project are posted at the ICE website, <http://www.ucl.ac.uk/english-usage/ice.htm>.) Already, work done on the ICE has informed such descriptive studies as the *Oxford English Grammar* by Greenbaum, with many more under development. A component of the ICE, the International Corpus of Learner English, will be reviewed in a later section (2.5).

The ICE project assembles text samples that represent educated language use; however, the definition of this notion is not left to the individual (subjective) decision of participating teams. Rather, the corpus will structure the language production of adult users of the national varieties of the regions. According to Greenbaum (1996a, p. 6) the texts included would be by speakers or writers aged 18 or over, with "formal education through the medium of English to the completion of secondary school." As the regional 1-million-word corpora will include written texts, identifying such factors will prove a rather difficult undertaking indeed.

2.2.8 Typology

We have seen a number of pre-electronic and electronic corpora, already noting some types: static and dynamic media, annotated and unannotated, as well as those containing written or spoken data or a combination of the two. The corpus development effort continues, and of course this subsection could review only a few of the most influential ventures. Table 2 presents a matrix of the typology of corpora, based on McEnery and Wilson (1996) and Kennedy (1998).

Table 2: A typology of corpora

By language	monolingual	parallel	
	L1	learner	
By representation	synchronic	diachronic	
	general	specialized	
By text type	written	spoken	combined
By storage	static	dynamic	
By notation	un-annotated	annotated	
By generation	first	second	
By status	set	developing	
By use	linguistic	applied linguistic	

The steps of developing these corpora and the technology used to maintain them will be reviewed in the following section.

2.3 Current issues in design and technology

Primarily, corpus linguistics offers quantitative studies of language use. It is concerned with the distribution of linguistic features within a set of texts or across samples. By using special corpus manipulation techniques such as word counts, single and parallel concordancing, linguists and applied linguists are better informed and can inform about the language they are studying. The evidence that a corpus can provide about the language, the quantitative information on frequency of word forms, on collocations, and lexical and syntactic patterns can then be applied in revealing the quality of the language studied.

2.3.1 Corpus development

All corpora are designed with a set of principles and using a sampling frame that adequately incorporates, and has the potential to explain linguistic variation across subcorpora and between corpora. The development of a sampling frame is required so that research may be able to use data that represents the population it intends to study. For this theoretical and empirical purpose, Biber (1994) suggested a cyclical model and a set of recommendations for testing the content validity and the reliability of the corpus. In this section, this model will be introduced, together with other procedures in sampling, annotation, and technical details.

The cyclicity of corpus development is a requirement as often, either the population to be represented or the text types generated cannot be defined strictly in advance. To be able to adjust preliminary concepts, a pilot study is required that can inform the effort of the population and language variables to account for. Theoretical analysis can confirm and refine initial decisions, but it may also introduce new sampling procedures. When this phase has been finished, the next step is corpus design proper. This involves the specification of the length of each component of the text (with minimum and maximum word counts), the number of individual texts, the range of text types, and the identification and testing of a random selection technique that gives each potential text an equal chance of being selected for the corpus.

During the third stage of the cycle, a subcorpus is being collected and the specifications are tested in it. This occurs in the fourth phase when an empirical investigation takes place with specifications studied and compared with the samples, and statistical measurements are taken to determine the reliability of representativeness. For any text that does not meet the requirements of the design, the specifications need to be revised, and either new design principles are identified or the problematic text is omitted. With each new sampling of a smaller unit of the corpus, constant checks and balances are in place to ensure the theoretical and empirical viability of the linguistic study that the corpus aims to serve. The Biber model is summed up in Figure 2.

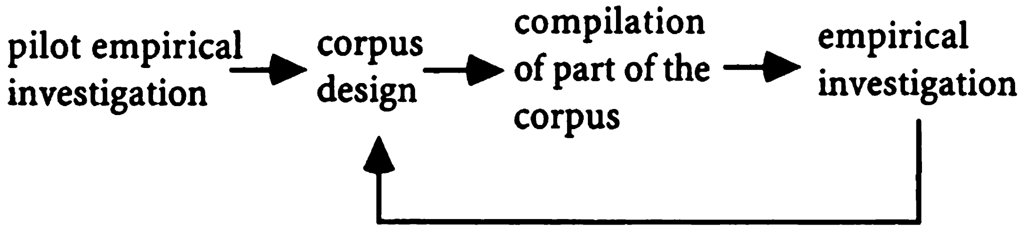

pilot empirical investigation
corpus design
compilation of part of the corpus
empirical investigation

Figure 2: Biber's (1994, p. 400) model of cyclical corpus design

Word frequency counts are strong indicators of reliability. For most general corpora, and especially those that aim to serve as bases of language teaching materials, such as learner dictionaries, establishing the frequencies of words is one of the main concerns. As this information has to be based on reliable sources, studies in representativeness provide a major contribution. According to Summers (1996), this information can then be applied in framing dictionary entries objectively and consistently, providing a dictionary that can list lexical units within a single entry according to frequency. Yet, she added, there is

> still a need to temper raw statistical information with intelligence and common sense. The corpus is a massively powerful resource to aid the lexicographer, which must be used judiciously. Our aim at Longman is to be corpus-based, rather than corpus-bound. (Summers, 1996, p. 262)

The compilation of small and large corpora was described in detail by Inkster (1997), Krishnamurthy (1987) and Renouf (1987a). One concern after the design principles have been set is that the spoken and written texts to be collected can be stored on computer; another is that what is stored there be authentic. The incorporation of electronic media poses little challenge: besides obtaining the permission of copyright holders, one needs only to ensure that the text is in a compatible format with the program used for accessing the corpus. The capture from CD-ROMs is one such relatively trouble-free area. But the compilation of non-electronic forms of texts, such as the transcription of spoken material and the typing in (or keying in) of manuscripts is far more prone to introducing error into the corpus.

Errors occurring during the entry of a text into the database should be avoided as this would defeat the purpose of representation. This is why developers need to put in place and regularly check procedures that help maintain an error-free corpus. The clean-text policy is one such procedure (Sinclair, 1991): manuscripts and other texts to be input are double-checked in the corpus.

Besides the procedural approach of designing a corpus and the need for limiting errors, the markup of the raw corpus is the third crucial area of dealing with general and specialized corpora. Most present-day corpora make extensive use of some annotation system that assigns one tag from a set of categories to units occurring in individual texts (Garside, Leech & McEnery, 1997). This process, the annotation of the corpus, aims to interpret the data objectively. Annotation can be viewed as adding a metalanguage to the language sample in the corpus, often in some form of the Standard Generalized Markup Language (SGML), an international standard.

By adding linguistic data to the raw text, a subjective element is incorporated in an otherwise objective entity. According to Leech (1997a, p. 2), there "is no purely objective, mechanistic way of deciding what label or labels should be applied to a given linguistic phenomenon." Leech focused on three purposes of corpus annotation:

- ➤ to enable linguists to extract information. Retrieving units in a corpus can be done with much more precision if word-class information is added;
- ➤ to offer further uses of the same corpus: once the grammatical tagging of a written subcorpus or the prosodic markup of a spoken collection is done, other research may benefit from the effort;

- to provide such additional values to the corpus as may be exploited by other uses; this is the multi-functionality purpose.

Tagging can now be done via computer algorithms employing designs of high sophistication, making annotation of orthography, phonetics, phonemics, prosody, word class, syntax, semantics, discourse, and even pragmatics and stylistics possible. An example of a grammatically tagged corpus may look like the one reprinted in Leech (1997a, p. 13, with word-class tags emboldened for clarity):

> Origin/**NN** of /**IN** state/**NN** automobile/**NN** practices/**NNS**
> ./. The/**DT** practice/**NN** of/**IN** state-owned/**JJ**
> vehicles/**NNS** for/**IN** use/**NN** of/**IN** employees/**NNS**
> on/**IN** business/**NN** dates/**VVS** back/**RP** over/**IN**
> forty/**CD** years/**NNS** ./.

Grammatical notation generally makes use of both automatic and manual techniques: special parsing computer software can be programmed to apply probabilistic techniques in determining classes of words. A second-generation megacorpus, the BNC, was annotated in such a way. It consists of two types of labels: header information (such as source of text) and the tagged text, using the system known as CLAWS (Constituent Likelihood Automatic Word-tagging System), which resulted in fairly reliable notation; according to Garside (1997), the accuracy rate was 95 percent or higher.

As an innovative empirical effort, Garside, Fligelstone and Botley (1997) provided an example of annotating discourse information in a corpus. Whereas most other levels of tagging can benefit from high technology, the area of cohesive relations poses major difficulties. Reviewing models of markup, the team worked out a fairly consistent method and an additional set of guidelines that may be further trialed and adjusted. Already, the notation system can describe such elements as antecedents and noun phrase co-reference, central pronouns, substitute forms, ellipses, implied antecedents, metatextual references, and noun phrase predications. Any unit not adequately captured is noted by a question mark. Although the authors recognized that the field of discourse annotation "is at a fairly immature stage of development" (Garside, Fligelstone, & Botley, 1997, p. 83), exploiting SGML and refining the tagging algorithm may achieve the sophistication of other levels of annotation.

2.3.2 Concordancers: Functions and packages

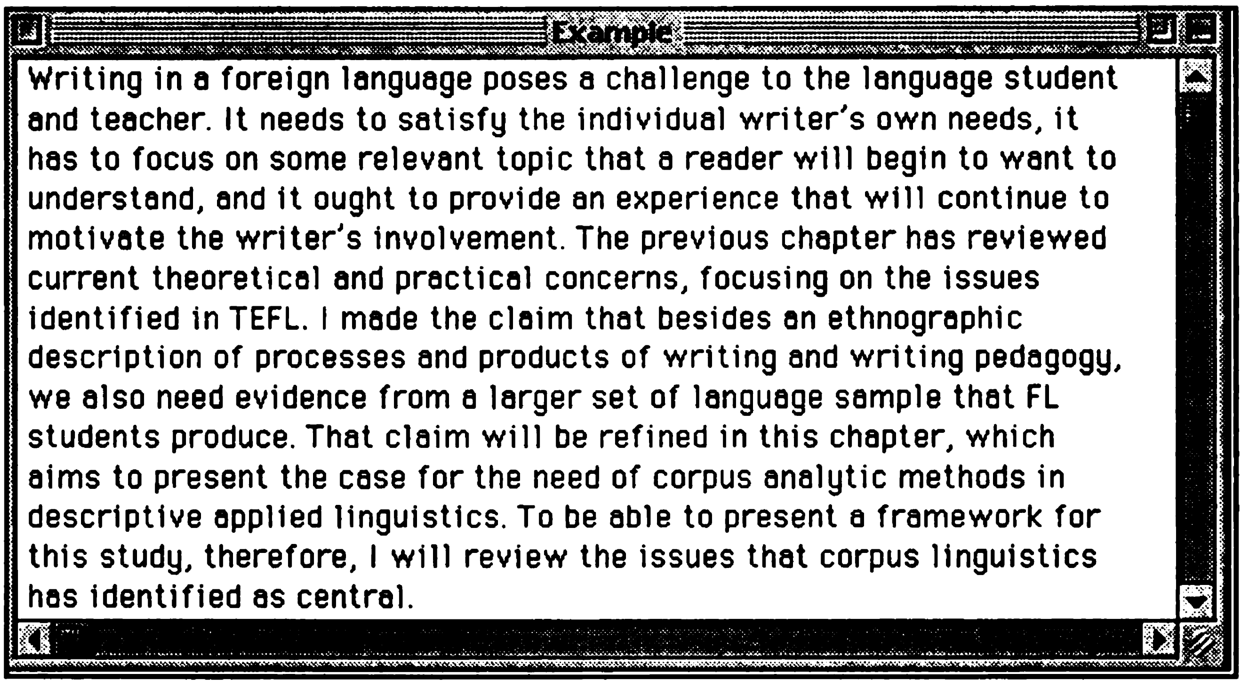
When a corpus, either in its piloted state or when the database has been checked for representativeness is assembled, the corpus linguistic analysis *per se* may begin. Typically, for this purpose a computer program is used, which has at its core either a general-purpose concordancing module or a dedicated software package developed to deal with specialized annotated corpora. Of the many publicly available concordancers, I will present five here. Each program has a set of special features (or tools, as they are often labeled) at their center; however, all such programs are similar in that their main functions can be broken down into the following four domains:

- opening a text file;
- generating a concordance output on screen and to print;
- generating various indices;
- saving files for later retrieval and editing.

Longman Mini Concordancer is a simple DOS-based application, suitable for the swift input of small corpora used in several language classes. *Mini Concordancer* served as the starting point for Vlaskovits's *Contour* (1996), a DOS-program designed for wide access in Central Europe (the author of the program received initial guidance and beta testing from me). Scott and Johns's *Microconcord* (1993) is another widely used application—its advantages over Longman's product include the feature that there is no internal limitation on corpus size and the modules that allow for editing concordance entries, enabling teachers to produce classroom materials. Scott's *Wordsmith* is a set of corpus linguistic tools, available for the Windows environment (Lázár, 1997). A powerful application, *Wordsmith* offers high speed and concordancing features that make it a popular program for work with large corpora.

The program used for the analyses presented in this book was *Conc* 1.7, a Macintosh application that can process text files limited only by the size of the computer's hard disk and memory allocation. Minimally, the program requires 512 kilobytes of memory, and 160 KB of hard disk space. *Conc* was developed by the Summer Institute of Linguistics in Texas in 1992. I selected this shareware utility for its user-friendliness and reliability: during the five years of using it, it has proved stable. As the corpus analyst inevitably will have to share disks with others, another consideration was that of platform stability: In the Windows/Intel world, shutdowns resulting from malicious computer viruses have become a frequent occurrence. By contrast, the Macintosh system is virtually free of such troubles. Here is a description of the features *Conc* offers, together with screen shots that illustrate them.

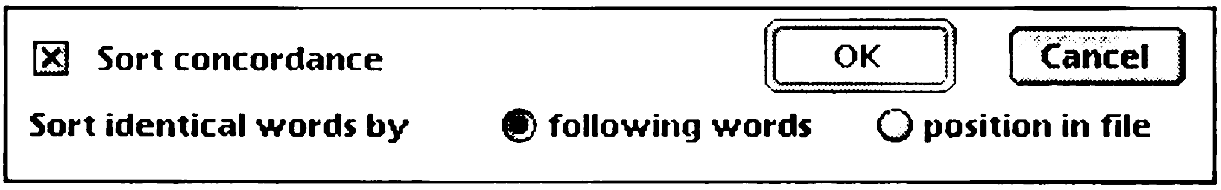
For the description, I selected an earlier version of the first paragraph of this chapter: a 153-word minitext. After saving this paragraph in text-only format, I launched the concordancer and opened the file. The window presented in Figure 3 appeared.



Writing in a foreign language poses a challenge to the language student and teacher. It needs to satisfy the individual writer's own needs, it has to focus on some relevant topic that a reader will begin to want to understand, and it ought to provide an experience that will continue to motivate the writer's involvement. The previous chapter has reviewed current theoretical and practical concerns, focusing on the issues identified in TEFL. I made the claim that besides an ethnographic description of processes and products of writing and writing pedagogy, we also need evidence from a larger set of language sample that FL students produce. That claim will be refined in this chapter, which aims to present the case for the need of corpus analytic methods in descriptive applied linguistics. To be able to present a framework for this study, therefore, I will review the issues that corpus linguistics has identified as central.

Figure 3: The example text in *Conc's* main window

When generating a word concordance output to screen, the user has the option of sorting identical words (types) according to the words that follow them or according to their position in the original file. As Figure 4 shows, I selected the former choice.

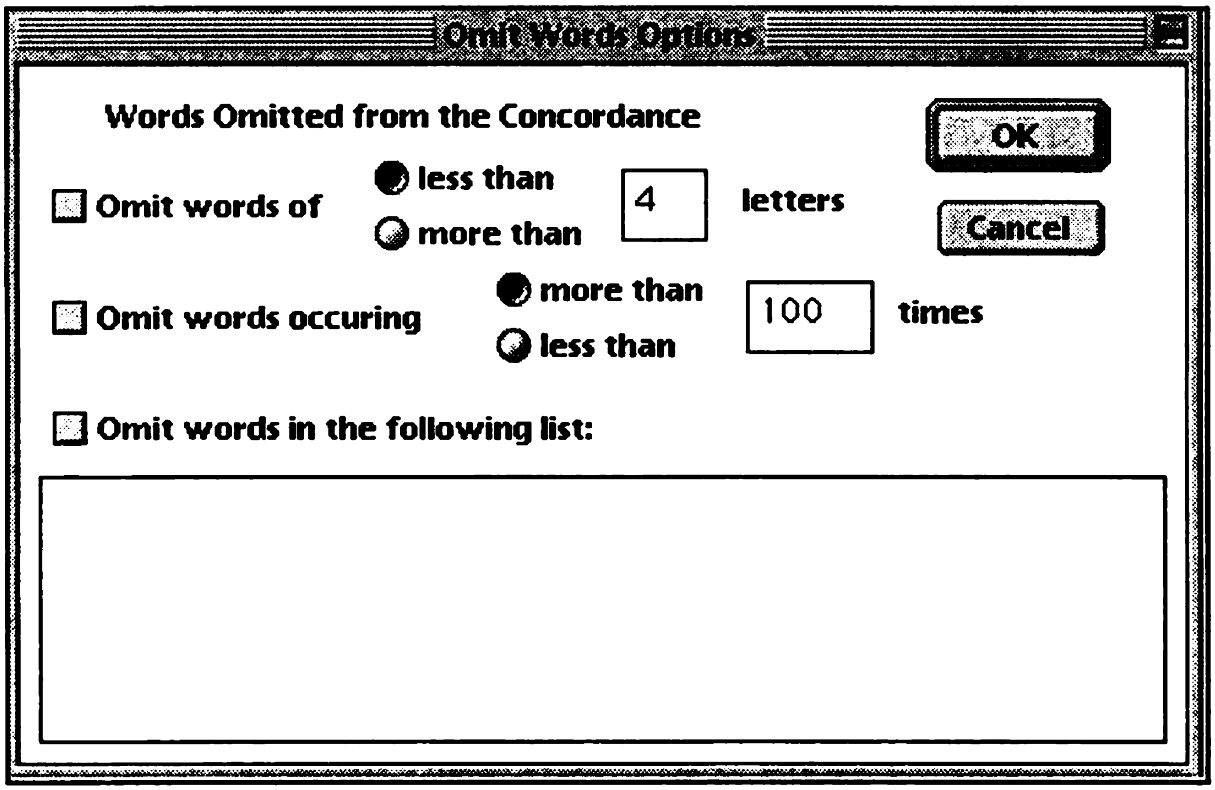


☒ Sort concordance

Sort identical words by ● following words ○ position in file

Figure 4: Part of the Sorting Parameters dialog window in *Conc*

As users may not wish to display all words occurring in a text, the program lets them deselect words from the concordance. Three options are available for this, as Figure 5 demonstrates. When none of the options are selected, the program performs a full concordancing of the text. A combination of word omission features, however, makes it possible to focus, for example, on *hapax legomena* (by selecting the *Omit words occurring more than 1 times* option) or on words that occur a number of times and which are longer than six letters.



Words Omitted from the Concordance

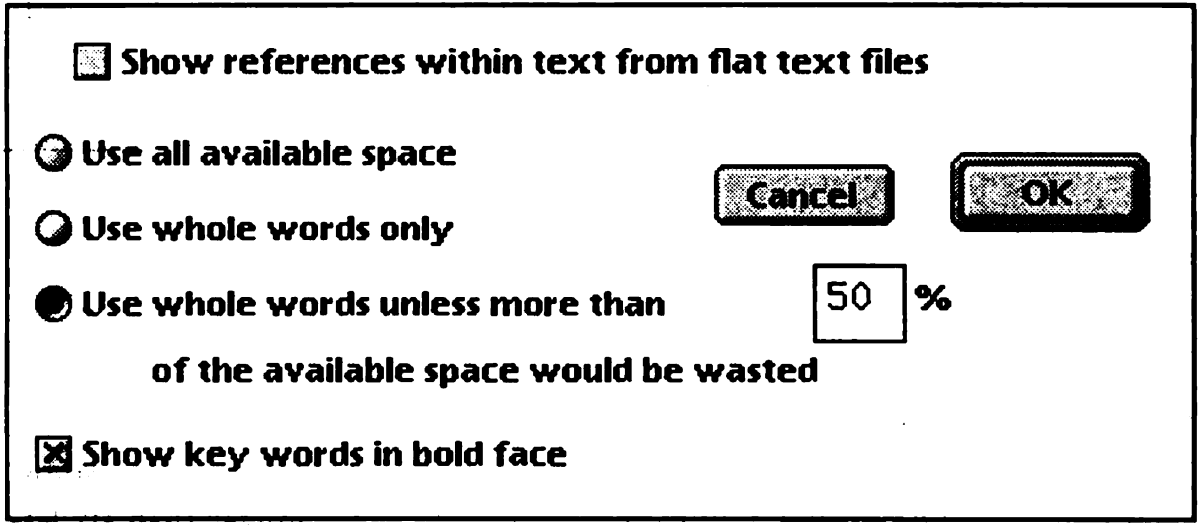
☐ Omit words of ● less than ○ more than 4 letters

☐ Omit words occuring ● more than ○ less than 100 times

☐ Omit words in the following list:

Figure 5: The dialog window where words may be omitted from the concordance

Another module lets the user define the filling of the display space, given in radio button options: all of it could be filled with the concordance lines, resulting in truncated words; only full words could be shown; or the program could compute a compromise between the two options. A typographical standard is presented in the check box to show key words in bold face (see Figure 6).



☐ Show references within text from flat text files

○ Use all available space

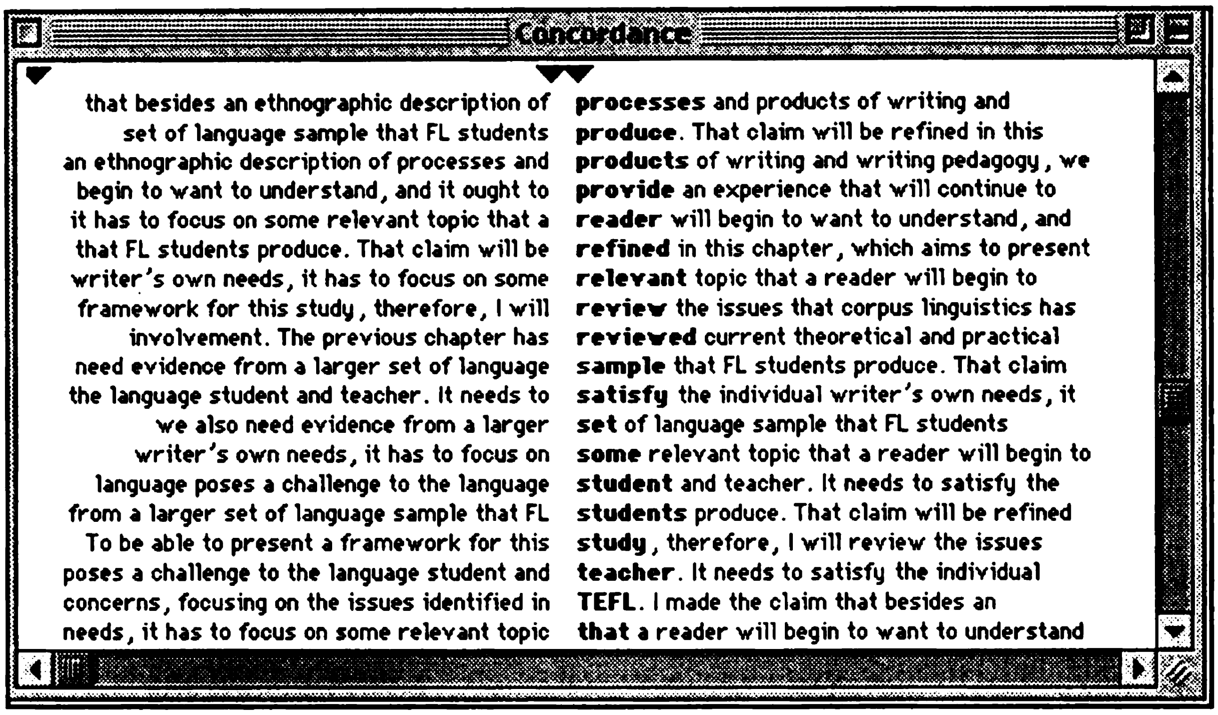
○ Use whole words only

● Use whole words unless more than 50 % of the available space would be wasted

☒ Show key words in bold face

Figure 6: Part of the Display dialog box on the Options menu in *Conc*

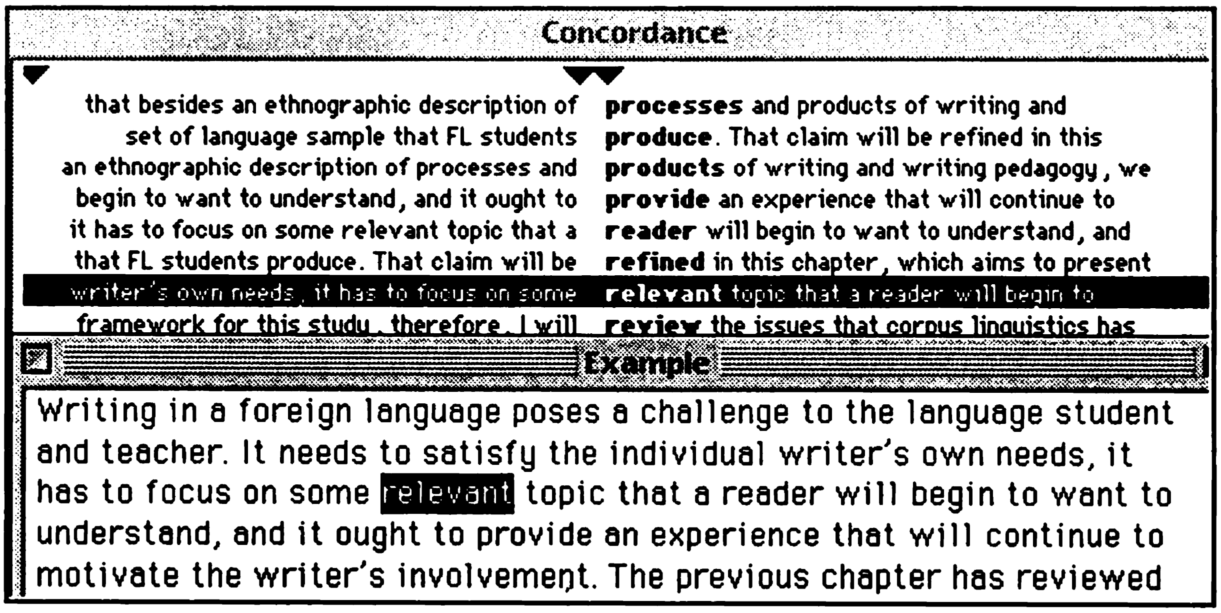
When such parameters have been set, the program is ready to sort the text file accordingly. A new window appears as a result, of which Figure 7 presents a part.



```
  that besides an ethnographic description of   processes and products of writing and
      set of language sample that FL students   produce. That claim will be refined in this
 an ethnographic description of processes and   products of writing and writing pedagogy, we
    begin to want to understand, and it ought to   provide an experience that will continue to
      it has to focus on some relevant topic that a   reader will begin to want to understand, and
      that FL students produce. That claim will be   refined in this chapter, which aims to present
   writer's own needs, it has to focus on some   relevant topic that a reader will begin to
     framework for this study, therefore, I will   review the issues that corpus linguistics has
      involvement. The previous chapter has   reviewed current theoretical and practical
  need evidence from a larger set of language   sample that FL students produce. That claim
   the language student and teacher. It needs to   satisfy the individual writer's own needs, it
          we also need evidence from a larger   set of language sample that FL students
      writer's own needs, it has to focus on   some relevant topic that a reader will begin to
     language poses a challenge to the language   student and teacher. It needs to satisfy the
  from a larger set of language sample that FL   students produce. That claim will be refined
    To be able to present a framework for this   study, therefore, I will review the issues
 poses a challenge to the language student and   teacher. It needs to satisfy the individual
  concerns, focusing on the issues identified in   TEFL. I made the claim that besides an
   needs, it has to focus on some relevant topic   that a reader will begin to want to understand
```

Figure 7: Part of the Concordance window of the program

It is in the Concordance screen that the user can first study the co-texts of the keywords, shown in bold face, and centered as key word in context (KWIC) concordances. When a co-text does not reveal sufficient information, and thus should be enhanced with the fuller context, one can switch between the main window and the Concordance window. With the appropriate line of the concordance output selected, the main window can be superimposed and the full sentence studied. This is shown in Figure 8.

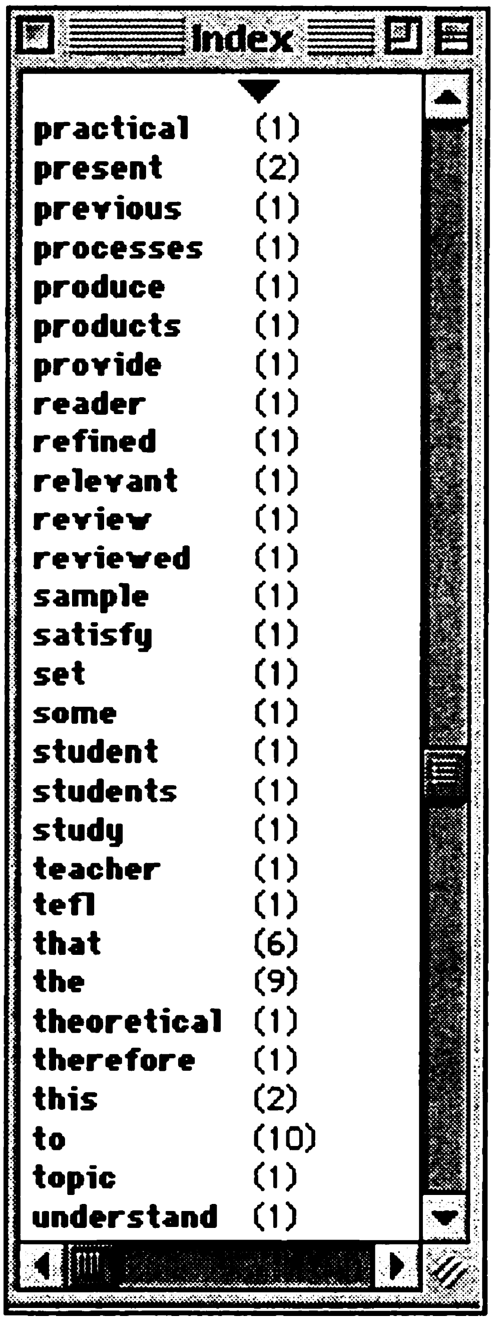


```
  that besides an ethnographic description of   processes and products of writing and
      set of language sample that FL students   produce. That claim will be refined in this
 an ethnographic description of processes and   products of writing and writing pedagogy, we
    begin to want to understand, and it ought to   provide an experience that will continue to
      it has to focus on some relevant topic that a   reader will begin to want to understand, and
      that FL students produce. That claim will be   refined in this chapter, which aims to present
   writer's own needs, it has to focus on some   relevant topic that a reader will begin to
     framework for this study, therefore, I will   review the issues that corpus linguistics has
```

Writing in a foreign language poses a challenge to the language student and teacher. It needs to satisfy the individual writer's own needs, it has to focus on some relevant topic that a reader will begin to want to understand, and it ought to provide an experience that will continue to motivate the writer's involvement. The previous chapter has reviewed

Figure 8: The Concordance and the main windows

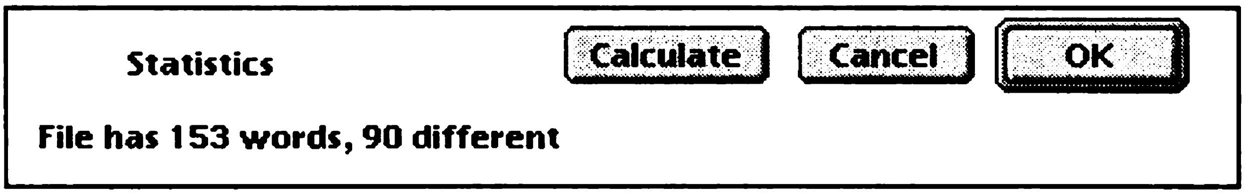
Conc 1.7 can provide one type of index for texts: alphabetical. As this can be saved to a text file, a database program can be used for sorting words by frequency. (This procedure will be described in Chapter 4). Figure 9 displays a screen shot of part of the Index window. First-occurrence word lists and frequency lists can be generated directly in other programs.



```
practical    (1)
present      (2)
previous     (1)
processes    (1)
produce      (1)
products     (1)
provide      (1)
reader       (1)
refined      (1)
relevant     (1)
review       (1)
reviewed     (1)
sample       (1)
satisfy      (1)
set          (1)
some         (1)
student      (1)
students     (1)
study        (1)
teacher      (1)
tefl         (1)
that         (6)
the          (9)
theoretical  (1)
therefore    (1)
this         (2)
to           (10)
topic        (1)
understand   (1)
```

Figure 9: The Index window's scrolling list of the alphabetical index of the file

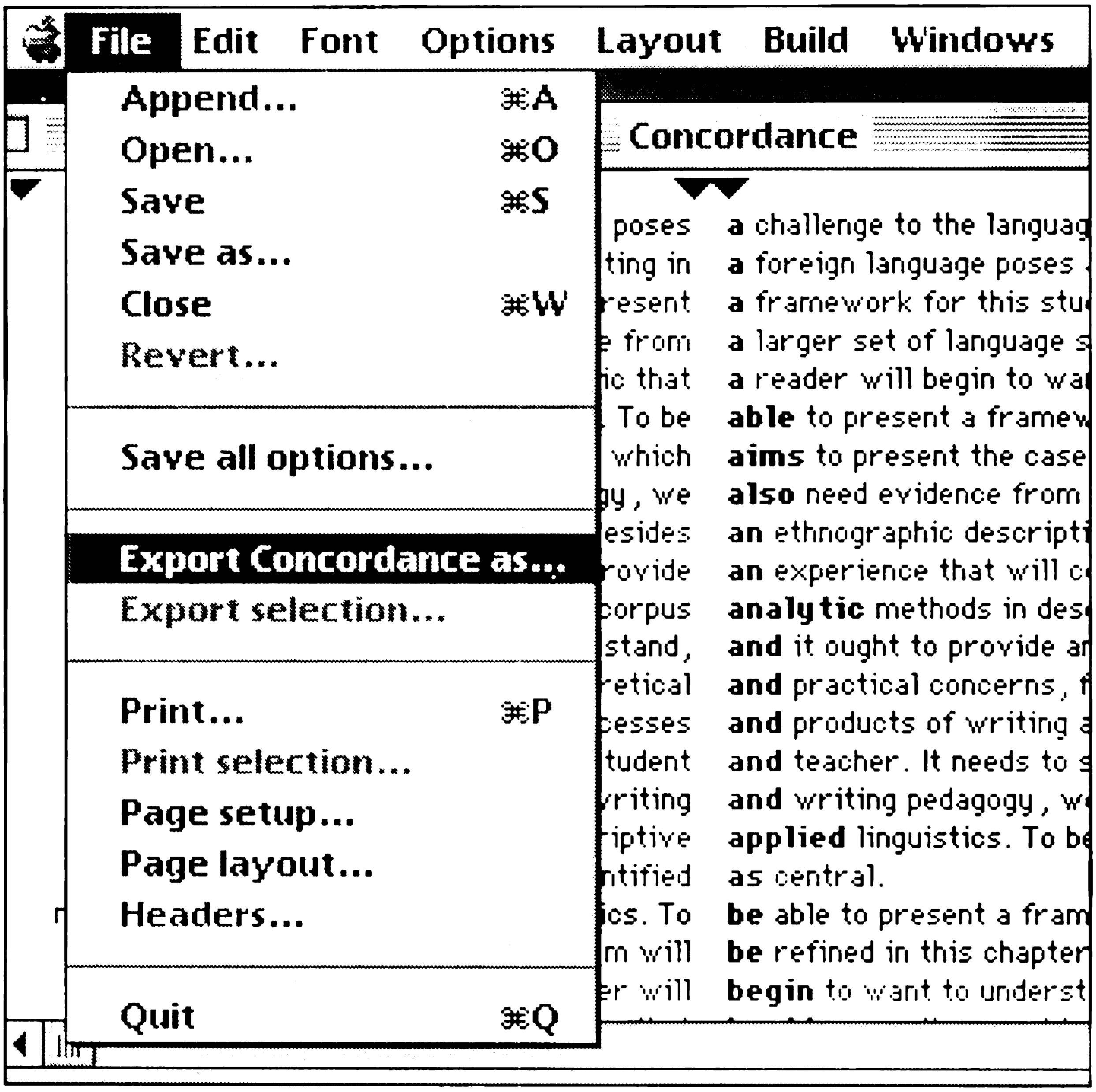
Simple statistical word count information is also provided. The example paragraph used contained 153 tokens, 90 types (see Figure 10).



File has 153 words, 90 different

Figure 10: Part of the Statistics window on the Build menu

There are several file management options that *Conc* 1.7 provides. New files can be added to texts, another may be opened, a selected concordance can be saved or printed, current parameters can be saved as default options. Of course, the full concordance can be exported, too (see the menu selection screen shot in Figure 11).


File
Edit
Font
Options
Layout
Build
Windows
Append... ⌘A
Open... ⌘O
Save ⌘S
Save as...
Close ⌘W
Revert...
Save all options...
Export Concordance as...
Export selection...
Print... ⌘P
Print selection...
Page setup...
Page layout...
Headers...
Quit ⌘Q
Concordance

```
 poses   a challenge to the languag
ting in  a foreign language poses
resent   a framework for this stu
e from   a larger set of language s
ic that  a reader will begin to wa
 To be   able to present a framew
 which   aims to present the case
gy, we   also need evidence from
esides   an ethnographic descripti
rovide   an experience that will c
corpus   analytic methods in des
stand,   and it ought to provide a
retical  and practical concerns, f
cesses   and products of writing a
tudent   and teacher. It needs to s
writing  and writing pedagogy, w
riptive  applied linguistics. To b
ntified  as central.
ics. To  be able to present a fram
rn will  be refined in this chapter
er will  begin to want to underst
```

Figure 11: The File menu of *Conc*

2.3.3 Principles and techniques in corpus analysis

The use of concordancing programs such as *Conc* provides the raw data for corpus studies. In any corpus linguistic endeavor, the units of analysis have to be defined after this so that conclusions made about the corpus are reliable and valid. Biber, Conrad and Reppen (1998) called these units the observations, each text in the corpus being one such unit. Another type of observation is a single linguistic feature across texts in a corpus. As the primary contribution of corpus linguistics to our knowledge of language use is aimed at a quantifiable piece of information, such studies need to be carried out on a solid statistical basis so that we can identify significant variables.

To achieve this aim in a comparative study that investigates a set of linguistic features across texts in a corpus or between corpora, Biber, Conrad and Reppen (1998) described the procedure whereby a so-called normalization of linguistic variables is performed. In essence, this involves the identification of a unit of the text that will serve as the basis of comparison. Table 3 shows one example of such a normalized comparative analysis. In this analysis of three news items and three conversations (identified as text files in the first column and as labels in the second), the length of each text is given in a word count (in column 3). For each of the three observations (verbs, adjectives and pronouns), the unit of analysis was one thousand words; the numbers indicate the occurrence of these types in each of the six texts per 1,000 words. Normalization applies a simple formula. The frequency of the observation is divided by the word count and multiplied by the unit in which the linguistic feature is analyzed. Normalization, then, refers to the process of establishing comparability among observations. According to Biber, Conrad and Reppen (1998 p. 263), it is a "statistical process of norming raw frequency counts of texts of different lengths." The results of normalization (the rates) for these observations are the quantitative data that can be compared across the texts, using statistical methods.

Table 3: An example of normalized comparative analysis (based on Biber, Conrad & Reppen, 1998, p. 273)

Text ID	Register	Word count	Rate of past tense verbs	Rate of attributive adjectives	Rate of first person pronouns
n1.txt	news	2,743	47.4	68.1	3.1
n2.txt	news	1,932	49.2	63.0	9.2
n3.txt	news	2,218	42.2	74.8	7.1
c1.txt	conv	2,197	32.2	43.1	62.6
c2.txt	conv	2,542	37.4	36.3	59.1
c3.txt	conv	2,107	36.8	39.7	58.7

One type of the frequently extracted statistical information is the mean score of individual items. Not only can the normalized frequency information on individual variables within a text be informative, but also the mean score, for example, of text length within a register and across registers. Once mean averages are computed, comparisons can be made. Studying Table 3, for example, we have evidence to suggest that news tends to have more past tense than do conversation text types.

Statistical measures such as the mutual information score and the T-score, the analysis of variance (ANOVA) of lexical collocation, and chi-squared counts are used to determine whether a linguistic phenomenon occurs merely by chance or whether it is statistically significant. Corpus linguists have increasingly sought to establish whether any observed difference between normalized frequency counts is the result of chance, or whether there is statistically significant correlation between them. Such measurements have long been applied in other social sciences, and there has been growing linguistic interest in them (Biber, Conrad, & Reppen, 1998; Clear, 1993; Kennedy, 1998; Koster, 1996; McEnery & Wilson, 1996).

2.4 Data-driven learning: CALL with classroom concordancing

The previous sections have outlined the theoretical justifications for the use of large computer corpora in language description and the procedures of the approach in describing linguistic phenomena in a valid and reliable manner. As the corpus revolution occurred during the technological advances of the eighties and early nineties, it is not surprising that the practitioners of the language teaching approach commonly known as Computer Assisted Language Learning (CALL) have attempted to apply the results of corpus linguistic research and innovation. Besides, as interest in building and exploiting L1 corpora has continued to grow, so too has the initiative to collect L2 scripts for corpus development purposes. This section will provide an overview of the convergence of CALL and corpus linguistics in data-driven learning, together with current techniques of applying the approach in language education. The section following this discussion will present the rationale and aims of learner corpus projects.

2.4.1 Computer assisted language learning

The field of CALL (and the related discipline of information technology) has been the domain of much classroom innovation, especially in the U.S., but also one that has not been able to come to terms with its inherent dependence on behaviorism. CALL brought personal computers into the language class, established self-access centers, developed courseware that aimed to individualize grammar practice, and contributed to the technological know-how of teachers and students. It can be viewed as an approach to language teaching that incorporates the procedures and theoretical foundations of several methods. Early applications relied heavily on drill-and-practice exercises familiar in the grammar-translation tradition, on the exploitation of authentic materials, with more humanistic approaches and a need for more interactivity appearing lately.

In both the language software and its way of delivery, many CALL practitioners assumed that extended time spent online would result in better performance. Although there was little scholarly attention focusing on the effectiveness of CALL in the 70s and early 80s (reviewed by Chapelle & Jamieson, 1989), anecdotal evidence and the enthusiasm of scores of language educators and of students continued to attract financial and pedagogical investment. Stevens (1989), however, remarked that much CALL experience in the U.S. and elsewhere failed to revitalize the behaviorist orientation that assumed that learning will take place when discrete steps are planned properly. This is somewhat surprising, considering the amount of work put in this enterprise, and the expansion of the approach supported by such organizations as TESOL and IATEFL. Arguing for a shift in this paradigm, Stevens called for computers and software in language education to be viewed and applied as facilitators of what he called humanistic learning.

This call for a pedagogical change meant that CALL software and its application had to be based on much more concrete applied linguistic principles. Although attention to sound methodological grounding was called for as early as 1986 by Jones, much CALL business remained within the confines of the grammar-translation tradition. Stevens (1989), aiming to synthesize SLA theory, specifically the hypotheses of Krashen (1985), summed up the features that were worth exploiting as follows. First of all, CALL software had to be able to create intrinsic motivation for the learner. In other words, such courseware would need to be relevant to student needs, offer authentic tasks, and create a no-risk environment, resulting in a low affective filter. Second, he proposed that CALL applications develop more fully the interactive potential of the technology. For example, programs can do this by adjusting their routines based on the input of the individual student, a principle gaining ground in computer-adaptive testing much more effectively than in teaching. Finally, Stevens made a call for non-CALL programs; the value of eclecticism lay, he argued, in that software "designed for other audiences and purposes" (1989, p. 35) could and should be adopted in the language class.

Wolff (1993) shared this view of applicable technologies in language learning. Also concerned with more direct integration of SLA research, he identified four principles for exploiting information technology in language education (p. 27):

- the provision of a rich, motivating learning context;
- the application of materials that take account of individual learners' strategies;
- the aim to assist learners in discovering processing and learning strategies;
- the goal of developing autonomy in learners.

How this takes place in specific educational contexts, however, needs more research. In the Hungarian secondary-school system, Nikolov's (1999) study found no evidence of information technology being applied. Teachers reported lack of access to high technology that schools did possess, but it was unclear who owned them and how they were to be used for what purposes. According to Sankó (1997) much more administrative, pre-service and in-service training is necessary for any large-scale integration of information technology in Hungarian education. It remains to be seen how the current re-evaluation of the promising educational project of the Sulinet Program (Gadó, 1998) can facilitate the further dissemination of the technology.

Where CALL has been introduced either in an isolated project (Horváth, 1994a, 1997a; Turi, 1997; Rósa, 1995) or as a school-wide undertaking, it has helped provide a pedago-technological innovation that has facilitated the acquisition of computer skills, thus providing a practical spin-off to language education. In this regard, CALL has been instrumental in connecting generations of students and teachers in the community of computer literate people.

2.4.2 Discovery in data-driven learning

We now turn to data-driven learning (DDL). The basic principle of this approach to language teaching, especially at intermediate and advanced levels, is that learners need to discover new knowledge about language themselves, rather than being told answers to their questions. Pointing out that much of what goes on in a traditional question-and-answer session arises from the fact that the teacher knows the answer, Johns (1991a, 1991b) posited that there are linguistic queries that the teacher cannot solve with any degree of precision without access to a large corpus. If the teacher has the corpus, it is time the students had the same opportunities. DDL teachers, then, came to act as an interface between CALL and corpus linguistics: the teacher became a facilitator by planning the overall scheme of a course, but the students were given the initiative in exploring authentic examples.

DDL is viewed (Farrington, 1996; Sinclair, 1996, 1997) as a possible "new horizon" in CALL because it offers the foreign language student opportunities to engage in authentic tasks in a low-risk environment, truly interacting with authentic texts, and using appropriate tools. In short: DDL, in many ways, incorporates the values Stevens (1989) set forth. Without the rapid development in the field of corpus linguistics, however, and without its many lexicographic and grammar applications, the approach would not have become so effective. Johns (1991a) attributed the growing interest in DDL specifically to the COBUILD project.

DDL, which may be regarded as a subdivision of CALL, first appeared in the late 80s, early nineties in Johns' work with international students studying at British colleges (1991a, 1991b). Drawing on the results that CALL had established in the U.S. and the U.K. (Higgins & Johns, 1984; Pennington, 1989), he helped set up a program that would provide what he called "remedial grammar" tools and training for science students. Johns argued that advanced EFL students had a need to directly exploit the growing evidence a corpus was able to provide. He offered a model (shown in Figure 12) to explain the nature of language awareness processes taking place in such a context.

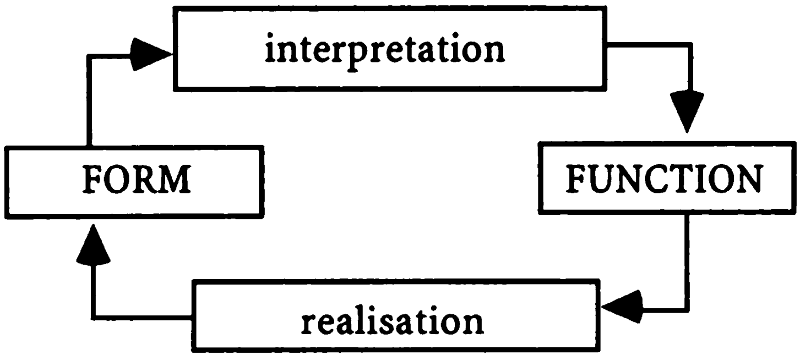


Figure 12: Johns's model of data-driven learning (1991a, p. 27)

As Johns was primarily concerned with the development of language awareness as it related to the needs of advanced students, he hypothesized that those who aimed to develop accuracy in the foreign language had to be able to understand the relationship between how functions of discourse are realized in forms, and how these forms are interpreted to satisfy them. Data is crucial in such a process: rather than inventing examples to explain to students how this happens, students and teachers need hard evidence of how forms are used in context. This is the rationale for the central position of data, with the roles of the student enriched by that of the researcher during the participation in classroom concordancing activities (such as those described in Tribble & Jones, 1990).

Data is authentic unmodified language extracted from a corpus (Johns, 1991b, p. 28). In Johns's remedial grammar and academic writing classes, students were actively involved in accessing, manipulating and exploring this data, partly by online classroom concordancing, and partly by participating in individual and pair work activities based on new types of exercises developed to take account of the data. One corpus used in the project was a 760,000-word sample of the journal *New Scientist*.

Data drives learning in the sense that questions are formed in relation to what the evidence suggests. Hypotheses are tested, examples are reviewed, patterns and co-texts are noted. The collaboration that evolves between students and the teacher who may not know the answer without also consulting the corpus carries a further innovative element of this approach. Students also have the opportunity to focus on clearly defined units in the data (Higgins, 1991; Kowitz & Carroll, 1991; Stevens, 1991). A spin-off of the approach was presented by Johns (1997a): new CALL programs, such as his *Contexts*, can be designed by incorporating concordance tasks piloted in the classroom.

The materials developed are another outcome of the approach. The technique of on-line concordancing has allowed for the generation of new task types, such as the one keyword, many co-texts activity, or the concordance-based vocabulary tasks described by Stevens (1991). Corpora also allow for the development of innovative and potentially effective approaches to and applications of pedagogical grammars (see, for example, "The Internet

Grammar of English," 1997; Hunston & Francis, 1998). Also, research investigates how what is presented in traditional language coursebooks may or may not be supported by the evidence of the corpus (Sinclair, 1997; Mindt, 1996, 1997). As DDL and corpus evidence in general become mainstream, as was suggested by Svartvik (1996), new FL materials, too, will benefit from the approach.

2.4.3 Applications of DDL

The researching student testing hypotheses about language with the help of data will, however, continue to need guidance from the expert teacher (Owen, 1996). For this need, Johns has recently suggested another interactive technique: the assistance by the "kibbitzer" (1997b). This essentially means that he is making available to an international audience the queries students had when working on dissertations and writing chapters. Students would identify a lexical, syntactic or pragmatic problem, and Johns would look up the corpus to assist in dealing with it, essentially providing a parallel concordance. Patterns in the data are highlighted, and a suggestion is made on how to revise the problem item, with the student being ultimately responsible for the final decision. Such an approach to revision appears to be beneficial, but there is yet scant empirical evidence to support claims about its effectiveness. One report, by Hadley (1997), attested that in a Japanese beginner EFL class, DDL proved a welcome transition from traditional sentence-based grammar tuition procedures.

Gavioli's (1997) example offered yet another insight into the application of concordancing activities in language education. She introduced multilingual corpus analysis processes and interpretation tasks designed for a course of translators in Italy. Gavioli emphasized the importance of consulting reference materials to test hypotheses about language use. By analyzing and interpreting data in a corpus, and by corroborating their own discoveries, students can become the ones who describe features of language, rather than being offered such descriptions. The singular contribution of these applications of corpus materials in language education is the exploration of authentic texts that raise awareness of significant patterns used in natural contexts. As suggested by Kennedy (1998), such inductive use of corpus texts in classroom concordancing helps FL students to "locate...all the tokens of a particular type which occur in a text...and note the most frequent senses" (p. 293), thus discovering collocational and colligational features. Leech (1997b) and Kirk (1996) were among those positing such applications as experimentation with real language, besides recognizing their value in academic study. Kirk underscored the change this brought in language teachers' roles: as teachers' roles are enriched by being providers of an authentic resource, they can coordinate research initiated by students (1996, p. 234). Clearly, this has the additional benefit of empowering students, mostly on intermediate and advanced levels, so they can gain experience in a new skill, too.

Another value of DDL lies in the manner in which teachers can establish and maintain a classroom-based research interest themselves. By applying corpora in their syllabus design and class materials development efforts, they are bridging the gap between research and pedagogic activities, a trend welcomed by Dörnyei (1997) and Ellis (1995, 1998), among others. One example of such involvement was offered by Tribble (1997), who described an innovative use of a multimedia product whose text component was used as a corpus. The author proposed that teachers who find it difficult to access large corpora or who do not regard the use of one as relevant can use multimedia encyclopedias as language learning resources. Targeting EFL students beginning to work with academic writing, the syllabus incorporated the multimedia product *Encarta*, a set of hypertexts, movies and graphics containing such diverse text types as, for example, articles by experts in the fields of physical science, geography, history, social science, language and performing arts. Tribble claimed that using this resource not only caters for diverse student interests in the writing course but can result in their recognition of different text organization and lexical preferences in descriptive and discursive essays, process descriptions, physical descriptions and biographies.

2.5 Learner corpora: Issues and implications

2.5.1 The International Corpus of Learner English

Most DDL activities are directed toward the manipulation of L1 corpora. They involve the tutor and the students in work similar to that done in the development of reference materials based on corpora; they contribute to a growing awareness of how users of the language studied apply the idiom principle; and they focus on improving the accuracy of the learner. Research has begun to address issues related to the development of learner corpora, too. Such projects began in the early nineties, partly to satisfy a need to verify or refute claims about transfer from the mother tongue to the foreign language. Among these drives, the Louvain-based International Corpus of Learner English (ICLE) was the forerunner, with part of the Longman Lancaster Corpus (LonLC) and the Hong Kong University of Science and Technology Learner Corpus following suit. Conceived by Granger (1993, 1994, 1996), the ICLE collection of written texts by advanced students of EFL aims to be the basis of lexical, grammatical and phraseological studies.

The main objective is to gather objective data for the description of learner language, which Granger (1998a; in press) saw as crucial for valid theory and research. Besides, the ICLE's contribution has been in directing attention to the need for observation of this language so that the notion of L1 transfer may be analyzed under stricter data control. The obvious potential outcome is for materials development projects, which will help specific classroom practices. (*Longman Essential Activator*, 1997, was among the first dictionaries to incorporate learner data derived from the LonLC.) Focusing on error analysis, and interlanguage (Selinker, 1992), the ICLE-based project enables researchers and educators to directly analyze and compare the written output of students from such countries as France, Germany, the Netherlands, Spain, Sweden, Finland, Poland, the Czech Republic, Bulgaria, Russia, Italy, Israel, Japan and China.

Part of the ICE project, the developers of ICLE identify the origins of interest in the analysis of learner language in early error analysis SLA studies. Granger pointed out (1998a) that although the investigations and theoretical explanations made about learner errors were grounded in data observation, the corpora for those studies did not take full account of the variables that affected the samples. For example, the number of students, their learning experience and often non-comparable test elicitation techniques raised doubts about the reliability of some of those observations. By contrast, the ICLE project has worked out a system of sampling scripts that allows for more reliable studies in the description phase as well as in contrasting individual subcorpora and a subcorpus with an L1 corpus.

With each script, detailed information is recorded in the contributor's profile. This not only ensures that the data comes from a valid source, but also allows for specific analyses of types of language use in clearly defined subcorpora. The descriptors include, according to Granger (1996, p. 16):

- biographical information: nationality, age and gender
- English learning experience: years of formal English studies and stay in an English-speaking country
- other learning experience: knowledge of other languages
- task- and text-related details: conditions of writing the script (test or non-test, timed or untimed, and use of reference tools).

2.5.2 The composition of the ICLE

The target word count of the ICLE is two million words. The scripts are primarily argumentative essays, covering a variety of topics, with a smaller set of scripts made up by literature examination essays (see a list of the essay titles recommended for national contributors in Appendix A). As the aim is to collect and analyze authentic learner scripts, the designers pointed out in their call for submissions that essays should be "entirely the student's own" and that "no help should be sought from third parties." This specification, however, raises two problems: one theoretical, the other pedagogical.

First, as a number of the assignments do not appear to involve much of the students' own deliberation as they present an argument that they need to support, no matter what their own positions, the validity of a text being a student's "own" is dubious. Even if students have the chance of choosing a title or a theme, they cannot "entirely own" their writing as they play a limited role in deciding on the focus of their essays. For this reason, the title "Europe" may be regarded of the suggested ones as the most authentic: it does define a clear enough focus, allowing students to develop an argument which is truly their own, yet specific for any lexical or rhetorical analysis when the text becomes part of the corpus.

As for the pedagogical implications of the preferred mode of submitting a student's "own" essay with "no help...sought from third parties," the authenticity of the task may be lessened. With so much written production viewed and undertaken as a collaborative process effort in the L1 field, it is somewhat surprising that no peer or teacher involvement is allowed. The specification also raises the issue of audience: the themes appear to favor the production of writer-based prose; yet the task is defined as an argumentative one where awareness of the position of the audience is crucial. Furthermore, why deny the opportunity of consulting a reader before the script is finalized if one were to follow, even for such a basically product-oriented enterprise as corpus development, a process syllabus? Considering the role that editors, colleagues and publishers play in the finalization of the written work of L1 authors (represented in L1 corpora), it stands to reason that such restriction in the development of L2 corpora may bias the comparative analyses.

These constraints notwithstanding, the ICLE has ushered in the time of interest in more specific analyses of learner language. Each of the national subcorpora will be about 200,000 words, allowing for grammatical and lexical investigations, but small for research into words and phrases of lower frequencies (Granger, 1996, p. 16). However, the project has been instrumental in helping an international team of researchers and teachers to join forces in the field (Ringbom, 1998; Lorenz, 1998; Virtanen, 1998; Petch-Tyson, 1998; Kaszubski, 1998, among others), and in leading the way to new inquiries: for the development of more specialized ESL and ESP corpora. Another area where the ICLE has motivated research is the advanced spoken learner corpus and the intermediate corpus, both under development. Work on L2 corpora is gaining recognition, and the practical implications of these efforts may be seen shortly in the new reference and teaching materials that take account of L2 learners' language use (Gillard & Gadsby, 1998; Granger, 1998a, 1998b; Granger & Tribble, 1998; Kaszubski, 1997, 1998).

2.5.3 Other written learner corpora

Besides the large-scale work of the ICLE and the LonLC, there are several other projects that have attempted to capture what is significant in learner texts. Of these endeavors, Tono's (1999) and Mark's (1997a, 1997b, 1997c, 1998) work merits recognition. Both are individual teachers' initiatives, but the aims and the applications are slightly different. The Tokyo Gakugei University Learner Corpus consists of 700,000 words written by lower-grade and upper-grade Japanese students' of EFL. One of the largest such collections in Japan, it has been used primarily for interlanguage error studies (Tono, 1999). The Meiji University Learner Corpus is smaller, made up by

220,000 words (Mark, 1997a, p. 93). Mark's interest focused on exploiting the data in syllabus design, helping students in examinations, and materials development. This latter objective was conceived as especially important because textbooks available for advanced Japanese students of EFL did not seem to reflect the needs arising from the status of their interlanguage (Mark, 1997a).

2.6 Concluding remarks

In this chapter, I have presented the case for employing corpora for language description and education. Describing corpus linguistics as an empirical study of naturally occurring language use in context, I have evaluated the theoretical contrast between generative linguistics and text-based language analysis. I have reviewed the development of various types of L1 and L2 corpora and recent work done in the field internationally. The scope of application has widened, with corpora set to affect the way language tests are validated (see, for example, Alderson, 1997; and Horváth, 1998c). Besides, teacher education and materials development can also benefit from corpus linguistic techniques (Bocz & Horváth, 1996; Hughes, 1997; Minugh, 1997; Renouf, 1997; Wilson, 1997).

Interest in applying corpora in linguistic analysis and materials development is on the rise in Hungary, too. Studies that are partly or entirely based on such corpora as the Bank of English represent a new trend in current Hungarian linguistics. Among these, Andor (1998), for example, applied a sample from this corpus, together with data elicited from forty native speakers of English, in the study of the mental representation and contextual basis of ellipsis and suggested that a combined use of psycholinguistic and corpus linguistic research methods would enable linguists to arrive at more valid and reliable conclusions. Csapó (1997) studied the viability of the convergence of pedagogical grammars and learner dictionaries, Hollósy (1996, 1998) reported on work to develop a corpus-based dictionary of academic English, whereas Szirmai (2000) investigated translation equivalence by using corpus linguistic methods.

The framework of DDL and the increasing interest in analyzing learner English on the basis of learner corpora will be applied in the following chapters: the next describing and analyzing writing pedagogy at the English Department of Janus Pannonius University, and the fourth giving an account and analysis of the JPU Corpus. The study of learner scripts contributes to the authenticity of writing pedagogy: those who collect, describe, and analyze L2 texts can test, in a valid and reliable way, hypotheses of the effectiveness of writing pedagogy. Also, such collections can serve as a basis of an innovative type of learning material that can be applied directly in the writing classroom.

Chapter 3

WRITING PEDAGOGY AT THE ENGLISH DEPARTMENT: PRODUCING PROCESSES

> *...many writers are paralyzed by the thought that they are competing with everybody else who is trying to write and presumably doing it better. This can often happen in a writing class.... Forget the competition and go at your own pace. Your only contest is with yourself.* (Zinsser, 1998, p. 79)

Introduction

A teacher undertaking to help students of FLs develop a language skill needs to have clear concepts of their needs, the requirements of the discourse community, and the variety of procedures whereby they can be met. The ability to place a skill and its many subskills in the wider context of language learning and use is an additional prerequisite. As with any skill at any level, the development of EFL writing skills at university also has repercussions for continued growth and motivation. Clearly, the stakes are high: students may become more or much less motivated to study and perform, depending partly on the opportunities they have had in preparatory courses.

I have been teaching EFL at the English Department of JPU since 1989. Originally hired to run first- and second-year Language Practice classes in the undergraduate program, I have also participated in the development and teaching of language development courses in the Russian retraining and the postgraduate in-service programs. Besides, my responsibilities have included the design, administration, piloting and analysis of two types of language proficiency test.

In all of these activities, the contribution I have attempted to make to the quality of education at the institution has been in the development of the personal narrative and academic expository writing skills of students. To be able to present a coherent analysis of relevant factors and procedures in the space available I have had to make a concession: This chapter will focus on undergraduate writing courses, even though the JPU Corpus comprises scripts by other students. This means that I have had to exclude lessons learned in Language Practice courses for undergraduate and Russian retrainee students, as well as writing courses for in-service postgraduate students, from whom I have received contribution to the corpus. However, as

Chapter 4 will show, the majority of scripts do come from the student population discussed here.

This chapter will introduce the development of writing skills at the ED of JPU. To be able to appreciate the work students have done to develop personal and academic writing skills and to see the context of their scripts presented and analyzed in Chapter 4, we need to identify the curricular and pedagogical factors that have shaped this performance. After listing the types of data used for the discussion of these factors (in Section 3.1), the chapter will focus on the major components of writing pedagogy. The description and evaluation of pedagogical concerns and of the curricular status of academic writing (3.2) will be followed by a detailed evaluation of the process of developing the syllabus for recent writing skills courses (3.3). Finally, in Section 3.4, future directions will be drawn on the basis of this discussion.

3.1 Data and participants

To provide a description of the institutional and curricular role that EFL writing pedagogy plays at JPU, I will use qualitative and quantitative data that comprises the following sources:

- the Language Development curriculum specifications of the English Department;
- syllabuses of recent Formal Writing and Writing and Research Skills (WRS) courses;
- students' portfolios containing narrative, descriptive and argumentative essays;
- students' research papers;
- writing textbooks reviewed, adopted and otherwise used during the development of the WRS courses;
- handouts used in WRS classes;
- notes on classroom activities;
- notes on employing various techniques and formats to provide feedback on student writing;
- records of students' performance in a number of WRS courses;
- notes on students' activities in office hour meetings;
- questionnaires and other instruments developed to assess students' attitudes to courses.

With the exception of the first source of data, the curriculum, these materials and documents have been prepared continuously as I have participated in various teaching and testing activities. All are authentic records of the individual activities they represented at the time—applying them for research analytical purposes will enhance the validity of the ethnographic objective of this discussion. The enterprise is unique: to my knowledge, no other

Hungarian writing tutors have attempted to evaluate their pedagogy in writing. It is hoped that the results of this endeavor will motivate further empirical research in the field, for the benefit of all concerned.

A total of 336 students have participated in the writing courses I have taught in the past three years, representing the majority of students who have gained admission to JPU since 1996. Of these participants, 204 took undergraduate WRS courses, with the remaining 132 pursuing postgraduate studies.

3.2 Pedagogical concerns and writing in the JPU ED curriculum

As Chapter 1 showed, interest in raising standards in writing pedagogy has contributed to the re-assessment of the role of several factors. One is the discourse community that shapes the modes of communication and socializes the novice student. Another factor is the identification of writing skills that make up the construct of writing. With the continuing influence of process-based and humanistic approaches to language education, teachers of writing in diverse social and educational contexts are addressing more effectively the theoretical and practical concerns of their profession. A third factor is represented by the relationship between writer (student) and readers (students and teachers): their interaction results in a rich and motivating experience, which is essential in continued growth.

3.2.1 Principles and their sources

When undertaking to participate in writing skills development in university EFL courses, I had already been teaching courses that contained a writing component. However, it was my participation in proficiency testing projects that first formally introduced me to student writing on a department-wide scale. For two years, I had learned the trade of marking student scripts before launching two sections of a Formal Writing course in 1996.

Apart from my role as teacher and tester, I had for some time been collecting student scripts for action research purposes. I became familiar with the concerns of students, was able to observe their decisions in writing, and began to develop a set of materials that exploited a growing corpus of learner English.

Yet another strand of my concern with student writing derives from various activities that aimed to help provide a forum of student voice. This line of interest was represented by two types of journalistic activity: founding and co-editing, with a colleague, Paul Olchváry, a JPU English magazine, *The Pannonius Post*, and editing several classroom magazines for students enrolled in undergraduate and Russian retraining Language Practice courses (such as *SnaX*, *Every Thursday*, *Talent*, and *The Friday Gazette*). I aimed to provide classroom materials that were authentic in the sense that they (1) communicated my motivation, (2) attempted to enhance students' integrative motivation, and (3) aimed to widen the scope of communication.

Such experiences have appeared to contribute to students' involvement with their own discourse communities and in the classes. For example, *The Pannonius Post*, between 1990 and 1996, helped initiate dozens of students into the art and craft of article writing, editing, and publishing. Interviews, reports, news stories, poems, short stories and reviews by students and faculty were published, contributing to the ethos of the department. Classroom magazines, such as *Talent*, invited students to explore the campus and discover talented peers in one field or another. In such extracurricular projects, students seemed to benefit from the discovery of knowledge that they found relevant to learn about and to publish, which was an especially valuable factor given the potential risks that the university credit system posed in group forming.

My involvement in these curricular and extracurricular projects was complemented by a third type of activity that bears directly on my role as a writing teacher: editing. In 1996, I became co-editor (together with Nikolov Marianne) of the Hungarian ELT and Cultural Studies journal, *Novelty*. This publication was in its second volume when the publisher, the British Council in Budapest, approached us to consider taking on the role. The daily tasks of soliciting articles, reading them, suggesting changes in focus and tone, the technical skills of establishing the use of a standard referencing system, and the contact to be kept with contributors, readers and the publisher provided me with experience and skills that are central for the writing teacher: both the wider issues of constructing and reconstructing meaning, visualizing structure, appreciating a solid research design, arriving at valid conclusions, and reverberating with readers; and the finer details of understanding and evaluating sentence- and word-level authorial choices, and establishing consistency in spelling and punctuation.

My work as co-editor positioned me as a suitable candidate with various types of teaching, extra-curricular and editing experience. With each new writing course syllabus prepared, I aimed to incorporate what I had learned so that my pedagogical concerns were met: that students would participate in classes that gave them opportunities to express and explore themselves, and that they would be equipped with skills that would enable them to continue to improve.

3.2.2 Writing in the curriculum

Educational curricula identify a field of study, its content and structure, and specify the goals and requirements for individual components. University curricula of individual departments also specify the input and the output of the courses and establish relations between other curricula. They are in constant revision as new needs arise and as units of education can cater to address those needs. Wide variation, however, can sometimes be seen in terms of explicitness of goals and methods.

The 1998 curriculum caters for two types undergraduate and two types post-graduate course of study (*Tantervek*, 1998). Each of the four types requires the passing of an entrance examination. In the undergraduate course, a centrally designed written exam is administered to high-school graduates, followed by an oral exam developed and assessed by department staff. By contrast, students wishing to gain admission to the postgraduate course are required to possess a teacher's diploma and pass an oral exam.

3.2.2.1 The undergraduate core curriculum

The curriculum of the undergraduate course is also controlled by its output options: students either study for a first degree in English Linguistics, Literature and Education or in Linguistics and Literature. The latter is further divided into two options: major and minor. The difference between the two options is in the number of elective course credits to be completed: in the major, 54 electives are to be chosen, whereas only 24 in the minor. What is common is the core: in both options, this offers a set of 66 credits. Thus, 120 credits is the requirement for the major, and 90 for the minor option. Table 4 presents the divisions of the core curriculum.

Table 4: The eight divisions of the core curriculum

Division	External prerequisites	Credits
Language Development (LD)	none	16
British Culture (BC)	none	6
American Culture (AC)	none	6
Linguistics	none	18
Applied Linguistics	none	2
British Literature	11 credits from LD, BC, AC, and AS	9
American Literature	same as for British Literature	6
Anglophone Studies (AS)	none	3

In terms of the specific content of the eight divisions, two set prerequisites for students for taking courses. The others also have prerequisites, but these are set from within. The majority of courses in the core can be taken independently of courses in the other divisions, with students making up their own timetables based on the information they receive from the curriculum and the separate list of courses issued each semester. There are no external prerequisites for Language Development courses either. As can be seen, this first division is one of two strands that are given most weight in the core curriculum. Together, Language Development and Linguistics contribute over half to the core. This is illustrated by the pie chart (see Figure 13).

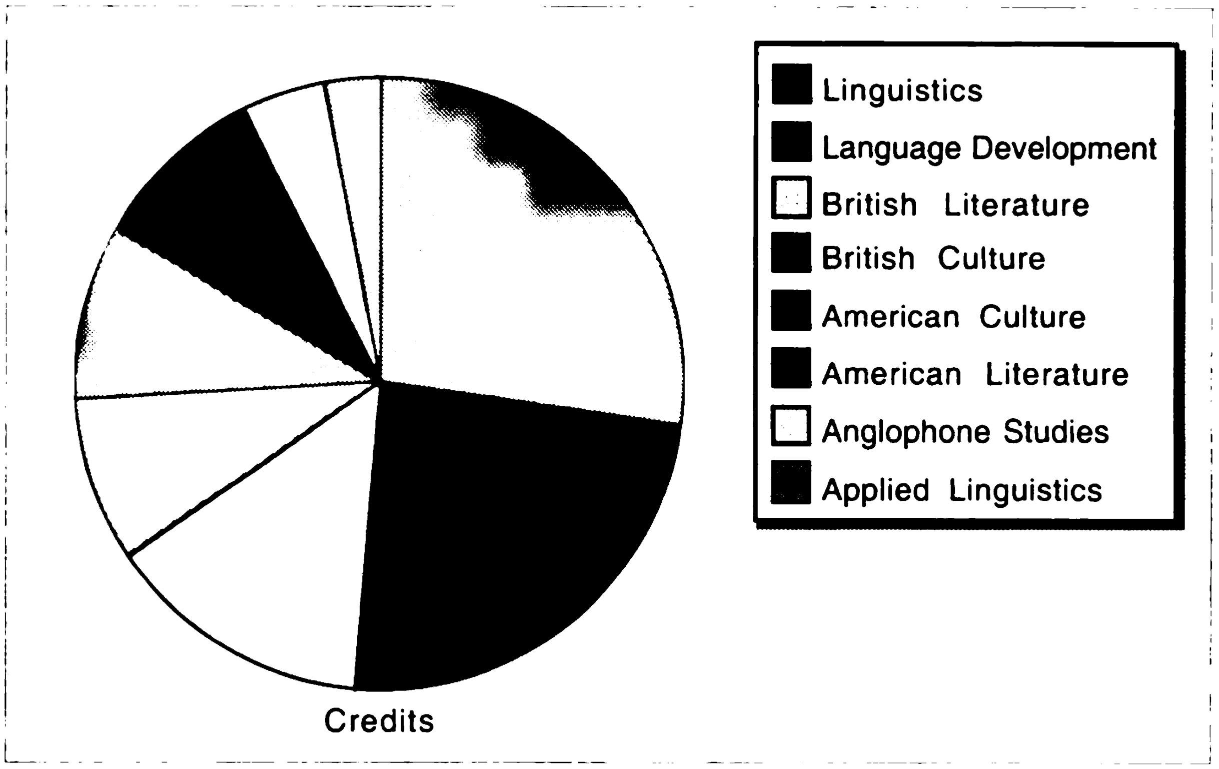


Figure 13: The respective weight of each of the eight divisions in the core curriculum

The Language Development is made up by eight courses, as illustrated in Table 5. As the prerequisites column shows, the main course in the division is Language Practice, making up a half of all credits. It is also the only course that sets registration requirements.

Table 5: The framework of the current Language Development division

Course	Credits	Prerequisites
Language Practice 1	2	none
Language Practice 2	2	LP 1
Language Practice 3	2	LP 2
Language Practice 4	2	LP 3
Writing and Research Skills	2	none
Linguistics Discourse	2	none
Literature Discourse	2	none
Culture Discourse	2	none

The core curriculum places much emphasis on language development. Of the 16 credits to be completed in the LD division, eight come from Language Practice seminars, two from the WRS course, with an additional six represented by introductory courses to the study of linguistics, literature and culture. The WRS course represents that part of writing pedagogy at the ED which is specifically devoted to writings skills. Most other courses within and outside the division include written assignments in their requirements.

However, an analysis of such content and requirements is beyond the scope of the current study—future department-wide longitudinal action research could reveal the role of such requirements and text types.

3.2.2.2 The postgraduate curriculum

The official name of the postgraduate program is Supplementary Training for English Language and Literature, offered to college graduates. Two types of course operate: one for English major graduates, and another for graduates of lower-primary teacher training colleges with a specialization in English. The major structural difference between the undergraduate and postgraduate curricula is that the latter does not make a distinction between core and electives.

The preamble of the curriculum specifies the output of the course, the method of training, and the sequencing of the course types. Graduates earn a degree that qualifies them to teach English language and literature at high schools. Participating in correspondence courses, they study four types of subjects: applied linguistics, linguistics, and the literature and culture of English speaking counties.

As the Faculty of Arts operates a credit system, this is adopted in the postgraduate course as well. However, students enrolled in the program do not have the option to choose courses—this being the result of the correspondence type of education. The constituent courses are specifically designed for these students, and are not open to others. Theoretically, students in the postgraduate program could register for others, but because the groups meet only five times a semester, there is no practical relevance of this option.

A total of 48 credits have to be completed for a degree, a fourth of which come from one of the four areas of study. After the fourth semester, the last period of study is available for students to complete their theses—during this period they are required to consult with their advisors.

3.3 Syllabus development

A syllabus is the most important official document of a course of study. It has to be based on the curriculum it aims to support, it records objectives and methods of reaching them, the input and output requirements, and it provides a basis on which to compare various stages of development of an educational program. It is also a piece of technical writing that has multiple audiences: first, the faculty that oversees the validity of the approaches and objectives; the students it addresses; and the administration that files such materials. Designing a syllabus that is based on solid educational principles, offers reasonable flexibility, sets manageable targets, and on top of it even enhances the motivation of the primary audience, the students, is especially crucial in a writing course. It is the first piece of authentic writing the students receive from the tutor—the content, style, and even typography of such a document communicates much about what students can, cannot, and should not expect.

The Writing and Research Skills course had its origins in the Formal Writing course established at the ED of JPU in 1986. The development of the syllabus of this course in its early stages relied on product-based approaches to writing. It was after such beginnings that I embarked on my own writing teaching career. Throughout my activities I aimed to incorporate those findings of the field and my own experience that appeared to better contribute to success.

The following, mainly qualitative, study is based on the records I have kept of five undergraduate courses run between the Fall of 1996 and the Fall of 1998. In the description and analysis of the course syllabuses, I will compare and contrast the courses in terms of objectives, tasks, techniques, text types, readings, feedback, evaluation and students' views.

3.3.1 Objectives

The first writing course inherited the name Formal Writing. I did have doubts about the appropriateness of this designation, yet it was not possible in the transition period to change it. The objectives that the Fall 1996 syllabus communicated (see "Course Description" in Appendix B) included the development of skills in "university formal writing assignments," with three distinct text types identified: in-class expository writing, "longer" take-home assignments, and the thesis. Two of these text types are fairly concrete examples of the academic writing tradition, whereas the "take-home assignment" is a less distinct genre.

Specific writing subskills were also identified: of the five listed in the syllabus, the one that appears the most relevant in terms of syllabus development was the last set of subskills—appreciating, analyzing and commenting on other students' writing in "a professional manner." Unless students were given opportunities to share their scripts, the writing teacher would run the risk of creating a vacuum, instead of creating a forum. By accommodating peer reviews of scripts, I aimed to develop a sense of community in the two groups of students.

The tone is formal, with students addressed in the third person plural. The variety of additional information, such as time and place of office hours, the telephone number, and the internet address of selected course materials, however, added a personal dimension to the document.

One seemingly immaterial syllabus-writing decision merits reference, before we move on to the next document—the position or role that the writing teacher identifies with. This can be detected in how the name of the teacher is introduced in the appropriate heading of the syllabus. Hungarian university tradition seems to prefer the position of the "instructor," partly perhaps as an effect of Anglo-Saxon academic preferences. Of the many conscious decisions

I made in designing my first WRS syllabus, the change in denomination was one. Instead of referring to my role as that of an instructor, I took the position of "tutor."

There were two reasons for this. For one, this was the term I had used in earlier courses, and I saw no reason to want to change. For another, and this is the more important aspect, I never viewed the act of helping students to learn to write better as an activity that can be achieved by instructing. That approach seemed to offer little in the way of negotiating meaning, allowing for personal differences in learning style and strategy, and I saw it as seriously limited in its potential of establishing a learning environment that would engage sustainable development. Opting instead to act as the "tutor" of students, I argued, told the students that I considered myself an expert in the field but that I was primarily concerned with individuals and teams of students to be motivated in discovering the power of writing for their own benefit.

The term "tutor" is about the only detail that is common between the first and the Spring 1997 syllabuses. Reflections of the positive results and shortcomings of the earlier course, and the application of the theory and empirical research with which I had become more familiar by the time I was producing plans for the new course, enabled me to introduce innovations that were far-reaching. One of these was the decision to officially change the name of the course. It was no longer a "Formal Writing" course of study, but one that focused on "Writing and Research Skills." As we will see in a later section (3.3.3.2), the first WRS course also included a research element, even if at that time it was far from being integrated into the texture of the course. The change in name reflected a change in approach and content. For the first time in the history of JPU ED writing pedagogy, there was a course that operated with reasonably specific academic terms. (See the Spring 1997 syllabus in Appendix C.)

These terms were used in the "Aims" section of the syllabus: the course, offered to three sections of students, proposed to address and improve writing and research skills that were to be developed during the seminars. It emphasized success: the course would "empower [students] to achieve" it in such discourse types as were seen as essential in the design, planning and execution of descriptive and review essays and research papers. In terms of writing processes, the stages of conceiving, structuring, editing, drafting and presenting were outlined.

The communication of the syllabus was still relatively formal, with the tutor referring to himself in the third person singular, and to the students in the third person plural. But the classes were now termed as "meetings," which occurred in two of the three sections in the Arizona Room of the university, a computer network facility that promotes dynamic and effective group work made possible by the *GroupSystems* courseware. In terms of content, another innovation was the introduction of the concept of plain English. As can be seen in the syllabus, the reference to this quality of writing appeared in the "Course themes" section. In later courses, the concept gained central position.

By the Fall semester of 1997, eighty-five students had taken these courses. The goals of motivating students to experiment and of establishing a firm base on which future development was possible had been established. In designing the new syllabus (see in Appendix D), I aimed to emphasize the need for both extensive reading and writing. Another syllabus design element that can be considered new was the use of the concept of a "center" which would provide a framework for the study during the semester.

The course was identified as "The Fitness Center," a place where the three groups of students would be assisted in "putting [their] writing skills into good shape by allowing [them] to work out and get the right amount of nutrition and protein." These metaphors were meant to communicate to the participant that writing economical, clearly structured texts could be achieved. As will be shown in the next sections on task and text types, although goal setting used terms that may have confused some students, the texts to be produced were the most concrete to date.

This syllabus was the first to break away from the formal tone tradition: the tutor welcomed participants to the course, spoke to them directly, and at the end expressed the hope that students would have a "useful and memorable experience."

The lessons learned in the fall of 1997 further motivated development. The Spring 1998 course can be seen as a stage that had established what appeared most effective approaches and content, including the continued centrality of aiming to assist students in producing plain and transparent text in English for personal and academic purposes. My dual role of teacher and editor, as well as my exploration of the theory and practice of writing pedagogy, had by now confirmed that this was a feature of writing I aspired to cultivate in students' writing.

Specifically, the syllabus made reference to the development of fluent, accurate, and plain written English. It also communicated the goal that the seminars would encourage experimentation with "topics, genres, audiences, and purposes." The output of the course was identified in proficiency in writing four types of text, of which the personal descriptive essay was the new example. (See the syllabus in Appendix E.)

The tone is similar to that of the syllabus in the preceding semester: it addresses the student as a stakeholder, a participant, using simple and clear language. It also continues with the metaphor of the "center," but this time it is a "writing center," as opposed to the "fitness" center a semester earlier. The reason for the change was that, although the WRS course continued to focus on "low-fat" English and energetic text, the term "fitness center" was regarded as politically incorrect. By the time I was preparing the syllabus, I received word that at least one student with a limited physical condition had become an English major.

Another notable feature of the text of the syllabus is that this was an example of paying attention to layout and packaging: icons and symbols provided visual information and aimed to facilitate cross-referencing between the weekly program of the sessions and the requirements. The titles of the sessions, quotations from one of the required readings, aimed to further raise students' curiosity.

It was after such work that the most recent WRS course I designed opened, in two sections, in September, 1998. In the first session, students received the syllabus presented in Appendix F. The "Description" section retained the elements that had become the staple of the course, but was complemented by a focus on "opinions, observations in personal narrative and descriptive essays" as the text types that the course aimed to help students develop. Experimentation, writing processes, and the research elements were included as the other building blocks.

A procedural innovation was the employment of student assistants. Students from the previous course were asked to consider volunteering to team-teach a session with me. Eight students expressed such willingness, a relatively high number, given the fact that this was not widely practiced at the university and that the offer was made when students were working on the revision of their research papers.

The closure of the syllabus is an example of how a teacher can frame a document of this type: if the audience is greeted at the beginning, a final personal remark seems to be relevant. In this instance, I expressly made the point that I was looking forward to the "time we will be spending together, and to your ideas and texts. I wish you a memorable and exciting time in the writing center." The syllabus, for the first time in the history of its development, operated with the first person plural pronoun, placing the students and the tutor in the context of a shared community.

3.3.2 Tasks and techniques

When objectives are identified in the syllabus, teachers reflect on past experience of what worked and what needed adjustment, and consider the professional literature, attend conferences to revitalize their teaching and cooperate with other colleagues. Other sources of monitoring progress are inviting peers from the same department to observe classes, and eliciting and acting on students' feedback. The objectives that the WRS courses set were to be reached by classroom and out-of-class activities. A review of these two major types of tasks and techniques will follow in this section.

The reason for dividing the activities into two categories was the relatively short time available for group meetings. Courses had an average span of thirteen weeks, with 90-minute sessions a week. As early as the first course in 1996, this was supplemented in two ways. First, office-hour meetings were always announced and students made welcome in them. I regarded these meetings as essential for the fulfillment of course goals, especially because the average group had 20 students. The other way of making more time available was that the course did not end when the semester was over: volunteering students received encouragement to revise their papers in exam periods. Although I have not kept continuous records of all office hour meetings and all students' revision choices, the majority of students came at least once to the office meetings, with many choosing to frequent these occasions throughout the semester and beyond. A similar tendency was shown for revising: the overwhelming majority of the students decided to revise.

3.3.2.1 Classroom techniques

The tasks applied in the sessions were tightly connected to the text types developed. As the syllabuses of the past five semesters indicate, the majority of sessions were devoted to workshop tasks intended to promote group learning and sharing. Some of these tasks were present in all semesters, others in a few, yet others in one only.

Tasks and techniques tended to follow a cycle: introductory sessions inquired about students' experience of reading and writing. Pair and group discussions were initiated to establish a cooperative network, one where students were willing to share their ideas, orally and well as in writing. As my writing pedagogy aimed to follow a process approach, cyclicity meant that after the introductory sessions, the different levels of text construction were dealt with as discrete elements and holistically.

The emphasis was on student participation: the course aimed to achieve sustainable development, which I hypothesized would be possible by fostering a classroom where questions, critique, and opinion are raised freely.

Besides this element of the classes, a few sessions incorporated a lecture component where I presented views on writing, often supplemented by illustrations from students' scripts. The lecture part aimed to make students aware of the larger issues of writing: processes in writing, audience, purpose, writer's voice, and plagiarism, and it also aimed to establish a link between individual sessions and the overall purpose of the course. Such presentations tended to include a metaphor: to bring fundamental issues closer to real experience, I devised several ways of describing the nature of writing. In one instance, the metaphor even became the central element of the course—in the Fall 1997 course, each element of the WRS was designed by the images incorporated in the metaphor.

In these approaches, I was led by theory and practice: different learning strategies need motivation from a variety of sources—the high-level cognitive load of construing writing quality and processes can be experienced via lower-level stimulus. The practical consideration was that such presentations and the ensuing discussion and application contributed to a lively classroom, with enhanced group dynamics. In developing writing habits and attitudes, images can be applied as a framework to plug the gap between the familiar and the unfamiliar. They can also motivate students to devise their own metaphors, share them, and attempt to use them as personally relevant elements of writing strategies. (See a brief discussion of the photographer, the plane flight, and the slim plain English metaphors in Appendix G.

A culmination of the metaphor approaches can be seen in the five T tips: these presented paragraph-level conventions and notions of signaling a paragraph with indentation, what its topic is, in what tense the ideas are presented, how the tone of the writer exposes the topic, and how one paragraph may contribute to the unity of the text. These basics were incorporated in the Spring 1998 course, and then a semester later they appeared as the tips. They specifically aimed to provide students with a technique that was easy to remember and which could inform revision. The tips are a simple checklist of five questions the writer can ask in developing or revising a text:

> One Tab: Have I indented this paragraph?
> One Topic: What is the one topic I discuss?
> One Tone: Does the text speak in one voice?
> One Tense: Do I use one tense? If not, do I know why I change them?
> One Target: Where do I go with this text?

Three of the tips are relatively simple to follow: spotting one tab, identifying one topic, and checking tenses require little effort, yet they can make a difference in organization and reader appeal. The tips on tone and target are more subjective matters, but in the long run, they can become part of how a student reflects on writing.

These processes materialized in classroom and take-home assignments. Although the 90-minute session format did not allow for much in-class writing, all projects were discussed in the classes, either in small groups or by the whole class. A few sessions, however, experimented with group writing in class. An activity of this type was done in the Spring 1997 semester. One of the three groups had the sessions in a regular classroom, whereas the other two in the Arizona Room. The task aimed to provide students with practice in writing unified paragraphs based on topic sentence prompts that they were required to discuss. In the traditional classroom, pairs and small groups of students negotiated content and development and then produced subsequent drafts. The individual paragraphs were collected, with the full text typed up for next class for revision. In the Arizona Room, the *GroupSystems* software allowed for pairs to work concurrently on individual paragraphs, by using the Group Writer tool of the facility.

Group Writer is one of several options of *GroupSystems* that facilitates negotiation. Originally developed for conducting business meetings, it promotes dynamic and effective meetings. Divided into the facilitator's server and the participants' workstations, the system connects anonymous users who can work individually or in small groups, responding to questions and participating in other tasks. Responses are typed in and sent to the server, which collects participant input and displays it for all. They can then be applied for small group face-to-face discussion, a process that lowers anxiety and can result in settling an agenda more efficiently than by using traditional methods only. As I had used this network tool earlier in Language Practice and other courses, I had an opportunity to evaluate its usefulness in education. Especially suitable for such a purpose are the modules of Brainstorming, Categorizer, Questionnaire, Group Dictionary, and Group Writer.

After students learned to use the workstations, I sent them the topic sentences and asked pairs to discuss and write their paragraphs. With five topic sentences sent to the participants, they could choose their own. Once a pair began writing a paragraph, the rest of the group worked on other segments of the text. When a draft paragraph was prepared, it was sent to the server, which in turn channeled the text back to all participants. In this way, everyone was able to contribute to the effort, it was instantaneous, and pairs were also able to comment and change the texts by other pairs.

The key advantage of this type of use of the Group Writing tool is its reliance on teams. Also, the text can be printed when the drafts are sent to the server, which students can take home and work on individually.

Other tasks that relied on cooperation belonged to two types: commenting on students' writing and co-authoring texts by students and by a student and the teacher. The former task gained increasing weight as the syllabus of the course was modified; as the section on Readings will show, a marked emphasis was laid on students' opportunities to read their peers' scripts. The latter was first attempted in the Fall of 1997.

Reflection on peers' text was part of the test given in the Fall 1998 semester. Students were instructed to select one of the portfolios from the previous semester, which were part of the reading set, and discuss a positive feature in it. These reflective scripts showed different foci of attention: styles and opinions, emotions and facts received evaluation, enabling the teacher to assess students' coverage of reading and to incorporate insights in modifying readings for future courses. They also represented cooperation on the receptive pane. The productive aspect of this process was practiced in the other type of cooperation: co-authoring essays.

Writing is often conceived of as a solitary activity: the author commits to paper thoughts, ideas, and opinions that seek expression. But writing in academic and other fields often takes place as an effort by more than one person; in fact, writing intended for a public always involves at least two people: the author and the editor.

Working on a theme by sharing an experience will result in growing consciousness of reader-based prose: contributing writers, when such partnerships are formed voluntarily, can provide insights that the solitary writer may not possess. For this purpose, the WRS course introduced the task of co-authoring essays as one of the many options. Beginning with the Spring of 1998, this meant either a script written by two students, or by a student and the teacher. The next section and the one on text types will present more details on this task.

3.3.2.2 Out-of-class activities

In sessions, a variety of individual, pair, and group tasks were applied. This part of the course was complemented by meetings in office hours throughout the five semesters so that individual students' needs and problems receive dedicated attention. But office hours also began to develop into meetings for small groups of students. Participants who could schedule such meetings may have found it useful to supplement their reading, writing and course participation with this opportunity of discussing their own agenda with other students and me.

The Faculty of Arts requires that teachers make five hours of contact time available for their students for such meetings every week of a semester. Officially, this was how much I minimally spent in the office. I was also willing to schedule appointments with students in different slots.

Meetings were held in my department office, their times specified in the syllabus and announced in classes. I attempted to do this so that students would feel welcome—my goal with the meetings was to provide a course that ran parallel with the sessions. Especially in the last two semesters, which applied the writing center framework, the meetings came close to establishing such a form of interaction. This seems to have been appreciated by the majority of students; although I did not keep continuous records of their attendance, the office did become a busy meeting point by the middle of each semester. Of the many types of activity that took place in these consultations, I will present two types: one involving the optional task of co-authoring essays, the other the application of technology.

As discussed in the previous section, students were asked to consider writing at least one text with their peers and another with me. Several students chose to write pieces with their peers, and quite a few with me, too. The rationale for the task was to provide an authentic information-gap writing task for both the students and the tutor: by cooperating on developing a text initiated by either party, they may learn about each other and about each other's writing strategies as well.

One of the students who found time to participate in this project was Polgár Judit in the Fall 1998 WRS course. When Judit came to one of the office hour consultations, she told me she wanted to write an essay with me. I asked here whether she would be interested in drafting a narrative essay on her name, which I could complement with a similar draft. We agreed we would try and swap texts when done. This was how the parallel type of the co-authored essay was developed. (See the essay in Appendix H.)

Another way students exploited the time in the office was technological. As they were requested to submit most drafts typed or printed, students with no computer literacy were helped either by their peers or me. Introducing them to the operating system of the computer and the use and functions of the word processor served a practical purpose. Others came to search for materials on the internet or to explore concordancing programs.

As we shall see in the next section on text types, the research paper requirement included the compilation of a reference section, one that contained all sources cited in the main text in a standard form. The WRS course was the first at the ED of JPU that required the use of either the standard of the American Psychological Association or the Modern Language Association. Writing a References or Works Cited section is no easy undertaking for the novice writer—in fact, even academics, I realized as editor of books and *Novelty*, with a long publication experience tend to ignore or be unaware of this requirement.

Several worksheets and activities were designed to help students with this task. Multiple copies of the publication manuals of the APA and MLA were ordered for the library. The updated version of the department's thesis writing guidelines (Horváth, Nikolov, & Turner, 1997) incorporated sections and illustrative examples on the issue. To help students even more, I designed a set of two simple computer programs to generate APA- and MLA-type reference lists (Horváth, 1999c, 1999d).

Finally, office hours set aside time for students to familiarize themselves with the spelling and grammar check modules of the word processor.

3.3.3 Text types

The tasks and activities described earlier aimed to serve the purpose of raising awareness of the importance of the clarity and concreteness of writing, both in personal and academic writing. Course themes were sequenced in such a way that about the first two-thirds provided a warm-up period in which students would familiarize with concepts and develop their personal strategies and schedules. Although the optional text types in each of the five semesters showed variation, a number of them remained constant elements. The warm-up was promoted by class discussions of personal attitudes to reading and writing, workshops, peer reviews and debates as well as by the development of each student's own journal or portfolio. After this phase, the academic writing module focused on writing the type of text that was to become the primary type of course assignments in the future: the research paper. In this section, I will present and analyze samples of these two types of text. (An overview of these task and text types was provided in Horváth, 1995a, 1995b, 1996a, 1998e, and 1998f.)

3.3.3.1 Personal writing

As the majority of entry-level English majors do not have extensive experience in writing, the WRS aimed to provide opportunities for discovery of tones, contents, approaches and effects, making the process of writing an enjoyable and personally rewarding experience. For this reason, personal writing was given much weight throughout the five semesters. Narratives, descriptions, and arguments were the text organizing elements described, discussed and practiced in the sessions. As the section on Tasks and techniques showed, the application of the metaphorical techniques aimed to illustrate and assist in the processes of theme selection, narrowing down and execution. I encouraged writers to submit multiple versions, and to do that so that not only the teacher but other students were given a chance to read and comment texts. The overwhelming majority of students appeared to be highly motivated to participate in this process. As implied in several course evaluation questionnaires, students maintained an interest in continuous writing, and saw in the teacher's motivation to read several hundred scripts a semester a purpose for their writing and writing development.

Theoretically, the warm-up with personal writing was seen as a phase fundamental for later work. From the students' point of view, this was seen as crucial in (1) establishing positive attitudes to writing, (2) providing practice in designing, planning and drafting clear and concrete texts, (3) helping them develop effective relationships so they had trust and willingness to share scripts, and (4) motivating students to want to revise so they had practice in text-, paragraph-, and sentence-level revision. From the teacher's point of view, the four outcomes were equally relevant, but they were complemented by a reader's curiosity of these students' ideas expressed in the scripts.

Dozens of personal text types were designed over the past semesters to achieve these goals. Of these, I will present the ones that proved most effective, explaining the underlying pedagogical and rhetorical considerations. Each of these text types was presented as an option for students' portfolios, but students were free to choose among them, as well as formulate their own themes and purposes.

3.3.3.1.1 Reflective essay based on a quote

Salamon and Zalotay's 1996 collection of quotes served as a source of content and inspiration. Students would select one or two quotes they liked or disliked and, after introducing the text and its author, provide a personal reflection and opinion. This type of writing represented one of the few choices that focused on argumentation. Its advantage may be seen in two skills. First, students who chose this option experienced both scanning and skimming, providing them with an authentic reading goal. Second, students were able to focus on their own meaning and opinions.

3.3.3.1.2 Descriptive essay on student's dictionary (variation: on the thesaurus)

This text, and the corresponding task, aimed to build skills in narrative and descriptive writing, focusing on relevant learning content. In introducing students' dictionaries, the theme would enable writers to share their views on a writer's tool that the authors knew well, and even to get to know them better.

3.3.3.1.3 Expository, argumentative, and personal narrative or descriptive essays based on the theme selection table

The proficiency testing of writing skills at the ED of JPU was conducted by using a theme selection table that incorporated over a hundred themes (see the 1999 example in Appendix I). As most students participating in the WRS courses were to take the Filter test that included such a component, I distributed previous tests so that students could familiarize themeselves with the instrument and the evaluation scheme, locate themes they wanted to write about, and even generate their own tables. Part of the course, then, aimed to help students in preparing for this high-stakes exam (Horváth, 1996b; Szabó, 1996).

3.3.3.1.4 An essay on any theme, but with two introductions and two conclusions

Sessions that dealt with the issues of presenting a theme effectively in introductions and arriving at a closure were supplemented by the task of writing a text on a topic chosen by the student, and then writing one more introductory and one more concluding paragraph. This option was particularly well received by students, as they were given an opportunity to experiment with different approaches—with those that were practiced in the sessions, and with their own techniques.

3.3.3.1.5 The miniature essay

This text type was first introduced in the Fall of 1997. My main purpose was to edit and publish a book of short essays by students for students. I aimed to help students focus on economy of expression: the total number of words set at 100, it invited authors to describe a concrete event, phenomenon, idea in concrete terms, using specific vocabulary. Students' response was not overwhelming, but I did receive over thirty such essays, which will be incorporated in a new writing textbook (*Take-off*, in preparation). One advantage of producing such a text is that it requires observation, a skill transferable to other text types, both personal and academic.

3.3.3.1.6 Completion of a task in any writing textbook

Writing coursebooks were part of the reading element of the courses, but their relative value diminished over time. However, students were shown, and encouraged to consult, a large number of such material so they could address their own needs in their own time and in consultations. For this purpose, one of the text types was to either complete a drill or a composition task in any of the textbook the students found relevant. The advantage of the option was seen in its contribution to students' learning strategies. Scanning and skimming also played a role in the completion of this task. Besides, familiarity with a freely chosen book can be regarded as a potentially effective mix of free voluntary reading and conscious learning.

3.3.3.1.7 Essay on tape

As we have seen earlier, the WRS course placed a premium on process, such as in the multiple-draft setup and portfolio assessment. Another approach to process writing is to continually develop tasks and texts for multiple purposes. This was implemented first in the Spring 1997 semester. The section on Tasks and techniques introduced the teacher's metaphors. Students were also encouraged to work out their own. This took place in the sessions in which participants discussed how writing about a personal learning experience outside school may help people discover an ability that they may transfer to other fields. Each student chose to write a personal reflective narrative on such an event. The scripts were so powerful and let readers (including me) learn so much about the students that I decided to combine this text with a revision technique: recording the essay on audio tape.

The follow-up task invited students to choose one of three drafts they had submitted before: the learning essay, a script based on a theme selection table, or the one about their own essay metaphors. After reading their scripts and the teacher's commentary, they were required to make any revisions they deemed necessary. Following this phase, students had to read out their own scripts and record them on tape. The rationale was that the aural experience may make students aware of other potential needs for change. By listening to a text, we may realize an unintended sentence fragment, an awkward term that "does not sound good," and other features that can and should be revised. Students expressed overwhelming support for the task, even though producing the tapes posed technical problems to many.

3.3.3.2 Academic writing

Personal essay writing may not be a type of discourse required by many university courses, yet its importance was validated in the past semesters. Students continued to be motivated to share their opinions and discoveries concerning a wide range of themes and fields. In collecting scripts for inclusion in their portfolios, their first books in English, they reflected on the work they had done, the interaction they had with other students and the teacher. They did so after reading authentic texts and revising their syntax, vocabulary and focus.

But the final output of the course was not so much text production based on personal experience and opinion but on observation and analysis of an academically relevant subject. The research paper requirement was thus conceived, for the first time in the history of writing courses at the ED of JPU, in the Fall of 1996, to provide a transition between personal and academic writing, and between the course and the rest of the university studies.

With language, organization, and revision skills practiced and improved, the next task was to conduct a small-scale authentic research project. The small scale of the project meant that students had to have enough time, about five weeks, to decide on a research question, formulate a plan and produce a first draft. The authenticity of the project was concerned with its coherence within the course; this being a WRS course, its research options had to do with the subject matters of its syllabus.

Four of the five semesters offered several choices for this component; the exception being the first one, when students had to write about one topic: the analysis of newspaper articles published on the day they were born. The task involved the location of a relevant source in an accessible library, the selection of the data based on the research question, and the use of reference material about journalism.

In each of the other courses, at least five options were presented, with the ones listed in Table 6 becoming constant elements by the fall of 1998. As the table shows (the same as what students received as one of the handouts in the course), there were six specific themes with corresponding data and suggested reference material. An open choice was also provided for students who wished to explore other opportunities.

Table 6: Research paper options in the Fall 1998 course

Theme	Data	Reference
An analysis of the content of essay introductions	At least 6 essays in *75 Readings*	Smalzer; Zinsser
A survey of feedback types	Interviews and questionnaires	Grundy and Li; Zinsser
An analysis of one aspect of a newspaper published on the day you were born	articles in a daily paper	Bell; Reah
Observing the Writing and Research Skills classroom	observation, interviews and questionnaires	syllabus
Concrete language in students' portfolio scripts	scripts, comments	Raimes; Zinsser
A survey of English majors' interpretation of plagiarism	questionnaires	Pennycock; Horváth
Open Option (to be negotiated with HJ)		

The research paper length requirement varied between 2,000 and 1,200 words—initially it was longer but was reduced in later courses. Students were to follow a standard structure: Introduction, Method, Results and Discussion, and Conclusion (IMRSC), supplemented by a list of their references either in APA or MLA format. In each course, relevant aspects of the department guidelines were to be followed.

As can be seen from this brief description of the research paper, the task and the text type aimed to plug the gap between the WRS course and future courses. The reference conventions it introduced and explained were practiced extensively. Separate task sheets and class forums were used to initiate students to academic modes of delivery.

What this component inherited from the earlier phase of the course was a continued emphasis on clarity and simplicity of language and ideas, and the multiple-draft process approach. Although the course officially ended by the time students submitted their first drafts, the majority welcomed the opportunity of revision and continued to submit second, third, and, in rare instances, fourth versions.

In terms of research design and data applied, the courses aimed to introduce students to two basic types: presenting quantitative and qualitative results. This component of the course functioned as initiation into basic decisions researchers have to make when they embark on a project.

3.3.3.2.1 Analysis of newspaper content

As Chapter 4 will show, this was the single most popular choice throughout the five semesters. An option in each course, this task invited students to consider a specific aspect of the newspaper issue published when they were born. Applying the IMRDC structure, students were to present the results of their library research based on the analysis of text that was relevant to their chosen focus.

3.3.3.2.2 Analysis of peers' writing

As students were writers and readers in the WRS courses, and as reading peers' texts was a priority, this task made it possible for students to explore others' portfolios when they had been finalized. This option allowed students to familiarize even more with various types of writing, so that their repertoire of approaches and strategies may be enriched.

3.3.3.2.3 Surveys among students and teachers

Issues like validity and reliability of the research effort were highlighted in those sessions that aimed to provide help for students who chose to conduct questionnaire and interview surveys among students and teachers. The option was incorporated in the research paper task pool so as to enable students to gather information relevant to their studies and to experience the need for searching for and presenting such information based on a reliable study.

3.3.4 Readings

Research has shown that one factor that greatly contributes to proficiency in writing is the amount of reading successful writers do (Krashen, 1984). This can include reading activities for pleasure, as in free voluntary reading, or reading dedicated to a specific learning goal. The complementary processes of reconstructing meaning in reading and constructing meaning in writing continued to be central in the WRS courses in the past three years. Most syllabuses comprised reading materials of three types: coursebooks selected because they appeared to contain well-designed texts and tasks, publication manuals, and authentic essays and studies.

A total of thirty-eight titles were employed in the past three years. As we have seen in earlier sections, some of these became the basis of classroom pair and group activities, with a number of them also featured in various tests.

One innovation in this regard was the increasing emphasis accorded to students' writing. As early as the Fall 1996 semester, an essay by Schubert Gábor (1996) was featured in the syllabus; the essay, the first student contribution published in *Novelty*, was included to introduce students to the idea of analyzing a course syllabus, and to serve as a possible model for student writing in which the voice of the author was clear, supported by the semi-fictional nature of the experience described in the narrative part of the text.

Schubert's article was the first of many student scripts used in WRS courses. Not all of these appeared in the Readings lists—as the course developed, students themselves began to share their own essays as well, which was facilitated by the course folders placed in the department library.

To provide an overview of the types of texts used as reading materials in the courses, I have prepared the following table (Table 7). It structures the readings according to the three types and presents them chronologically. Note that I had omitted years of publication to economize on space for titles where such information is not necessary for identification—I indicate readings by authors, editors, or titles. The full publication information is provided in the References.

Table 7: The three types of reading materials in the five semesters

Semester	Coursebooks	Manuals	Authentic essays and studies
Fall 1996 7 titles	Hubbard; Marius & Weiner; McCrimmon	Nikolov & Turner	*JPU essays;* Schubert; peers' scripts
Spring 1997 10 titles	Arnaudet & Mary; Berry (1994); Gray & Melis; Hubbard; Marius & Weiner	Nikolov & Turner	Horváth (1996b); Hurtt & Boylan; Pinker; Schubert
Fall 1997 10 titles	Gray & Melis; Hubbard; Strunk & White; students' own choice	APA; Gibaldi	Hurtt & Boylan; Kurdi & Horváth; *Novelty* (1997); Zinsser (1988)
Spring 1998 11 titles	Smalzer; students' own choice	APA; Gibaldi; Horváth, Nikolov & Turner	Eco; Geresdi; portfolios; Salamon & Zalotay; Schubert; Zinsser (1988)
Fall 1998 15 titles	Smalzer	Gibaldi; Horváth, Nikolov & Turner	Babarci; Bacskay Demeter; Földesi Hurtt & Boylan Grundy & Li Horváth (1998b); Rácz *Research papers* Vadon; Zinsser (1998)

The courses also introduced students to the use of an important writer's tool, the thesaurus, which the majority of students had never used before.

Besides these resources, several other texts were reviewed in designing courses and made available for interested students in office hours. These extra materials included one of the first descriptions of plain English for educational purposes, by Gowers (1953). Study-skills handbooks such as those by Smith and Smith (1990) and Sotiriou (1984) complemented composition texts from the U.S. and U.K.: Arnold and Harmer (1978), Clouse (1986), Crews (1987), Elsbree, Bracher and Alitzer (1977), Evans (1998), Gere (1985), Gould, DiYanni and Smith (1989), Hall (1988), Hamp-Lyons and Heasley (1987), Hansen (1987), Hult (1986), Legett, Mead, Kramer and Beal (1988), Leki (1989), Madden and Rholck (1997), McMahan and Day (1984), Rackham and

Bertagnolli (1988), Raimes (1983b; 1996), Rankin (1972), Schenk (1988), and Weiner (1973).

Of the textbooks published in Hungary, the most recently used were Csomay and Szerdahelyi's (1997) process-syllabus resource for advanced students and Kiszely's (1998) collection of texts and tasks. Besides, the chapters of my own developing course material, *Take-off* (Horváth, in preparation), were made available to students.

3.3.5 Feedback and evaluation

As Chapter 1 discussed, the literature is divided on what constitutes best practice in providing feedback on student writing. The pedagogical validity of expert feedback, however, has not been questioned. In this section, I will explain my approaches to and practice of feedback provision, which will be followed up by the evaluation of students' participation and other work in recent courses.

3.3.5.1 Feedback techniques

Students received continuous assessment on their work in prompt feedback to their writing. In the Fall 1996 semester, this took the form of conducting a dialog in each student's journal. In the other courses, students were required to submit a portfolio of their scripts, a selection of their drafts. Written feedback was supplemented by discussions in sessions and in office-hour consultations.

By writing on an author's script, the editor-teacher becomes a co-author of the text. This relationship necessitates professionally sound and useful comments, which are clear, specific and which lead the student to want to reflect on the advice. Commentary has to give an authentic view of the reader's impression of the content and overall quality of the text. To achieve these aims, my practice involved two types of comment: (1) handwritten notes in the margins focusing on sentence- and paragraph-level issues and notes at the end summarizing overall impressions, and (2) typed reviews.

As far as the portfolio scripts are concerned, most comments were written in hand on the scripts. Besides, I applied the technique that was later also discussed in Grundy and Li (1998): to save the original script from becoming an illegible mixture of main text by writer and subtext by reader, I used Post-It notes. These could be flipped over or removed when revising. Another traditional technique was using pencils: this even allowed students to erase comments they did not agree with.

In all of my work on feedback, I aimed to focus on positive features so that students were able to build on them while addressing weaknesses. Also, by reading my feedback, students became co-authors of my writing, which I considered another authentic text type.

When a portfolio was presented for evaluation, I had seen most scripts at least once in their earlier versions. The purpose of the typed feedback was to provide one more reading material to students that was special in its detail, and hopefully useful. As for the comments on research papers, the feedback followed the categories of evaluation. Before students received the options for the research paper task, they learned about the evaluation criteria. Extensive comments were given on all first drafts. Tables 8 and 9 show the version used in 1997 and 1998, respectively.

Table 8: The evaluation categories of the research paper in 1997

Category	Max.	Score
Identification of field and research question in Introduction	1	
Clarity of research in Method section	1	
Clarity and appropriateness of reporting findings in Results and Discussion section, including appropriateness of citations	2	
Relevance of implications in Conclusion	1	
Appropriateness of form of References	1	
Syntax, spelling, punctuation and vocabulary	3	
Double-spaced and stapled	1	

Table 9: The evaluation categories of the research paper in 1998

Criteria	Max mark	Your mark
Clarity of introduction and research question	1	
Clarity and appropriateness of method	1	
Clarity and relevance of results and discussion	2	
Relevance of conclusion	1	
Clarity and appropriateness of language	3	
Citations and Works Cited	2	

The combination of spoken and the two types of written comments, although a most time consuming effort, appeared to contribute to students' willingness to participate in classes and to revise. Also, by setting an example with my own motivation to respond promptly, with most written feedback provided within days of receiving a script, I aimed to communicate my own motivation to students. Further empirical research, however, is necessary in the field: both qualitative and quantitative data should be gathered to establish factors that most effectively contribute to improved writing. Also, as will be explained in the next chapter, the use of teachers' typed feedback can be extended to form part of the annotation of a learner script, thus facilitating a systematic study of the nature, typology, validity, and reliability of such commentary.

3.3.5.2 Evaluation

In any course of study, teachers assess the progress and achievement of the students. The basis of the assessment is some sample of skills or knowledge covered in the course, whereas the results can serve evaluative and diagnostic purposes. Informal assessment of participation was done on a continuous basis in all of the WRS courses; this was based on data on students' attendance and holistic assessment of their work in the sessions. In awarding a final grade to students, the achievement was tested in the texts student submitted.

A major decision to be made in such assessment is concerned with its basis; the two distinct types are holistic and analytic. I chose the latter option to enhance the transparency of the course: as all scripts were scored by me, students had to know the constituent categories I evaluated.

In the past five semesters, four types of assessment categories were applied in the courses. As Figure 14 illustrates, their relative weight changed across the five courses, with participation being modified least, and the test the most. The Spring 1998 course was an example of the four categories receiving equal weight.

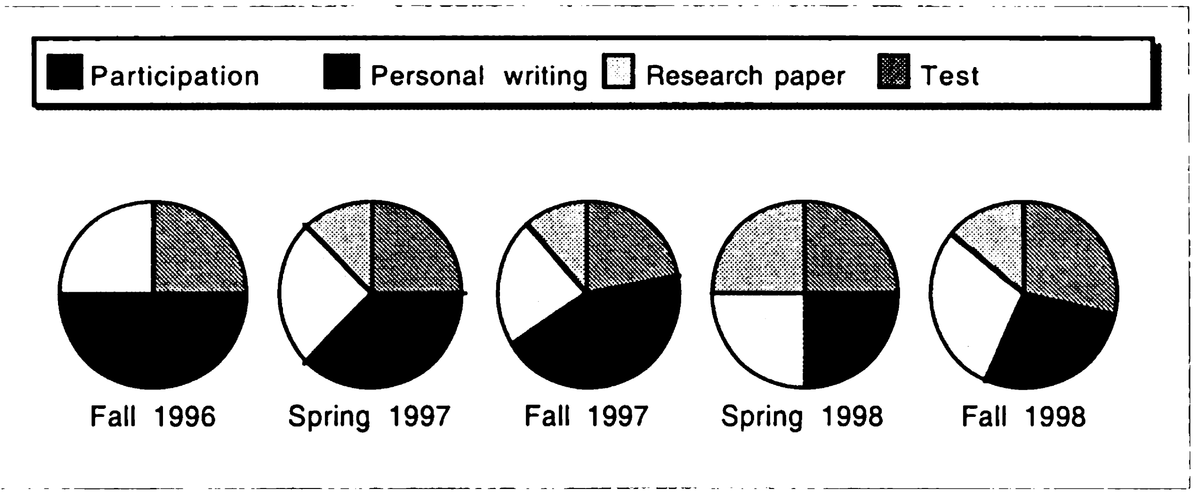


Figure 14: The relative weight of assessment categories across the five courses

Each of the four types of activity assessed provided information on students' achievement, and thus were integral elements of the final picture that emerged.

Student involvement in achievement testing is also an option. This was elicited twice in the course, with the most recent project occurring in the fall of 1998. It involved the evaluation of the portfolio, which was assigned a maximum mark of ten, on a holistic scale. The requirements I considered in assigning a grade to a particular collection were the following: regularity of writing during the course, the number of scripts (a minimum of five), the application of readings and of the five T tips, and evidence of effective revision.

Overall, on the basis of the information gathered from the participating students, it appears that not only were students successful in their portfolio projects, but the majority also regarded the evaluation as fair. As a tutor of these students, I was glad to see a marked agreement between the two scores. But to be able to add to the reliability of this part of the study, further investigation is necessary. In discussing preliminary findings of this project, several students suggested that in reporting a score to me, some participants may not have given the true score of their work. In a future project, student research assistants may need to elicit this information. Also, interviewing students could provide insights into the process of students' self-evaluation.

Another aspect of assessment is how levels of performance are compared. Most university courses appear to apply criterion referencing: in the syllabus the teacher specifies a grading scheme with percentages representing levels. This may be a valid approach in lecture courses involving a large number of students. However, in seminar courses norm referencing may be more valid from the point of view of the construct of seminar work. Comparing students' results with each other informs teachers of the work they have been able to do. Also, fine-tuning level setting may carry higher face validity.

For these two reasons, I applied norm referencing throughout the five semesters, deciding on required levels of performance for each of the four passing levels by consulting the graph of final scores.

3.3.6 Students' views

Students' course evaluations have become a regular procedure at the end of terms at JPU. They were introduced in 1995 to provide staff, students and administration with the information students share about each of the courses they completed. Besides this official procedure, several tutors have implemented their own feedback generating practices so that they may receive valuable insight from students into the effectiveness of their courses. As the results of the official evaluations take a longer time to process and tabulate, and recently have not even been released, tutors who need more immediate feedback have experimented with an unofficial yet simpler technique of eliciting student response to their courses. In this section I report the result of one such evaluation project.

Thirty students participated in the procedure of the Spring 1997 course evaluation. The three sections of the course, ANG 1601, 1602 and 1603, had a total of 36 registered students, of whom two had not participated in the last four to six classes. Out of 34 students, 32 were present in the last classes. Data was collected on May 12 and May 13, 1997, on the dates when students were submitting their end-of-term assignments.

My hypothesis was that students would express positive and negative attitudes to the course and that the information I would receive may be useful in planning next semester's syllabus for a slightly modified course.

In each of the three sections, students were asked to participate in the evaluation anonymously in writing. The questionnaire consisted of four categories that students were asked to rate numerically. They were told that they had the option of not completing the questionnaire or not submitting it. I administered and collected the questionnaires. Two students chose not to participate.

Students were asked to rate each of the following four evaluation criteria on a scale of 1 to 7, where 1 represented extremely negative, and 7 extremely positive views:

- ➤ Fairness of evaluation;
- ➤ Assistance from students;
- ➤ Assistance from the tutor;
- ➤ Usefulness of the course.

I identified these criteria as I regarded them as genuine indicators of students' satisfaction levels. Also, I hypothesized that the composite mean figures for fairness, student assistance and tutor assistance would correlate highly with the mean of the usefulness criterion.

After I collected the 30 questionnaires, I analyzed the data by statistical techniques, calculating means and STD figures.

Out of the 30 students who returned the questionnaire, 28 responded to the item on how fair they found the evaluation of their work in the course. In the three sections, students seemed to consider my evaluation fair; two gave the Fairness of Evaluation criterion a value of 4, five students gave it a value of 5, six students a value of 6, and twelve students gave it the top value, 7. Figure 15 presents the distribution of values for the fairness criterion.

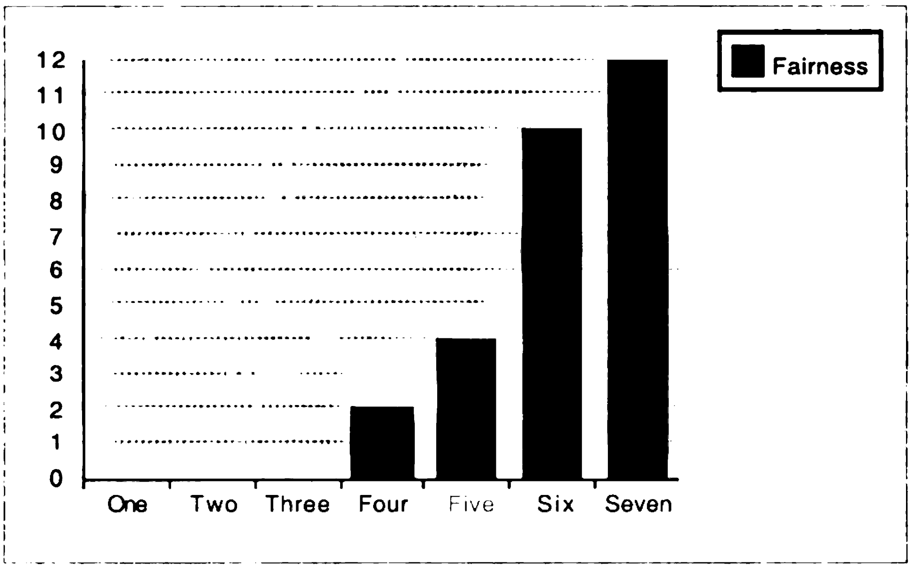


Figure 15: Number of students selecting values for the fairness of evaluation query (N = 28)

The second item asked students to rate how much assistance they received from other students in the group. All 30 students who took the questionnaire answered the question. With three students giving this criterion a value of 3, nine students a value of 4, ten students a value of 5, three a value of 3, and five a value of 7, the assistance students reported they received from others appeared to be somewhat lower than I expected. Figure 16 shows the distribution of values for the Assistance from Students criterion.

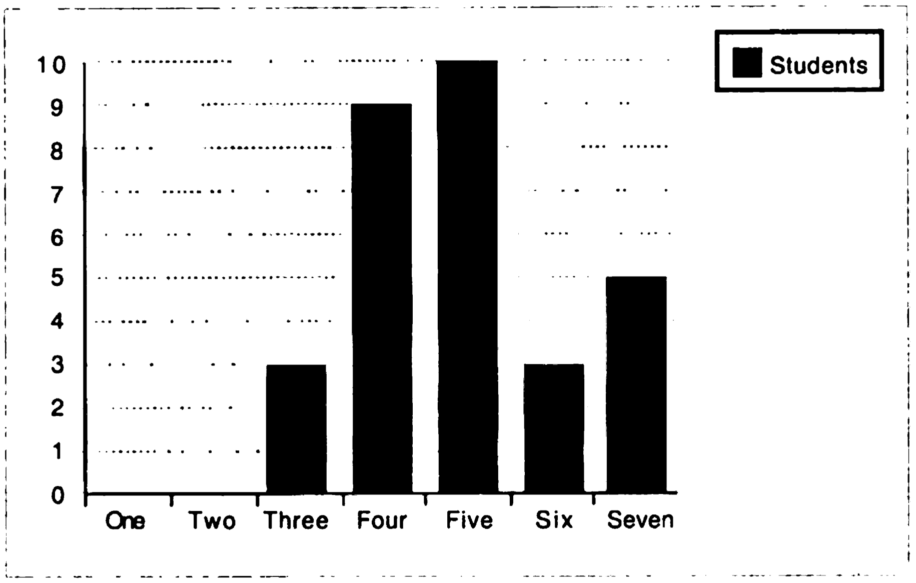


Figure 16: Number of students selecting values for the assistance from students query (N = 30)

The third category was Assistance from the Tutor. All 30 students who returned the questionnaire responded to this query. One student assessed the tutor's assistance by giving it a 3, two by giving it 5, seven by giving it 6, and twenty-one by giving it the top value, 7. Figure 17 demonstrates the distribution of values for the assistance from the tutor criterion.

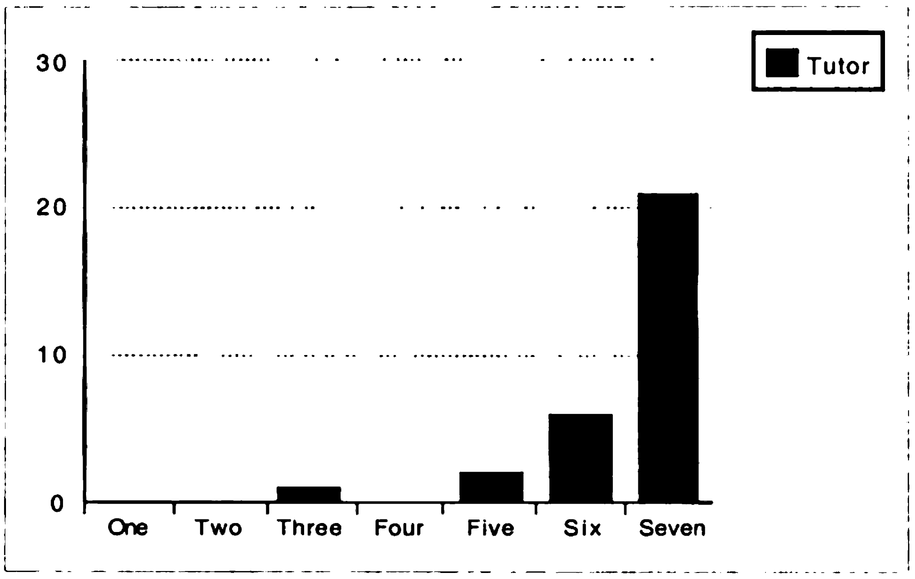


Figure 17: Number of students selecting values for the assistance from the tutor query (N = 30)

The last course evaluation category in the questionnaire invited students to assess the usefulness of the writing course. Again, all the 30 students returned their questionnaires by assigning one value to this category. One student gave it the median value, 4, nine the value of 5, fourteen the value of 6, and six the value of 7. Figure 18 shows the distribution of values for the usefulness criterion.

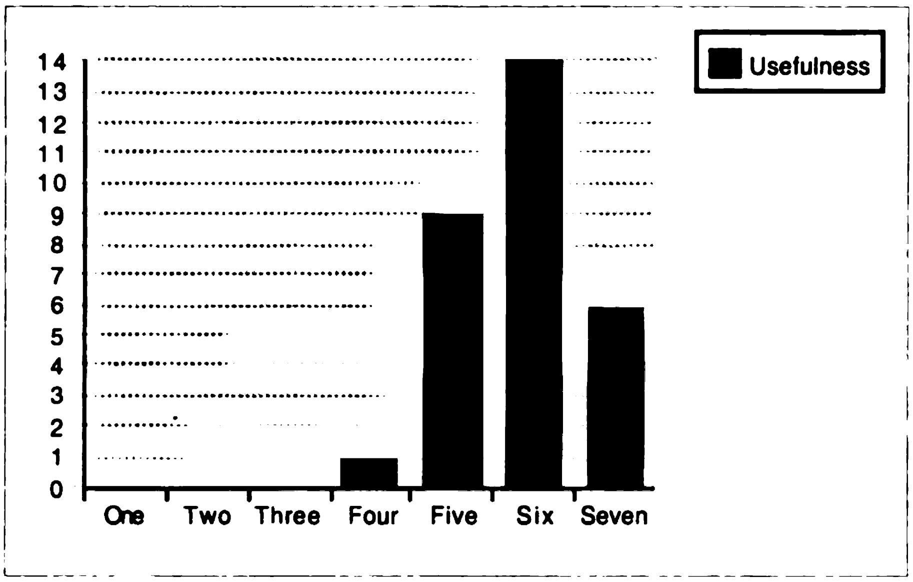


Figure 18: Number of students selecting values for the usefulness of the course query (N=30)

To obtain information on how students' evaluations differed from each other, I calculated the standard deviation (STD) figure as well. An STD can show how similar or different respondents' opinions are by comparing each respondent's rating with the mean. The lower the STD, the more uniform individual responses are; conversely, the higher this value, the more divergent the opinions. Although it is extremely rare that in any group all members would agree on all issues, I regarded the STD of the four criteria as another essential aspect of the reception of the course.

As Figure 19 attests, the most divergent opinions were expressed about the fairness of evaluation (1.79). The other three category STD figures were lower, with the usefulness category STD value being the lowest (0.79), showing that this was the evaluation category that elicited most uniform responses.

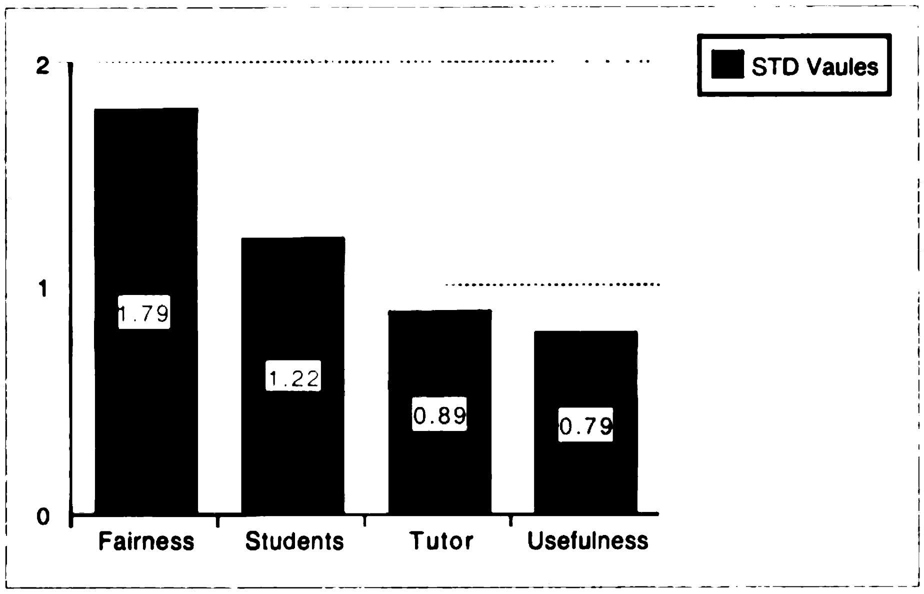


Figure 19: STD values of participants' evaluations of the four criteria

Another way of looking at the results is by calculating the mean figures of the category values. To be able to form an overall image of students' evaluation of these criteria, I conducted this calculation and found the following: The lowest mean was obtained for assistance from students (4,93). While this was the lowest value, it was still in the positive range of the scale. Students ranked the usefulness of the course criterion higher, as the mean figure for that category was 5.83. For the fairness of evaluation and assistance from the tutor categories the mean figures were 6.14 and 6.53, respectively. Figure 20 shows the rating of the four factors.

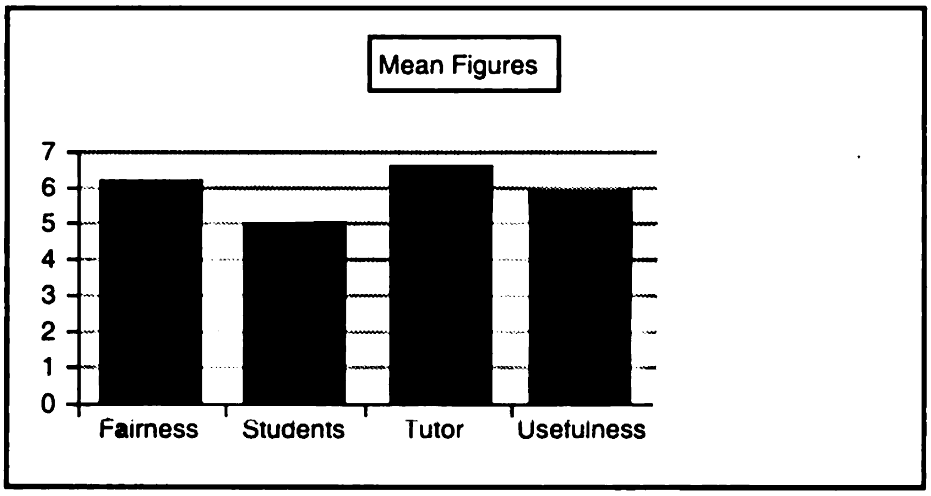


Figure 20: Mean figures of the evaluation of the four criteria

Finally, to assess the reliability of the results, I undertook a comparison analysis by calculating the means of the values assigned to the fairness of evaluation (F), assistance from students (S), and assistance from the tutor (T), and by comparing that with the mean figure for the usefulness of the course category. I hypothesized that the comparison would result in little if any difference between the two values if the results were reliable, but be markedly different if they were not. As Figure 21 reveals, almost no difference was found between the usefulness of the course and the composite of the other three factorial means.

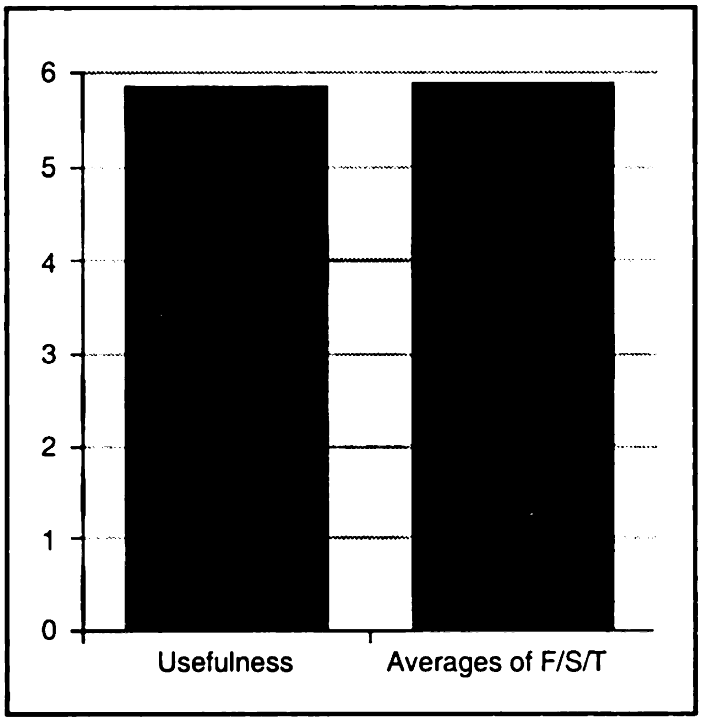


Figure 21: Comparison of the mean score for the usefulness criterion and the averages of the fairness, assistance from students and assistance from the tutor criteria

I obtained valuable information on students' evaluation of the three WRS sections. I hypothesized that students would share their positive and negative opinions in selecting values on the scales for each of the four categories. Most opinions students expressed about these courses were in the positive range of the scale, with only one student assigning one of the categories a slightly negative value (3, in assistance from the tutor).

3.4 Future directions

This chapter has positioned the WRS courses in the ED curriculum to set the context of the processes applied in recent undergraduate courses. Aiming to present an analysis of both qualitative and quantitative data, it has provided details of the curriculum, and of the innovations introduced in the teaching of writing. I reflected on and analyzed the development of the syllabus, classroom procedures, tasks and text types, and feedback techniques. The analysis of students' evaluation testified that the courses have benefited from the shift from a product-oriented approach to writing pedagogy to one that incorporated the practical implications of the process approach, especially in terms of text types and formative assessment.

As a result of conducting writing research and practicing the pedagogy of writing, I have benefited professionally and personally. Students' views and opinions have continued to shape the focus of the course. Their continued interest in participating in voluntary conferences in office hours provided further evidence that a teacher's motivation was a significant factor in maintaining and raising student involvement. Most importantly, their personal descriptive essays and research papers have helped me understand more clearly their views of the world, themselves, and the educational and linguistic issues they identified as essential. In so doing, they have also helped me want to become a better teacher writer.

The opportunities I have had as a writing teacher have gone beyond the classroom and the office hour. Several students have begun to pursue writing activities for their own pleasure and for various purposes. An increasing number of fiction and non-fiction writers have emerged. Over twenty students have published their articles, essays, and reviews in Hungarian and international newsletters and journals. To have been able to motivate and assist them has been another source of satisfaction.

The invigorating effect of reading a first draft, of discussing its merits and problems in class and privately, of reading a revised version of the script: the opportunity to be part of the development of an idea that a student is willing to share with peers and the tutor is among the greatest benefits a writing course can offer. Besides, the use of earlier groups' essay and research paper collections in new courses establishes a link between students, and thus helps maintain an authentic and relevant discourse community.

How this community is being developed throughout university was beyond the scope of this study. This area may be well worth the investigation in future action research, involving a representative sample of courses. Another type of follow-up study could investigate how what is acquired and learned in the WRS courses is applied in students' theses. For wider institutional appeal, and for encouraging cross-institutional cooperation, Hungarian colleges and universities could set up research and pedagogical projects.

Chapter 4

THE JPU CORPUS: PROCESSING PRODUCTS

> ***We need more genre-sensitive studies and more specialized corpora in addition to the larger representative corpora as a basis for analysis.*** (Kennedy, 1998, p. 291)

Introduction

The previous chapter has placed writing pedagogy in the JPU ED core curriculum and described and evaluated the procedures developed in the past semesters, focusing on undergraduate WRS courses. It has applied a balance of quantitative and qualitative data. In this chapter, I aim to provide a detailed description of written learner English by investigating quantitative data, the JPU Corpus. As indicated in Chapter 3, the majority of contributions have come from WRS course participants—the corpus, however, provides evidence of five main types of learner groups. Three of these have been undergraduate pre-service students in the last six years: those attending Language Practice, WRS and miscellaneous elective courses. The remaining two groups of participants have taken part in in-service language education: a few Russian Retraining students, and a larger group of postgraduate students.

A solid set of data was collected between 1992 and 1998, facilitating a quantitative analysis of the language produced. The approach followed in this chapter is based on the corpus linguistic assumption that the performance of a language community has to be investigated to capture probable features of language behavior, whose statistical and pedagogical significance can then be tested and validated.

Why and how the corpus was first conceived will be discussed in section 4.1, which also explains design principles, data input procedures, text types, and the three types of methods used for the empirical study. Section 4.2 then goes on to present the current composition of the main corpus, followed by the specific compositional details of the five subcorpora. After this book, section 4.3 identifies ten hypotheses of this part of the book. Descriptive and contrastive analyses were carried out, involving the full JPU Corpus, its subcorpora, and contrastive analyses based on the results of ICLE investigations.

The chapter then follows up to address the pedagogical uses of the corpus: section 4.4 introduces an application of data-driven learning, whereby students are assisted in submitting their own scripts to analysis. Specific examples will illustrate how this has been done in Language Practice, Elective and WRS courses for group activities and for individual study. The section briefly discusses miscellaneous other applications of the corpus.

I hope this presentation will serve as a valid basis on which to draw conclusions, in section 4.5, on the applications and limitations of a corpus-based study of written learner English—besides, I intend to suggest future directions where such endeavors may lead.

4.1 The development of the corpus

4.1.1 Conditions of and rationale for data collection

Bratislava hosted the 1992 TESOL Summer Institute (with "At the Crossroads" as its slogan), which I was able to attend for part of its duration. A large number of workshops were offered, among them two by Macey Taylor, a leading U.S. practitioner of Computer Assisted Language Learning. Having an interest in the application of word processing techniques in writing as well as in the design and pedagogical application of dedicated CALL software, I joined the courses. In one of them, Taylor introduced the participants to Longman's *Mini Concordancer* software by demonstrating the ease with which it processed small sets of text. It was in that session that terms I had learned earlier as an avid user of the first edition of the *Collins COBUILD Dictionary* materialized in front of me: I generated *concordances* using the *keyword in context* function, studied *co-texts*, and looked at the statistics on *tokens* and *types*. My first hands-on experience with the application made me want to learn more.

I saw in this program and the lexical and syntactic investigations it made possible a wealth of pedagogical applications. Imagining how JPU English majors in the Fall 1992 Language Practice course could benefit from its use, I began to read the literature on corpus linguistics and DDL. Saving my earlier essays and papers as ASCII, I loaded my first own small corpus of my own work and saw, fascinated, features I would not have thought I could see or wanted to see before. But now I could and did. And I was convinced students could and would, too. With two groups in September 1992, I became the first tutor at the ED of JPU to explore the potential of analyzing authentic native speaker (NS) text and non-NS text by computer.

As my primary interest was the analysis of learner English for language education purposes, I proposed to the students in the two groups that they submit their written contributions on computer disk (Bocz & Horváth, 1996). Looking back, the positive response continues to strike me as incredible. After all, those were not the times of wide access to computers—in fact, there were few even in department offices, with the first portable units just arriving. However, students consented, and I made time available for brief practical typing and word processing sessions. From that time on, there have been a growing number of students who have submitted texts on disk, permitting me to save their files onto the hard disk of the computers I used at the time.

The current status of the development of the JPU Corpus may be regarded as satisfactory for a linguistic and language educational study. It is the first to employ a large database of Hungarian learner English for descriptive and analytic purposes, which represent the ultimate rationale for corpus development.

Specifically, collecting students' scripts enables applied linguists to do the following:

- keep a record of students' performance, making longitudinal studies possible;
- submit the collection to theoretically and practically relevant analysis;
- extract linguistic and pedagogical information from the corpus;
- exploit the corpus for language education;
- compare and contrast individual learner corpora;
- compare and contrast learner corpora with L1 collections.

For the first option, a corpus can contain all the scripts students have written, and requires the cooperation of a team. The second, third, and fourth fields can be explored individually, as they have been in the DDL tradition (Johns, 1991a, 1991b; Horváth, 1994a, 1994b, 1995a). The fifth and sixth areas often necessitate team work nationally and internationally (Granger, 1998b).

In the rest of this chapter, I will restrict the investigation to demonstrating what I considered relevant analyses given the individual undertaking of the project.

4.1.2 Corpus design principles

As the presentation of cycles in corpus design (in Chapter 2) has pointed out, when one is attempting to collect texts for principled linguistic study, factors such as purpose, language community, text types, representativeness, encoding, and storage facilities need to be investigated. Preliminary aims and composition requirements may need to be modified in the light of pilot studies that test how representative the sample is.

In my effort, I was led by the following considerations. I envisaged a corpus that would

- be about half a million words;
- represent written English by JPU students in the courses that I taught;
- permit generalizations on student written production;
- incorporate a variety of text types;
- not reveal the identity of any contributor of a specific script to the public;

- be based on such submissions as are voluntarily contributed.

I set the size of the intended corpus at 500,000 words to collect at least half the size of first-generation corpora. Although that target has not yet been reached, the current size is rather close. Also, other learner corpus projects indicate that a smaller size is sufficient (Granger, 1993, 1996, 1998a; Kaszubski, 1997; Mark, 1998). As will be shown shortly, the current size of the JPU Corpus is twice as large as a subcorpus of the ICLE. In terms of the second criterion, all components come from courses I taught between 1992 and 1998. The reason for arguing that this sample may allow for generalizations on other writing by other students at the institution is that the majority of scripts come from students in WRS courses and from those participating in in-service postgraduate education. Combined, these contributors represent the majority of learner population at JPU in the past three academic years.

As for the third criterion referring to text types, a representative sample of different genres has been collected, with corpus linguistic and pedagogical aims in what can be regarded as sufficient balance. None of the students have been asked to allow me to reveal their authorship of any examples to be shown in this chapter—the names that appear in the Acknowledgments cannot be linked to the scripts. Finally, all text samples that appear in the current version of the JPU Corpus are voluntary contributions—most solicited by asking students to sign a permission form. Details on these six considerations will follow in the rest of the section.

4.1.3 Data input

Texts were sought for inclusion in an unnamed collection between 1992 and 1993. Between 1993 and 1995, students were told that their contributions would be incorporated in the Pécs Corpus. The name was changed to JPU Corpus in 1995 so that it more realistically identified the endeavor. The flow chart in Figure 22 illustrates the process of incorporating individual learner texts. As the chart illustrates, two types of data were recorded: the script itself saved to computer disk, and the information on the student and the course of origin for the script.

From the figure it is perhaps evident that the JPU Corpus is a semi-annotated collection: it has author, gender, year, course, and genre information tagged to it, but it does not take advantage of any of the robust tagging techniques available today. There is a disadvantage and an advantage to this lack. Without word class or grammatical tags, the corpus cannot in its present form allow for fully reliable, automatic processing and information output. However, in the vein of Fillmore's (1992) claim on the "armchair" linguist and the corpus linguist having to exist in the same body, this limitation may be viewed as a potential advantage: the partial reliance on intuition, based on pedagogical practice and observations, and on linguistic evidence may make up for the present lack of the tagging component. (However, as Labov, 1996, suggested, when intuition and introspection are employed, the following principles should be observed: the consensus, the experimenter, the clear case, and the validity principles.)

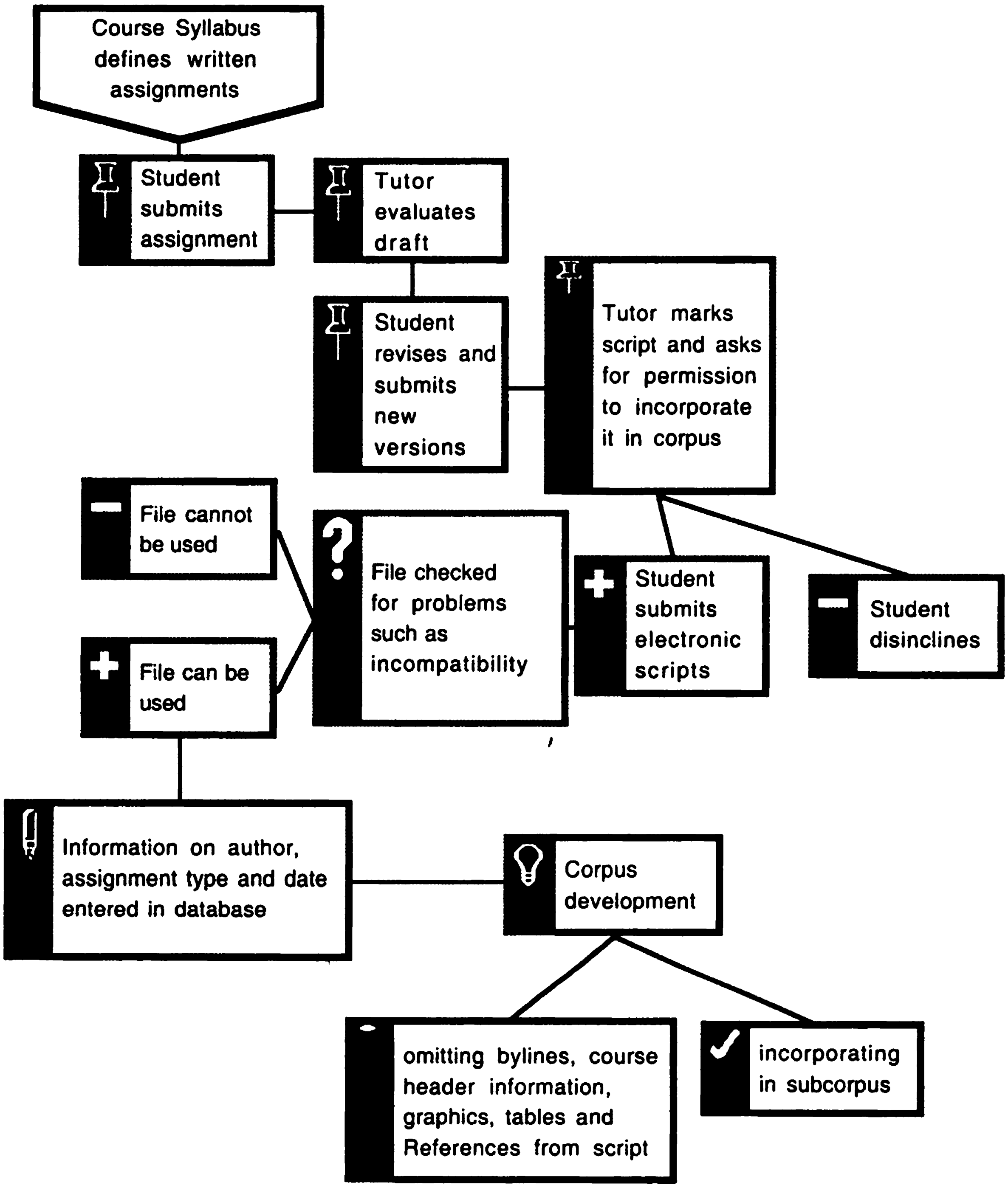

Course Syllabus defines written assignments
Student submits assignment
Tutor evaluates draft
Student revises and submits new versions
Tutor marks script and asks for permission to incorporate it in corpus
File cannot be used
File checked for problems such as incompatibility
Student submits electronic scripts
Student disinclines
File can be used
Information on author, assignment type and date entered in database
Corpus development
omitting bylines, course header information, graphics, tables and References from script
incorporating in subcorpus

Figure 22: The process of data input

4.1.4 Seeking permission

At the end of courses students were asked to submit the electronic copy of their essays and research papers. I explained to them my purposes, saying that I aimed to analyze their scripts in relation to other students' contributions. In most instances, students were willing to do so.

In the early stages of the development, only oral permissions were sought. In each instance, submissions were sought after the students had received their grades for the course, so that their decisions may not affect evaluation. By letting me save a copy of a script, the students would consent to the act of incorporating the text in the collection. To enhance the reliability of the process, however, I introduced an authorization form in 1996, which was the time of bulk additions to a relatively small learner corpus. A copy of such a form appears in Appendix J.

Not only was the change a result of making the project fully legal, but it was also based on a socialization consideration. I made the move to ask for official permits so as to contribute to the sense of professional community among students and teachers. Familiarizing onself with the concept and practice of copyright was seen as an additional element of language education at the department. Further, the decision was supplemented by suggesting to students that they submit their printed assignments with a © notice. For one thing, not many students knew what exactly the symbol represented and how this related to academic standards of free expression and of text ownership. Some may even have found the proposal superfluous, thinking that the teacher was making too much fuss. But when one considers the problems of copyright infringement in many subcultures, and specifically the occurrence of plagiarism at Hungarian universities, my approach arguably promoted an authentic experience of being initiated into the scholarly community.

4.1.5 Clean text policy

Data capture was done relatively fast. As Figure 22 has shown, students who were willing to contribute to the corpus were asked to submit scripts on computer diskette. In the beginning, both standard size DOS-compatible disks were used, with the transition to 3.5-inch disks exclusively taking place in 1994. When I was handed a disk, I checked it for any problems such as viral infection and incompatibility. The former issue had been safely eliminated by early 1995 when I began to store scripts on an Apple Macintosh computer. Fortunately, viruses cannot engage their malicious operation across platforms; this was a crucial technical issue for the sustained development of the corpus. It also meant that once I had saved a student's file to the hard disk, no lurking viral programs were transmitted to the student's disk either.

However, incompatibility of proprietary word processing software code in the text file was harder to overcome. For the first two years, before word processing software became widely available in educational institutions, I had had to exclude texts that could not be converted properly. More recently, I have been using shareware programs for any text file that my word processing programs could not extract.

When the technicalities are taken care of, real work on text preparation for corpus inclusion can begin. This process serves three functions: recording contributor data in the corpus database, ensuring that the content of the file is compatible with the concordancing application, and editing the text for authenticity.

The first function presents no hurdles: I have used the computer's file system hierarchy to maintain the database. Figure 23 illustrates, via a screen shot of a window on the Macintosh desktop, the file hierarchy concept.

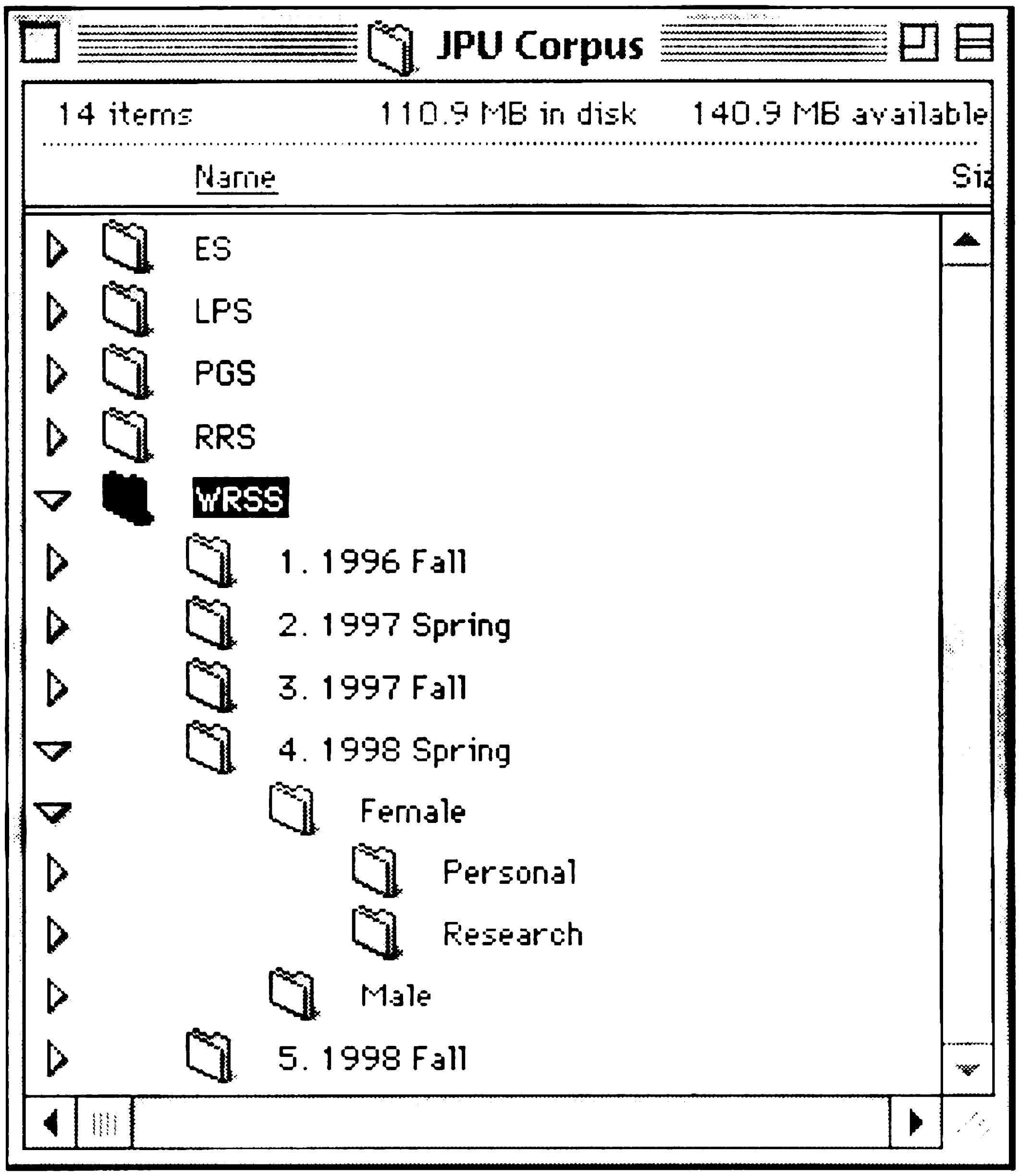
As will be detailed in section 4.2.1, the corpus is divided into five subcorpora. The screen shot shows one of the folders highlighted, and the contained folders listed, storing files by semester, then by gender, and finally by text type.

The second function is also relatively straightforward once the file is saved locally: *Conc*, as most other concordancing software, can process data saved as ASCII, or text-only files.

The third function, however, is much more time-consuming, given the short experience most students have had with word processing. Much as one of the requirements for most submissions in the past five semesters has been for students to check their texts for typing and spelling errors, some have continued to submit files that needed careful editing. Deciding whether an error was a typing or a spelling mistake has not always been easy. Yet, I have worked out a procedure that may be regarded as reliable.

I decided to take action and change text only if the error was clearly a typing mistake. This meant changing words like "langauge" to "language" or "teh" to "the." That is, transposed characters were always amended. The clean text policy of the JPU Corpus project meant that no other mistakes were corrected so that the data would remain as authentic as possible (a similar approach was employed for text handling in the ICLE project; see Granger, 1998a).

Finally, texts were edited by removing any author identification from the header, such as bylines, and components such as course codes, any graphics, tables and references.



```
JPU Corpus
14 items          110.9 MB in disk     140.9 MB available
          Name                                          Si
▷  ES
▷  LPS
▷  PGS
▷  RRS
▽  WRSS
▷      1. 1996 Fall
▷      2. 1997 Spring
▷      3. 1997 Fall
▽      4. 1998 Spring
▽          Female
▷              Personal
▷              Research
▷          Male
▷      5. 1998 Fall
```

Figure 23: A window of part of the corpus in the Macintosh file system

4.1.6 Text types

Two major types of text are represented in the corpus, which also account for most of the assignments that students submit at the ED: essays and research papers. In this analysis, an essay is defined as any non-fiction submission to a university course for which the method of gathering data is not strictly specified. Within this group, there are further divisions: personal reflective essays, narrative-based and descriptive writing, and a combination of essay and research paper for a content course. In this third type of text, the writer typically consults reference materials but the presentation of the ideas does not follow a standard research pattern.

In contrast, a research paper is a submission for which the writer has to follow academic standards: identifying a field of study and a research question, presenting the method for answering the question, and putting forth its data and analysis to answer it. It is typically supplemented by reference materials to be collected on the basis of the readings section of a syllabus and as the writer's own initiative. In most regards, the research paper can be viewed as a small-scale thesis, or as one of the body chapters in a thesis. (Figure 24 illustrates the curricular composition of the scripts.)

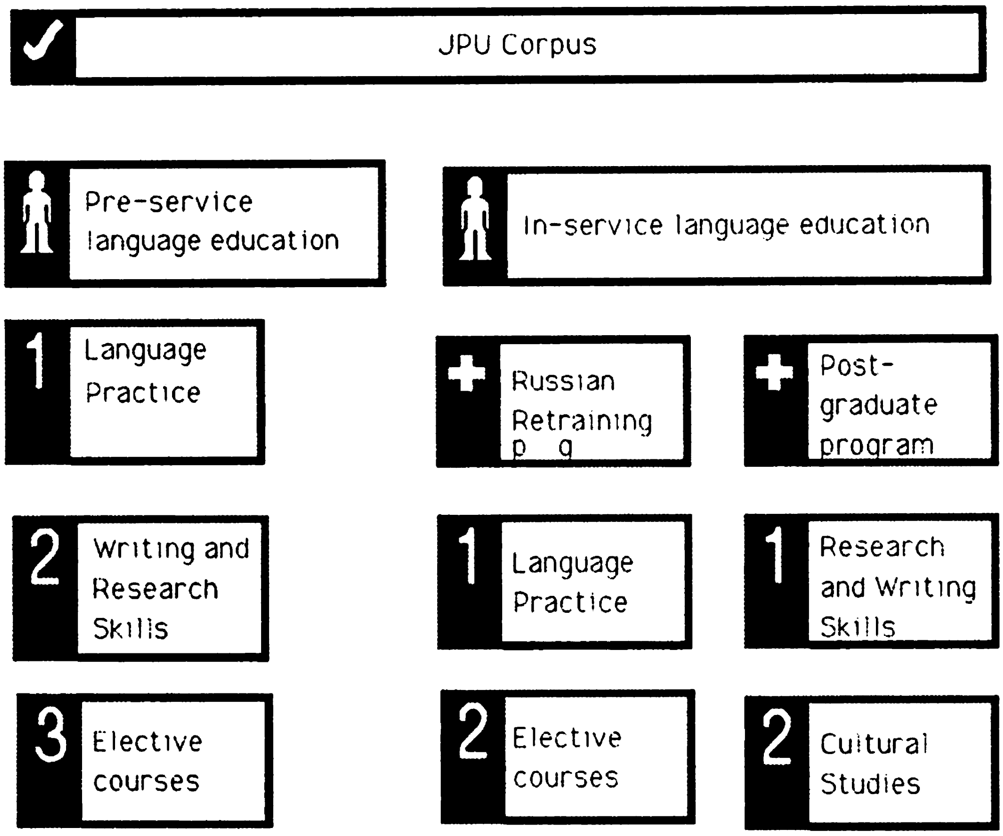

JPU Corpus
Pre-service language education
In-service language education
1 Language Practice
Russian Retraining
Post-graduate program
2 Writing and Research Skills
1 Language Practice
1 Research and Writing Skills
3 Elective courses
2 Elective courses
2 Cultural Studies

Figure 24: Curricular and course origin divisions of the scripts

4.1.7 Procedures applied

Hypothesis forming in corpus linguistics follows the cycles of the development of the data set itself. Phases of design specifications may be preceded by hypothesis building, but as work progresses, the linguist will gather insights into the composition of the data, and thus the research questions may change. This has been true of the JPU Corpus project as well. My overall hypothesis has always been that by submitting the data to detailed analyses, one can describe the written English of Hungarian university students as a social dialect of English. Rather than taking a dubious stance of underestimating the values of this language and calling it "Hunglish," I have preferred to construe of this interlanguage (Selinker, 1992) as a valid component of world Englishes (Phillipson, 1992a, 1992b; Quirk & Widdowson, 1985). With recent interest shifting to peripheral studies, corpus linguistics has the advantage of providing the evidence on how such languages are structured and what phenomena they exhibit.

In terms of actual procedures used, I have employed two corpus linguistic techniques, two statistical models, and one language educational approach. As for the corpus linguistic techniques, I distinguished between operations on the complete corpus and on various samples. Data processing was carried out via *Con* 1.7: I opened each of the 332 files in the program, sorted the text alphabetically right to the keyword, and saved the KWIC concordance. This stage provided the raw material for concordance analyses and other techniques. To collect information on the composition of the full corpus, I also saved the alphabetical index that the program can provide, together with information on tokens and types. The same procedure was performed on each of the subcorpora.

A limitation of *Conc* is that it cannot automatically produce a frequency list. This, however, posed no difficulty as a database application, *FileMaker Pro*, has this tool. I opened each of the alphabetical index files for the main corpus and the subcorpora and sorted the contents by the frequency of words.

With these operations done, I printed KWIC and frequency list pages to study their content. Online searches were also carried out.

It was at this stage that common corpus linguistic techniques were performed: KWIC analyses, calculating normalized frequencies, comparing most frequent word forms, and drawing up the statistics for lemmatized words in the main corpus and the subcorpora.

These steps were taken to have a large set of materials on which to test hypotheses—of which those that required statistical verification were loaded into a spreadsheet program to obtain significance information via the chi-squared test and ANOVA. The former model, used for observations in the JPU Corpus and between the PGS and the WRSS, makes no assumption about the normal distribution of data and can be applied for frequency comparisons based on different size corpora (McEnery & Wilson, 1996, p. 70). The latter is suitable for studying the effect of variables across three or more populations, using interval scales (Koster, 1996).

Finally, the third type of method used for this study comprised the production and evaluation of classroom worksheets that have been piloted in earlier courses, as well as the development of material to illustrate how such an approach can be exploited for guiding individual study.

With these methodological considerations, we can now move onto the specific details of the current state of the corpus.

4.2 The JPU Corpus

4.2.1 The current composition of the corpus

The 1999 version of the JPU Corpus contained 412,280 words in 332 scripts, each from a different student. This volume represents over twice the size of the individual national subcorpora contained in the ICLE, making the JPU Corpus one of the largest written learner English data sets. Earlier, some ninety students were represented by multiple scripts, but extra contributions were removed so as to avoid bias. Two courses of action were taken for this purpose. When a student submitted multiple versions of a script, the last one was incorporated. Alternatively, for students who participated in more than one course, the scripts for which they received the higher marks were included. As Figure 25 shows, each text is stored in one of five subcorpora, according to type of course the authors attended.

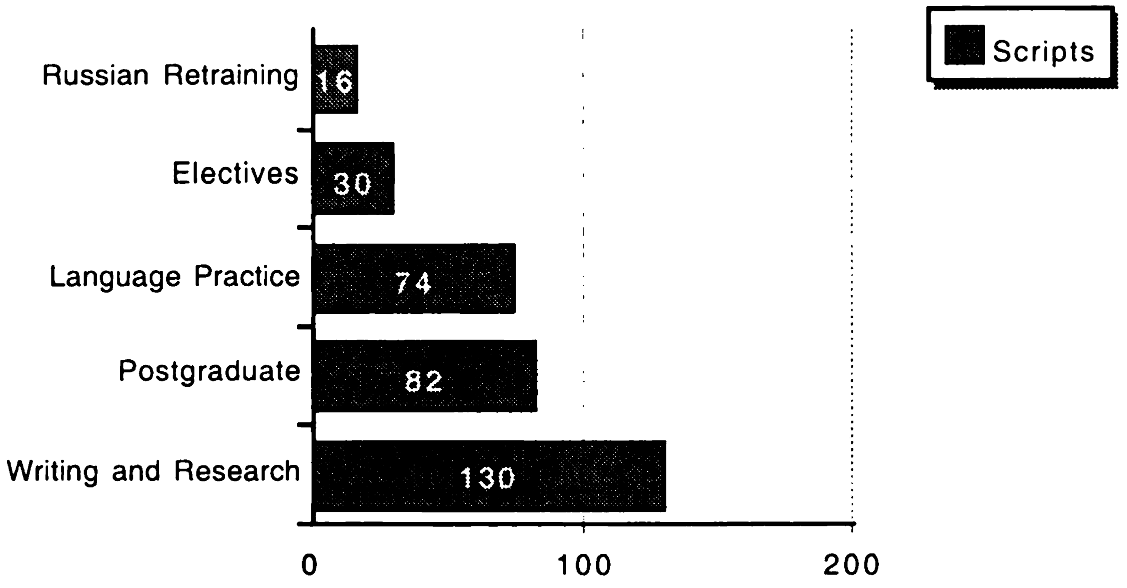


Figure 25: The number of scripts contained in the five subcorpora

The Russian Retraining subcorpus (RRS) is the smallest unit, with two types of text: Language Practice personal descriptive and argumentative essays by twelve female students and one male, and semi-research paper essays by three female students of elective courses. I consider this component of the corpus valuable even though its size is small: it records the performance of students who participated in a study program that has been discontinued since.

Somewhat larger than the RRS is the Electives subcorpus (ES), comprising 30 scripts. Most were submitted by females: 21 academic essays on CALL, Indian Literature, the application of the internet in language learning, and DDL. The other nine texts, by male students, are of similar types.

A significantly more representative sample is structured in the Language Practice subcorpus (LPS): the texts are personal descriptive, narrative or argumentative essays. This is also the subcorpus with the most significant male student population: 31 male and 43 female authors are represented.

The two most sizable subcorpora are the Postgraduate (PGS) and the Writing and Research Skills (WRSS) collections. In terms of number of scripts and types of words, the WRSS is more representative, with its 130 texts (by 106 female and 24 male contributors). The text types represented by the WRSS are personal essays (23), with the rest of the collection (107 scripts) made up by research papers. (For more details on types of research paper in the subcorpus, see the sections on hypotheses 9 and 10.) However, in terms of tokens, the PGS is larger: with 82 students (68 female, 14 male) contributing to this subcorpus, it is made up by 123,459 words. The relative significance of each of the five subcorpora is demonstrated in Figure 26: it charts the JPU Corpus by the number of scripts in them.

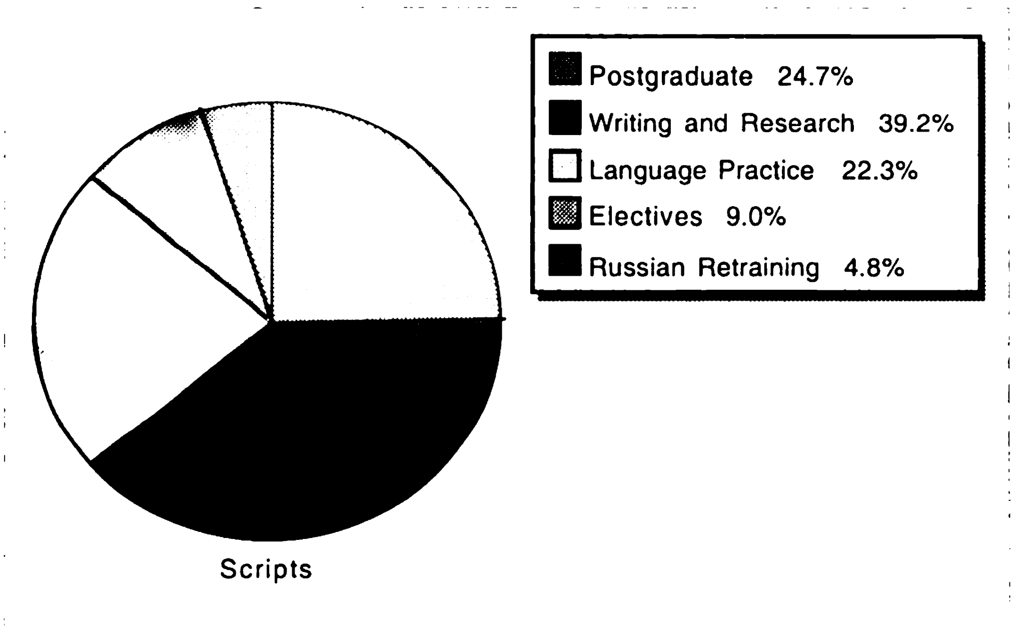


Figure 26: Distribution of texts in the subcorpora according to number of scripts

Figure 27 also illustrates the distribution of texts in the five subcorpora, this time calculated by tokens of words in them.

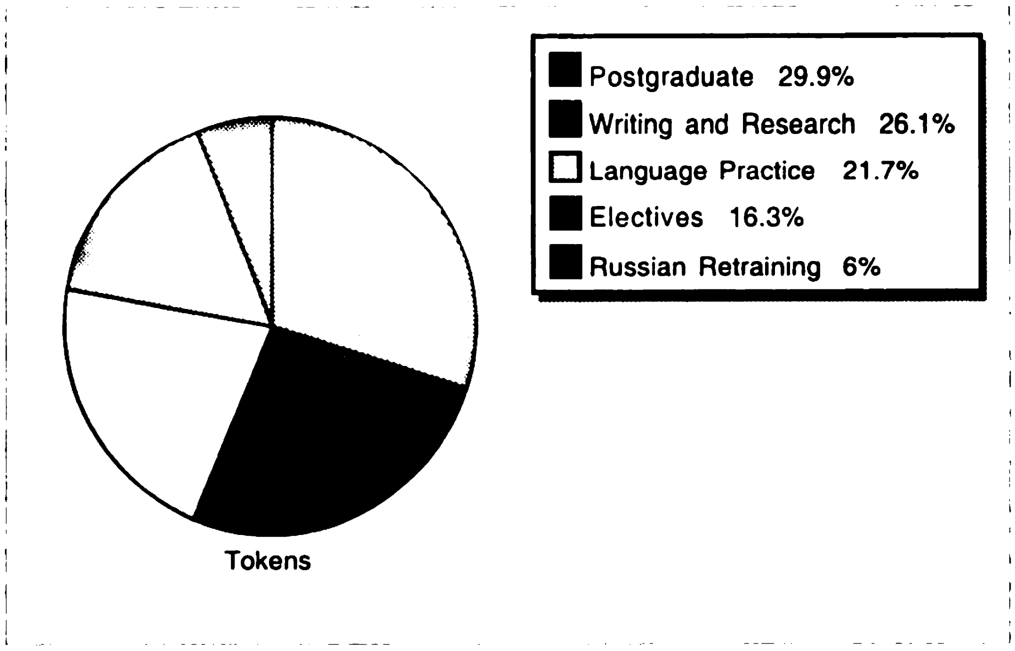


Figure 27: Distribution of the texts according to number of tokens in the subcorpora

Altogether, the five subcorpora are made up by 17,535 types of words (that is, distinct graphic word forms), a relatively high number. The PGS is ranked number one for both number of tokens and ratio (see Table 10); it already appears that the papers in that subcorpus contain relatively more homogeneous texts than the second largest, the WRSS.

Table 10: Statistics of scripts in the five subcorpora

Subcorpus	Tokens	Types	Ratio	Scripts
PGS	123,459	6,933	17.80	82
WRSS	107,752	8,666	12.43	130
LPS	89,396	8,260	10.82	74
ES	67,061	7,710	8.69	30
RRS	24,612	4,006	6.14	16

Table 11 shows gender representation in the JPU Corpus. As can be seen, over three-fourths of the students are women: 76.2% as opposed to 23.8% men. This appears to be in line with the general demography of the ED of JPU.

Table 11: Gender representation in the JPU Corpus

Subcorpus	Female	Male
PGG	68	14
WRSS	106	24
LPS	43	31
ES	21	9
RRS	15	1
Total	253	79

To provide a preliminary overview of the content of the corpus, Tables 12 and 13 list the most frequent words and the most frequent content words. In studying Table 13, one has to note that raw word forms do not provide sufficient detail on word class—as a result, tables listing raw frequency data represent only the basis of further analysis (cf. Kennedy, 1998, p. 97). For reliable lexical analysis, lemmatization has to take place.

Table 12: The 20 most frequent words in the JPU Corpus

Rank	Word	Frequency
1	the	32231
2	of	14757
3	to	11602
4	and	10835
5	in	9102
6	a	8526
7	is	6409
8	it	4149
9	that	4123
10	I	3695
11	are	3265
12	they	3195
13	not	3041
14	for	2981
15	be	2916
16	this	2759
17	with	2755
18	as	2732
19	was	2566
20	on	2521

Table 13: The 20 most frequent content words in the JPU Corpus

Rank	Word	Frequency
1	students	2164
2	writing	1552
3	essay	945
4	language	898
5	people	773
6	english	747
7	different	746
8	time	729
9	use	680
10	words	660
11	like*	651
12	paper	606
13	introduction	587
14	make	554
15	write	553
16	work	549
17	way	539
18	used	531
19	text	524
20	reading	506

Note: *Like* appears as a preposition and subordinating conjunction 371 times.

The twenty most frequent words total 15,494, or 3.76% of all tokens. In terms of content words, we can see that several words in Table 13 belong to the semantic field of writing; this indicates a marked use of such vocabulary, not surprisingly, in the WRSS and PGS (see also sub-sections on these two subcorpora later).

As attested by all corpus analyses, the most frequent word forms are represented by function words—this can be seen in Table 14, which lists the ten most frequently occurring types across the five subcorpora. The number one position of the definite article and the frequency of prepositions are not surprising; what is worth noting is the high rank of the first person singular pronoun in the PGS and the WRSS; the sections that describe the composition of those units will provide a reason for this occurrence.

Table 14: The ten most frequent words in the five subcorpora

Rank	Postgraduate	Writing	Language P	Electives	Russian
1	the (9615)	the (8912)	the (6640)	the (5352)	the (1679)
2	of (4357)	of (3980)	of (3178)	of (2561)	and (770)
3	to (3636)	to (2941)	to (2461)	to (1868)	to (691)
4	and (3297)	and (2835)	and (2174)	and (1758)	of (691)
5	in (2758)	in (2323)	a (1908)	in (1569)	in (569)
6	a (2596)	a (2165)	in (1852)	a (1389)	a (468)
7	is (1930)	is (1318)	is (1615)	is (1127)	is (418)
8	I (1761)	that (1165)	that (1051)	it (681)	his (273)
9	are (1180)	I (1127)	it (1018)	that (648)	he (272)
10	it (1124)	it (1110)	are (835)	be (549)	they (244)

In developing the JPU Corpus, one of my early aims was to test the accuracy of the use of the definite article, the most frequent word in any corpus; also, the word that appears to be least taught, relative to its importance and frequency. However, the sheer size of the corpus has made it a daunting task to conduct such an analysis on the present untagged corpus—still, as will be shown later in this chapter, such information was obtained on the RRS.

Over seven thousand of the word forms (7,522) occur only once in the JPU Corpus. As Table 15 illustrates, the most significant representation of such lexis can be seen in the Russian Retraining subcorpus—this adds support to the observation that the shorter the text, the most likely it is to be made up by such word forms.

Table 15: Rank order of the five subcorpora according to ratio of hapax legomena

Subcorpus	Number of hapax legomena	Ratio of hapax legomena
RRS	2070	8.41%
ES	3580	5.33%
LPS	3814	4.26%
WRSS	4163	3.86%
PGS	2854	2.31%

This tendency can be further highlighted by comparing the rank order of the subcorpora according to ratio of hapax legomena and number of tokens: see Table 16.

Table 16: Contrasting the rank orders of the subcorpora by hapax legomena (HL) and tokens (T)

Subcorpus	Rank by HL	Rank by T
RRS	1	5
ES	2	4
LPS	3	3
WRSS	4	2
PGS	5	1

Although my study cannot be concerned with comparing the lexis of the JPU Corpus with any large non-specialized NS corpus, I submitted the frequency list of the JPU Corpus to a rank-order analysis, based on Kennedy's (1998, pp. 98-99) table of the top fifty words in six corpora. Of these, I selected the rank-order lists for the Birmingham (Bank of English) Corpus, the Brown Corpus, and the LOB Corpus. Then I rank ordered the words that are common to the Birmingham and the JPU Corpus, to identify the word forms whose ranks showed similarity and differences. The two parts of Table 17 list the rank orders for the four corpora.

Table 17, Part 1: The rank orders of the most frequent words in three large corpora and the JPU Corpus: Ranking from 1 to 25 (Based on Kennedy, 1998, p. 98)

Word	Birmingham	Brown	LOB	JPU
the	1	1	1	1
of	2	2	2	2
and	3	3	3	4
to	4	4	4	3
a	5	5	5	6
in	6	6	6	5
that	7	7	7	9
I	8	20	17	10
it	9	12	10	8
was	10	9	9	19
is	11	8	8	7
he	12	10	12	40
for	13	11	11	14
you	14	33	32	58
on	15	16	16	20
with	16	13	14	17
as	17	14	13	18
be	18	17	15	15
had	19	22	21	47
but	20	25	24	26
they	21	30	33	12
at	22	18	19	34
his	23	15	18	44
have	24	28	26	25
not	25	23	23	13

Table 17, Part 2: The rank orders of the most frequent words in three large corpora and the JPU Corpus: Ranking from 26 to 50 (Based on Kennedy, 1998, pp. 98-99)

Word	Birmingham	Brown	LOB	JPU
this	26	21	22	16
are	27	24	27	11
or	28	27	31	22
by	29	19	20	33
we	30	41	40	42
she	31	37	30	70
from	32	26	25	29
one	33	32	38	28
all	34	36	39	45
there	35	38	36	36
her	36	35	29	93
were	37	34	35	39
which	38	31	28	27
an	39	29	34	31
so	40	52	46	65
what	41	54	58	49
their	42	40	41	24
if	43	50	45	60
would	44	39	43	74
about	45	57	54	30
no	46	49	47	84
said	47	53	48	317
up	48	55	52	81
when	49	45	44	54
been	50	43	37	107

After this introduction of major features of the corpus, I will present specific information on each of the five units. (The most frequent word forms occurring at least 100 times in the JPUC appear in Appendix K.)

4.2.2 The five subcorpora

4.2.2.1 The pre-service data

4.2.2.1.1 ES

The ES represents the smallest of the three undergraduate subcorpora. Made up by the scripts of thirty students, this subcorpus represents the early stages of corpus development, with 30 scripts collected between 1993 and 1996. In terms of content, as Table 18 indicates, educational issues dominate the majority of these texts: keywords such as *language*, *students*, *teachers* and *learners* feature in the most frequent words of the ES.

Table 18: Word forms occurring 100 times or more in the ES

the	(5352)	but	(284)	when	(146)
of	(2561)	from	(267)	time	(142)
to	(1868)	an	(265)	them	(141)
and	(1758)	which	(263)	if	(137)
in	(1596)	have	(257)	most	(137)
a	(1389)	one	(257)	some	(130)
is	(1127)	language	(252)	only	(129)
it	(681)	his	(240)	may	(127)
that	(648)	students	(240)	two	(127)
be	(549)	he	(215)	who	(122)
as	(489)	at	(209)	would	(121)
for	(489)	i	(201)	teacher	(119)
not	(481)	there	(188)	teachers	(119)
with	(480)	were	(184)	britain	(118)
are	(473)	had	(183)	her	(116)
this	(439)	more	(181)	out	(116)
was	(431)	its	(180)	learners	(112)
can	(411)	also	(168)	english	(107)
on	(393)	all	(165)	she	(106)
they	(380)	these	(154)	you	(104)
their	(349)	other	(152)	into	(101)
by	(327)	about	(149)	been	(100)
or	(316)	has	(148)		

4.2.2.1.2 LPS

The LPS is the second largest of the three undergraduate subcorpora. The 74 students contributing to it submitted their scripts over the longest period, compared with those in the other subcorpora: scripts from as early as 1992 and as late as 1996 appear in the LPS. Two types of learner English are included: scripts written as part of Language Practice courses in the core curriculum, and those by students in an advanced Language Practice course offered in the Spring of 1996. The table listing the most frequent word forms (Table 19) indicates a more heterogeneous topic base than that of the ES.

Table 19: Word forms occurring 100 times or more in the LPSS

the	(6640)
of	(3178)
to	(2461)
and	(2174)
a	(1908)
in	(1852)
is	(1615)
that	(1051)
it	(1018)
are	(835)
not	(779)
they	(749)
for	(742)
be	(729)
as	(655)
this	(645)
with	(620)
on	(558)
can	(516)
have	(515)
or	(474)
but	(457)
i	(457)
their	(457)
was	(457)
one	(377)
we	(361)
he	(357)
from	(351)
people	(346)
which	(337)
about	(331)
by	(327)
you	(327)
at	(323)
more	(318)
there	(302)
all	(300)
an	(300)
other	(281)
she	(281)
students	(278)
them	(278)
his	(264)
so	(255)
only	(250)
these	(248)
some	(244)
who	(243)
group	(242)
also	(241)
has	(238)
do	(237)
if	(233)
were	(231)
will	(226)
would	(223)
because	(205)
what	(205)
most	(200)
time	(196)
her	(194)
course	(187)
had	(187)
like	(187)
when	(181)
very	(164)
dallas	(163)
our	(163)
life	(162)
no	(161)
even	(156)
student	(155)
way	(154)
could	(153)
well	(151)
coffee	(147)
than	(143)
its	(142)
up	(141)
my	(138)
use	(131)
many	(130)
should	(129)
been	(128)
first	(126)
out	(126)
different	(125)
two	(124)
language	(123)
how	(119)
any	(118)
always	(115)
get	(114)
news	(114)
cards	(112)
much	(111)
new	(109)
children	(107)
good	(107)
important	(107)
those	(105)
world	(104)
every	(103)
such	(102)
your	(102)
family	(101)
make	(101)
after	(100)

4.2.2.1.3 WRSS

The largest undergraduate subcorpus contains texts by 130 students, mostly first-year JPU English majors who participated in the WRS courses between 1996 and 1998. It is by far the most representative of the student population, in terms of the number of students and number of texts. As Table 20 shows, the most frequent word forms include vocabulary related to the writing experience itself, as the majority of students are represented by their research paper submissions, rather than by their personal descriptive or narrative essays. The high frequency of the first person singular pronoun (1,127) indicates that the majority of authors of texts employed an active, rather than a passive, frame in discussing their themes.

Table 20: Word forms occurring 100 times or more in the WRSS

the	(8912)	at	(439)	three	(194)
of	(3980)	can	(427)	well	(190)
to	(2941)	but	(416)	so	(188)
and	(2835)	them	(397)	did	(187)
in	(2323)	by	(381)	words	(187)
a	(2165)	only	(373)	people	(184)
is	(1318)	these	(355)	student	(182)
that	(1165)	essays	(349)	english	(181)
i	(1127)	how	(341)	use	(181)
it	(1110)	had	(339)	we	(181)
they	(880)	more	(335)	paper	(179)
was	(848)	there	(332)	articles	(177)
on	(764)	first	(324)	time	(176)
not	(761)	my	(320)	different	(168)
this	(695)	two	(316)	research	(168)
for	(687)	all	(301)	hungarian	(165)
students	(682)	he	(270)	between	(164)
with	(650)	also	(261)	has	(163)
as	(638)	who	(252)	his	(163)
be	(620)	most	(248)	she	(163)
are	(584)	other	(248)	write	(162)
one	(584)	out	(234)	than	(161)
essay	(555)	because	(229)	introduction	(157)
or	(547)	news	(225)	used	(156)
about	(523)	what	(222)	would	(156)
their	(503)	some	(214)	make	(154)
writing	(497)	you	(208)	topic	(151)
were	(489)	when	(205)	word	(147)
an	(455)	do	(200)	up	(146)
from	(449)	if	(197)	verbs	(146)
which	(444)	could	(196)	year	(144)
have	(443)	will	(195)	many	(143)

course	(142)	same	(124)	text	(110)
paragraph	(140)	four	(122)	events	(109)
work	(139)	no	(120)	sentences	(108)
university	(138)	made	(119)	after	(107)
those	(130)	question	(119)	conclusion	(106)
found	(129)	article	(115)	each	(106)
should	(128)	its	(115)	papers	(105)
find	(127)	second	(115)	results	(104)
like	(127)	day	(114)	such	(103)
into	(126)	any	(111)	last	(102)
number	(125)	five	(111)	reader	(101)
page	(125)	way	(111)	sentence	(101)
information	(124)	writer	(111)	according	(100)

4.2.2.2 The in-service data

4.2.2.2.1 RRS

The 16 Russian Retrainee students took Language Practice and elective courses in 1995-1996. They represent the last groups of such students in JPU ED—and in the country. The low number of types resulted in no content words in the word forms occurring at least 100 times (see Table 21).

Table 21: Word forms occurring 100 times or more in the RRS

the	(1680)	their	(203)	i	(139)
and	(770)	for	(199)	be	(138)
to	(691)	that	(199)	which	(126)
of	(680)	not	(195)	or	(122)
in	(569)	are	(191)	from	(119)
a	(468)	as	(191)	on	(111)
is	(418)	this	(181)	about	(110)
his	(273)	was	(168)	have	(106)
he	(272)	with	(168)	by	(105)
they	(244)	can	(167)		
it	(210)	but	(144)		

4.2.2.2.2 PGS

Finally, the PGS is the largest in terms of tokens of all the five subcorpora. The 82 students submitting research papers participated in postgraduate Research and Writing Skills and Cultural Studies courses in 1997 and 1998. There is considerable variety of vocabulary in the most frequent types of words, as attested by Table 22, the majority of content words indicating a preference of such themes as writing and language education.

Table 22: Word forms occurring 100 times or more in the PGS

the	(9615)	how	(341)	you	(177)
of	(4357)	some	(341)	items	(176)
to	(3636)	english	(335)	up	(176)
and	(3297)	words	(335)	very	(176)
in	(2758)	text	(332)	part	(175)
a	(2596)	different	(331)	time	(175)
is	(1930)	were	(328)	paragraph	(170)
i	(1761)	book	(322)	writer	(170)
are	(1180)	write	(316)	between	(168)
it	(1124)	reading	(311)	well	(168)
that	(1059)	when	(309)	new	(167)
they	(942)	two	(299)	way	(165)
be	(869)	other	(289)	skills	(164)
for	(864)	tasks	(282)	unit	(164)
writing	(857)	only	(281)	written	(164)
with	(837)	will	(254)	papers	(158)
not	(823)	sentences	(253)	conclusion	(157)
this	(795)	do	(252)	could	(157)
my	(769)	had	(250)	same	(157)
as	(757)	essay	(246)	teaching	(157)
or	(731)	use	(246)	information	(155)
on	(694)	if	(244)	after	(153)
can	(692)	because	(243)	vocabulary	(153)
students	(680)	all	(241)	did	(151)
was	(662)	style	(234)	into	(150)
have	(660)	most	(232)	three	(149)
which	(584)	also	(230)	found	(147)
their	(569)	sentence	(224)	listening	(147)
about	(498)	topic	(223)	any	(143)
but	(479)	so	(218)	no	(143)
an	(476)	has	(215)	our	(143)
them	(451)	work	(205)	know	(140)
from	(448)	used	(203)	questions	(140)
we	(448)	research	(202)	word	(140)
one	(443)	texts	(202)	he	(139)
these	(443)	make	(197)	read	(139)
there	(442)	grammar	(196)	teacher	(139)
introduction	(396)	should	(196)	question	(135)
more	(393)	out	(193)	content	(134)
paper	(379)	exercises	(189)	ideas	(134)
language	(375)	reader	(188)	essays	(132)
by	(373)	task	(186)	its	(132)
what	(371)	find	(185)	results	(131)
first	(359)	like	(183)	too	(131)
at	(352)	each	(181)	activities	(129)

category	(129)
good	(128)
help	(128)
letter	(128)
subject	(128)
main	(126)
who	(126)
important	(125)
would	(125)
people	(124)
your	(124)
me	(123)
method	(122)
then	(122)
his	(121)
than	(119)
thesis	(119)
intermediate	(118)
second	(118)
present	(117)
teachers	(117)
number	(116)
aim	(115)
many	(115)
does	(114)
order	(114)
she	(114)
analysis	(113)
attention	(113)
knowledge	(113)
type	(113)
get	(112)
form	(111)
give	(111)
school	(111)
given	(110)
speaking	(110)
story	(110)
readers	(109)
categories	(108)
four	(108)
another	(107)
parts	(107)
according	(106)
course	(105)
general	(105)
level	(105)
following	(104)
types	(104)
both	(103)
may	(103)
exercise	(101)
made	(101)
mistakes	(101)
point	(101)
been	(100)
paragraphs	(100)
units	(100)
where	(100)

4.3 Analysis of the corpus

As a teacher of the students represented in the JPU Corpus, I was one of the readers of the scripts submitted. Receiving multiple drafts from the writers, I formed a view of the content and quality of these submissions, many of which I read repeatedly as students had made revisions. Studying and evaluating the scripts also gave me an insight into student writing that would inform the hypotheses tested on the basis of the corpus. A host of lexical and syntactic investigations are made possible by the corpus—the ones offered here represent what I regarded as pedagogically most relevant inquiries that I was able to conduct with the software available. As no tagging was performed on the data, these studies are restricted to those types of analysis that can be performed by reference to frequency information and lexical patterns identified in KWIC concordances.

4.3.1 Hypothesis 1

The first hypothesis suggested that the RRS will contain a number of inaccurate uses of the definite article. There were three reasons for this hypothesis. Of the sixteen students in the RR Language Practice course, fifteen used to be teachers of Russian, a language that employs no article. As Hungarian definite article usage is governed differently, there was a probability of marked negative transfer in the second foreign language, English. The second reason for such a hypothesis was that these students had a relatively short time to prepare for their university education, a condition that may not have been counterbalanced by the increased amount of Language Development tuition they received. The third reason was that this group did not enjoy the opportunity of submitting multiple drafts, and thus the chance of error was assumed to be higher.

To test this hypothesis, I generated the KWIC concordance of the RRS and analyzed the citations for the definite article. Of the 1,680 occurrences, 103 were eliminated, as these were quotations from various sources. Of the remaining 1577 citations, I hypothesized erroneous uses would reach about 100, or about every sixth in one hundred co-texts.

The hypothesis was rejected: the total number of errors in the use of the definite article was 43. The result shows the effectiveness of students' learning and applying the rules of using the definite article. However, as the study could not investigate the frequency of error of not using a definite article, the finding cannot be regarded as conclusive. Also, as co-texts cannot always provide sufficient information on context, the 1,577 samples may have contained more erroneous uses, which could not be determined on the basis of subjective parsing.

4.3.2 Hypothesis 2

In the second phase of the analysis of the corpus, transitional phrases were investigated—involving the full corpus and by comparing observations in the PGS and the WRSS. Hypothesis 2 was concerned with the distribution of frequencies of the following discourse markers: ***but***, ***however***, ***still***, ***yet***, ***on the other hand***, and ***nevertheless***. In particular, the hypothesis suggested that of these phrases the coordinating conjunction *but* would be most frequent, and that in sentence initial position this frequency would remain. For emphatic change of focus or argument, students were encouraged to employ the conjunction, besides opting for what appear to be more preferred choices in academic writing, such as ***however***, and *on the other hand*. Rather than using such wordy transitions as "however, it should be pointed out that" or "yet, it is important to note that," the simplicity of *but* often results in effective sign posting, as confirmed by such authors as Strunk and White (1979) and Zinsser (1998).

To test the hypothesis, the frequencies of these phrases were tabulated for the main corpus and the three subcorpora. The results are shown in Table 23.

Table 23: The frequencies of contrasting transitional phrases in the JPU Corpus and two subcorpora in sentence-initial position

Phrase	JPU	WRSS	PGS
But	308	75	61
However	138	23	47
Still	21	7	3
Yet	24	3	2
On the other hand	35	3	13
Nevertheless	25	5	7

As the table indicates, Hypothesis 2 has been confirmed: in sentence-initial position, the coordinating conjunction is most frequent in the main corpus and in the two writing subcorpora, with four of the transitional phrases represented by much lower frequencies.

4.3.3 Hypothesis 3

Clarity of written expression, in whatever genre, is enhanced by the use of concrete verbal phrases that accurately identify the reader's intentions and adequately cross-reference an earlier segment of the text. This is especially true of academic writing, which needs to operate with valid reporting verbs. However, this area appears to be a source of problems for the non-native writer, whose vocabulary may not be wide enough and who has not had extensive reading experience in the target language.

One early insight I gained as a writing tutor into both native speaker and non-native speaker academic texts was the frequent use of the phrase "mentioned above," and its many active and passive variants. I identified three potential problems with this usage. First, on many occasions, the act of mentioning appeared to be a form of hedging, referring to an important point in the argument made earlier. Instead of finding a "mention" of these points, I would often locate a discussion, a definition, an illustration. The first problem, then, was that of validity. The second reason I became interested in the phrase was related to the adverbial component. Referring to the antecedent as being "above" appeared to characterize most formal text types, such as those in the legal profession, and in instructions. Its use in academic writing may contain the intentional or unintentional desire to make the text more formal than one may consider necessary. The writing courses aimed to sensitize students to this issue so they could look for alternative expressions. The third problem area was maybe the most relevant from a linguistic and pedagogical point of view: what many authors referred to this way appeared in the previous sentence. While another frequent use of the phrase appears to be in concluding sections of papers, with the adverb being an all-purpose filler for "in this paper," the frequency of the phrase was also high in sentences making an anaphoric reference to a point in the previous sentence. In these contexts, simple deictic phrases would suffice.

Hypothesis 3 suggested that there would be a relatively high frequency of "above" in anaphoric verbal phrases, and that a significant verbal collocate would be *mention*. Further, the hypothesis claimed that in the PGS and WRSS these frequencies would drop, as a result of the practice students had in those courses. To verify or reject it, the hypothesis was submitted to the following analysis. First, I obtained the KWIC concordances for the variants *mentioned above*, *above mentioned* and *above-mentioned*. The frequencies of these expressions were recorded for the main corpus and the two subcorpora, as shown in Table 24.

Table 24: The frequencies of "mentioned above"/ "above mentioned" in the JPU Corpus and two subcorpora

JPU	WRSS	PGS
24	1	0

The hypothesis has been confirmed by the test, as shown in the table. To determine the level of statistical significance of the finding, however, I ran the chi-square test on the data. As is clear from Table 25, over 23 occurrences of the phrase were observed in the non-writing subcorpora (Rest of JPU). I tabulated this data, as shown in Table 4.16.

Table 25: Frequencies of the phrase in the non-writing subcorpora (RRS, ES, LPS) and the two writing subcorpora

WRSS	PGS	Rest of JPU
1	0	23

The chi-square value of 46.45 (df = 2) was significant ($p < 0.001$), lending support for the hypothesis that students in the non-writing courses used significantly more such phrases than in the writing courses. In this instance, it appears that both pedagogical and statistical significances were present.

In noting these occurrences I located a number of similar variants in the main corpus. These included two main types of phrase: past participle + above and definite article + above + noun phrase (such as *listed above*, *described above*, *detailed above* and *the above facts*, *the above criteria*, *the above writers*, and even, occurring twice, *the above paragraph*).

4.3.4 Hypothesis 4

Related to the previous area of investigation is the fourth hypothesis, concerned with the performative collocates of *I*. The study of this issue was necessitated by a potential pedagogical outcome: I wished to gather data on what the 332 writers of these texts identified as their aims and methods in their texts, either in explicit thesis sentences and statements of method or in topic sentences referring to a particular point made in the main body of the text. This information is necessary to form an overall view of the types of aims students identified for their scripts, and can serve as the basis of evaluating writing strategies in students' texts.

This hypothesis was a broad one: it suggested that aims would be primarily identified by the *would like + to infinitive* structure (Type 1). For statements of method and topic sentences, the *I will* construction would be more frequent (Type 2). To test this claim, I ran the KWIC concordance on the full corpus and analyzed the keyword *I*, identifying patterns that suggested significant collocates in the two types. Then I recorded the frequencies of the individual patterns and rank ordered the frequency of collocates. The results are shown in Table 26, with the frequency of the performative in parentheses.

There were a total of 44 occurrences of Type 1, whereas 93 of Type 2 patterns. Hypothesis 4 was confirmed: Type 2 expressions were more frequently associated with the modal auxiliary *will*. These were not only more frequent than Type 1 patters, but also showed a wider variety and more explicitness. (The pedagogical application of the finding will be discussed later in this chapter.)

Table 26: Thesis statements, topic sentences and statements of method expressed by the *I would like* to structure and *I will* in the JPU Corpus

I would like to	I will
	analyse / ze (11)
	examine (10)
	present (7)
	attempt (6)
show (5)	
(4) examine, focus on, point out	point at / out (4)
(3) analyse / ze, present	(3) discuss, focus on, give analysis/classification/tips, introduce, show, use
(2) emphasise / ze, find out, get answer	(2) check, concentrate on, deal with, demonstrate, describe, evaluate, investigate, provide data/view
(1) answer question, call reader's attention, clarify deal with, describe, explore, get to know, give suggestions, highlight, prove, stress, suggest, touch upon, try, write	(1) address, argue, compare, delineate, devote space for, emphasize, draw conclusion, have a look, highlight, list, make analysis, make attempt to find, monitor, report, shed light, study, sum up, summarize, survey, take the mean, tell, try, turn to

4.3.5 Hypothesis 5

Learners of EFL were found to overuse the pattern of the epistemic stem "I think [that]" in writing in a contrastive study of a sample of the ICLE L2 and an L1 student corpus (Granger, in press). The study found 72 occurrences of the phrase in the learner corpus, compared to only 3 in the native corpus. Granger hypothesized that the reason for this difference (termed "overuse") lay in students' differential concepts of spoken and written registers.

Hypothesis 5 investigated JPU students' use of the stem. The two corpora used in Granger's study (in press) were made up by 251,318 and 234,514 words, respectively. For comparative purposes, the combined subcorpora of the PGS and WRSS were used—these are valid sources for such data both in terms of text types in them and tokens: the combined length of the two subcorpora is 231,211 words. The KWIC concordance of *I think [that]* was captured for the PGS and the WRSS, and the frequency of the phrase compared with those in the other two samples. The result is tabulated in Table 27 (showing the frequencies normalized for 200,000 words).

The difference between the use of the phrase by EFL learners and native users was confirmed. As can be seen, the difference between frequencies in the L1 and the combined Hungarian learner subcorpora was markedly lower than between the ICLE and the L1 corpus.

Table 27: Frequencies of "I think that" in the three corpora

ICLE	L1 writers	PGS and WRSS combined
72	3	21

However, one is cautioned not to overgeneralize from the result that both L2 learner corpora contained higher frequencies of the phrase. The main reason for this caveat is that the relative frequency of *I think [that]* in the individual subcorpora is hardly significant. Also, we know little of the purpose and audience of the individual scripts contained in the ICLE and the native sample. In the PGS and the WRSS, the use of the phrase cannot be regarded as "overuse" unless one further explores these two text organizing principles. As this was not performed on the other two corpora, the hypothesis that learners overuse *I think [that]* cannot be confirmed—further studies are necessary. For the future analysis, the variables of purpose and audience have to be controlled and validated for both the L1 and the L2 samples.

4.3.6 Hypothesis 6

The use of the adverb *very* in written production has been the subject of a number of rhetorical and pragmatic analyses. Zinsser (1998) suggested that this adverb and what he called "little qualifiers" such as *a bit*, *a little*, *sort of*, *kind of*, *rather*, *quite*, and *in a sense* dilute one's style (p. 71). Explaining his professional writer's attitude in the context of purpose, he pointed out that "every little qualifier whittles away some fraction of the reader's trust. Readers want a writer who believes in himself and in what he is saying" (Zinsser, 1998, pp. 71-72). The issue is also related to the Gricean (1975) maxims of quantity and quality. As for the use of amplifiers and *very*, Granger (in press) hypothesized that when L2 learners "over-use" *very*, they compensate for their "under-use" of what may appear to be more specific amplifiers.

Hypothesis 6 was based on the experience that introduced to JPU English majors the notion that when aiming at concreteness in academic writing, authors need to review their use of such adverbs so that their intentions may be transparent to readers. As Appendix F shows, a component of a WRSS syllabus introduced the "Very-less week" program so as to make students aware of the issue. The hypothesis claimed that the adverb would still have a high frequency in the JPU Corpus, but that it would be less significant in the PGS and the WRSS.

As Appendix K reveals, *very* is ranked 83rd in the raw frequency list of the full corpus. To test the hypothesis on its distribution, I tabulated the frequencies for *very* in the PGS, the WRSS, and the non-writing subcorpora (RRS, ES, and LPS), and then calculated the chi square index to determine whether differences were statistically significant. When looking at Table 28, we can see that the lowest frequency was found in the WRSS, followed by the PGS, and that the highest figure was obtained for the rest of the corpus.

Table 28: Distribution of the frequency of *very* in the three subcorpora

PGS	WRSS	Rest of JPU
176	81	299

The chi square test revealed that the differences were significant ($\chi^2 = 128.9$, df = 2, $p < 0.001$), verifying the hypothesis: the WRSS scripts contained much lower frequencies of *very* than either of the other two subcorpora. Whether or not this tendency can be observed in the long run requires further study, however.

4.3.7 Hypothesis 7

Both in writer- and reader-based prose, authors are advised to look for ways to enliven their language by the use of specific expressions that carry their exact points and attitudes. McMahan and Day (1984), Raimes (1996), and Leki (1989), among others, made this point. Zinsser (1998) added that for such specificity to occur on the vocabulary and text level, one needs clarity of thought: in personal essay writing and in academic discourse, writers are advised to establish simplicity, rather than clutter. Critically reading one's own text, sharing with others, and monitoring the progress during revision are the stages of how this development takes place.

One form of clutter of thought and of expression, in both L1 and L2 writing, is the use of imprecise vocabulary that does not readily lend itself to interpretation. The writing pedagogical experience of the past semesters at JPU has familiarized me with the issue, and by reading and commenting on students' drafts, I aimed to enable participants to work on clarity and specificity. This is a long process. To investigate a part of the related segments of the JPU Corpus, I looked for the occurrence of five words that seemed to be frequent in student writing: two nouns, two adjectives, and an abbreviation: *case*, *thing*, *good*, *interesting*, and *etc*. Hypothesis 7 claimed that the frequency of these words would be lower in the PGS and the WRSS than in the rest of the JPU Corpus, as students in the WRS courses had the advantage of practicing learning and revising strategies for the avoidance of these vague terms.

To test the hypothesis, I obtained the frequency of the lemmas CASE and THING, and of the two adjectives and the abbreviation, and calculated the χ^2 value for each set of distribution. The results appear in Table 29.

Table 29: The distribution of and statistical information for the frequency of each of the five words in the three subcorpora

Word	PGS	WRSS	Rest of JPU	χ^2	df	p
CASE	121	76	146	8.77	2	< 0.05
THING	74	46	159	74.46	2	< 0.001
good	128	90	163	20.97	2	< 0.001
interesting	68	22	61	26.73	2	< 0.001
etc.	19	5	68	71.28	2	< 0.001

The table reveals that for each word, the differences of frequencies were significant; the lowest level for CASE, and for each of the other four observations, the high statistical significance level of < 0.001 was obtained. This verifies the overall hypothesis that in the writing subcorpora specificity of expression was not marred by the frequent use of these words.

4.3.8 Hypothesis 8

The last investigation involving the full sample of the JPU Corpus was concerned with two prefabricated patterns: *the fact that*, and *in order to*. The first of these often appears in both L1 and L2 texts with no apparent extra information contained in them. The third phrase is regarded by several sources as a redundant prepositional phrase that can often be substituted by the simple *to* infinitive (see, for example, Strunk & White, 1979; Raimes, 1996; and Zinsser, 1998).

As far as *the fact that* is concerned, Granger (in press) noted that L2 student writers demonstrate excessive "over-use" of the phrase, also citing Lindner (1992), who studied a corpus of German EFL texts and suggested that the high frequency of the phrase can be attributed to students' perception that expository and argumentative writing has to carry high "verbal factualness."

The hypothesis claimed that there would be lower frequencies for *the fact that* and *in order to* in the PGS and the WRSS than in the rest of the JPU Corpus. To test the hypothesis, the same procedure was applied as for testing the previous one. The results appear in Table 30.

Table 30: The distribution of and statistical information for the frequency of the two phrases in the three subcorpora

Phrase	PGS	WRSS	Rest of JPU	χ^2	df	p
the fact that	24	27	75	38.98	2	< 0.001
in order to	35	50	48	2.98	2	NS

Note: NS = not significant

As the table shows, part of the hypothesis was confirmed by the test: the phrase *the fact that* is significantly more frequently used in the three subcorpora than either the PGS or the WRSS. However, no similar trend was observed for the phrase *in order to*—the distribution of its frequency being fairly even. The second part of the hypothesis was thus rejected.

4.3.9 Hypothesis 9

So far, we have seen the results of eight investigations, highlighting various lexical choices students made in writing. They have involved the analysis of one subcorpus, the full JPU Corpus, contrastive studies across the subcorpora and the analysis that showed similarities and differences between the JPU Corpus and the ICLE. For the last two investigations, I selected the research paper samples of the WRSS. As noted in section 4.2.1 on the current composition of the JPU Corpus, the majority of scripts, 107, were submitted as the final research paper requirement of the course. This collection represents a valid basis on which to test hypotheses 9 and 10, the former related to introductions, the latter to conclusions.

The investigation of the types and composition of these first sentences of the introductions was motivated by the linguistic and pedagogical concern with the importance of drafting and revising introductory and concluding matter. By looking closely at this sample, we can gather useful information on students' choices, using authentic data that can be exploited for future language education (to be discussed in detail in the next section of this chapter).

Of the 107 papers, 33 discuss aspects of Hungarian newspaper articles published on the day students were born. As section 3.3.3.2.1 suggested, this option was designed to include a personal intrinsic motive for students to begin to want to do research. The high number of such papers seems to prove that the approach was successful. However, a large number of other content and method types are also represented in this subcorpus—these are listed in Table 31.

Table 31: Content and method types in the 107 research papers in the WRSS

Type	Number
Newspaper articles from the day student was born	33
Analysis of students' writing	30
Survey among students	20
Word processing for writers	4
Types of revision	3
Analysis of WRS course tasks, readings, procedures	2
Analysis of Umberto Eco's writing	2
Survey among teachers	2
Analysis of teacher's comments on portfolios	1
Analysis of essay test markers' comments	1
University syllabus analysis	1
Analysis of writing textbooks	1
Introductions in *75 Readings*	1
Analysis of introductions in *HUSSE Papers*	1
Analysis of narrative essay	1
Analysis of Zinsser's notion of simplicity	1
Models of paragraph	1
Analysis of structure in research papers	1
Proficiency test for high-school students	1

The hypothesis claimed that the type of introductory sentence chosen by students would affect the length and vocabulary of the first sentence. Besides, I aimed to gather descriptive information on the frames of the first sentences (Andor, 1985). To test the hypothesis, the first sentence of each introduction was saved as a separate document, which was then processed by the concordancer, also calculating tokens, types, and average sentence length in different groups: in short, the introductory sentences were treated as a mini corpus. Besides these measures, a table was also designed, listing the types of introductions observed.

The mini corpus of these sentences contained 1,946 words, of 579 types, a ratio of 3.36. The average length of a sentence was 18.18 words.

To test the validity of the hypothesis, I performed a content analysis of the sentences, using categories. Initially, I identified five categories to capture the types of frames of the introductions, representing different approaches I knew students employed in their texts. These included

- describing a *personal* incident related to the theme (e.g., "Having read the newspaper issue of Kisalföld of 14th September 1978, a whole new world opened to me.")
- identifying a relevant *historical* detail ("In June 1979 Leonid Brezhnev paid a visit to Hungary.")

- opening with a *narrative* ("The first thing that many people do in the morning is opening one of the daily newspapers and browsing among the articles."
- giving a *definition* of a field, an issue or a problem ("Students' opinion about syllabi can influence the popularity of courses.")
- beginning the text with *five* semantically germane nouns, verbs or adjectives ("Clutch, weep, glare, jerk, loathe.")

The last of these introductory frames was first employed and practiced, primarily for personal descriptive and narrative essays, in the WRS course in the Spring 1998 semester.

In categorizing the introductory sentences, I scanned them for traits of these frames. As some introductions did not fit into the original categories, new ones were set up:

- stating a matter clearly *obvious* for the intended reader, often containing determiners such as *every*, *each*, *all*, or adverbs like *always* (e.g., "Newspapers are used for informing the population about how the society works and what goes on all over the world.")
- stating the *aim* of the paper ("In this paper my aim is to compare two Hungarian daily newspaper issues...")
- defining the *method* of the investigation ("One possibility to gather information about a period of time is to read newspapers.")
- directly addressing the *reader* ("Reading old newspapers may make you realize what has and what has not changed during the years.")
- including a direct or indirect *citation* from a source ("According to Harris (1993, p. 81), a general point about writing is that it cannot be seen in isolation...")
- asking a *question* ("What is exactly a portfolio?")
- beginning with the *title* of a source ("Bits & Pieces.")

These labels were then assigned to the introductory sentences. To test the reliability of the categorization, the same procedure was conducted a second time. In only two instances was there a difference between the first and the second result, which were identified with a question mark, and the first and second label recorded. Altogether, I identified twelve types of introductions in the WRSS sample, with the 13th represented by the problematic examples. When these measures were taken, the frequency of types was rank-ordered. The results appear in Table 32. The table shows overwhelming preference for four types of introduction: those based on a definition, a personal incident, an obvious issue, and a historical detail. Altogether, the four types account for the majority of the papers, 83 out of 107.

Table 32: The rank order of types of introductory sentences in the WRSS sample

Rank	Type	Frequency
1	definition	47
2	personal	15
3	obvious	12
4	historical	10
5	aim	7
6	method	4
7	five	3
8	citation reader ? (obvious-definition; obvious-historical)	2
9	narrative question title	1

To confirm or refute the hypothesis that the type of introduction affected the length of the first sentence, I devised the following procedure. Of the 107 sentences, I selected the 83 that belonged to the most popular options. As the rest of the sentences were each represented by only seven or fewer examples, they were eliminated from the investigation, as their low frequency would not have given sufficient information on length distribution. After this, I calculated the length of each of the 83 sentences in the four main groups. When these indices were obtained, I determined the effect of the type on length via one-way analysis of variance (ANOVA). Table 33 presents the statistics.

Table 33: Results of the analysis of variance on the data of length of first sentences

Source	df	SS	MS	F	Pr[X>F]
Between	3	199.14	66.38	1.20	0.31
Residual	80	4410.10	55.13		
Total	83	4609.24			

Grand Sum = 1504.00 Grand Mean = 17.90

According to the figures in the table, the ANOVA findings are inconclusive: no significant differences were found (F = 1.20; p = 0.31). The type of sentence did not affect its length. This result points to the need to analyze the full introductory paragraphs, so as to reveal how type may affect its size and structure.

4.3.10 Hypothesis 10

Similarly to the importance of how a research paper opens the theme for the reader, in writing the conclusion's last sentence, the author has an opportunity to make a last and maybe lasting impression. In this investigation, I analyzed the final sentences of concluding sections of the 107 papers, looking for the same types of information as in the previous study. Hypothesis 10 claimed that there would be a number of types of concluding sentences, which in turn would affect their length and vocabulary. The procedures for testing this last hypothesis were the same as for the previous one.

The mini corpus of the concluding sentences was made up by 105 sentences—two fewer than in the introductory mini corpus, as two students did not include a conclusion in the submission. The sample contained 2,389 words, representing 818 types, resulting in a ratio of 2.92. The rounded average length of sentence was 23 words. When compared with the same statistics for the introductory mini corpus, we can see that concluding sentences tended to be somewhat longer, using more types of words on average than the introductory ones. However, the differences cannot be regarded as marked, as shown in Table 34.

Table 34: Descriptive statistics of the two mini corpora

Index	Introductions	Conclusions
Tokens	1946	2389
Types	579	818
Ratio	3.36	2.92
Average length	18.18	22.75

As for the typology of the last sentences, the following eight categories were set up initially:

- summary of a *qualitative* result (e.g., "The more senses are involved in learning, the deeper the learning will be.")
- summary of a *quantitative* result ("From the foregoing it is clear that all of the analyzed essays except for one or two are better than the average.")
- statement of practical implication ("I also learnt about the relationship between journalism and the political life.")
- identification of *limitation* of study ("As the other classes during the semester were more or less active than the one dealt with in this paper, this research paper and the results of it can be applied to this particular class."
- a direct or indirect *question* ("I wonder how many findings will apply to me and my peers in the future.")
- identification of *hypothesis* or problem for future study ("It could be used for finding out why some important information was left out from Hungarian papers, and what they were.")
- *non-sequitur* or irrelevant notion ("Only children were excited when they were waiting for Santa Claus to bring them presents.")
- stating the *obvious* ("Other sources can be used as well for doing similar research on this topic, which would certainly enrich knowledge about this field.")

Again, not all concluding sentences could be grouped under these headings. The three new categories added were

- *citation* (e.g., "Such an essay test might be a torture for those students who dislike essay writing, but it 'continues to serve as a challenge for a number of students who have shown excellence in writing.'—reports Horváth József....")
- addressing the *reader* ("Thank you for not leaving and reading the Research Paper.")
- *unclear* content or ambiguous ("With this paper I got the information, what I wanted to know.")

Each of the 105 sentences was coded, and the grouping double-checked. In the second analysis, the original division was found to be reliable.

Table 35: The rank order of types of concluding sentences in the WRSS sample

Rank	Type	Frequency
1	qualitative	47
2	practical	26
3	obvious	9
4	unclear	7
5	quantitative	5
6	question	3
7	hypothesis limitation non-sequitur	2
8	citation reader	1

The two most popular last statements in the mini corpus were represented by the qualitative and the practical outcome types. This result is in line with previous pedagogical experience suggesting that student writers favored these options. They also appear to be relevant for the types of research design the scripts were based on. However, the high ranking of the obvious type of sentence and of the unclear category calls attention to the need for more practice in the area of writing conclusions. As the next section on the pedagogical exploitation of the corpus will show, this can be facilitated by channeling back the information on students' scripts to the writing course, using authentic student texts.

Finally, to test the relationship between type of concluding sentence and length, I employed a one-way analysis of variance test for types. I used the sentence-length data for the qualitative and practical groups, and the combined length for the obvious and unclear types. The results appear in Table 36.

Table 36: Results of the analysis of variance on the data of length of last sentences

Source	df	SS	MS	F	Pr[X>F]
Between	2	862.29	431.14	4.34	0.02
Residual	86	8539.22	99.29		
Total	88	9401.51			

Grand Sum = 1978.00 Grand Mean = 22.22
Qualitative Mean: 23.36
Practical Mean: 24.23
Obvious + Unclear Mean: 15.62

The table shows that the analysis revealed a significant effect of type of concluding sentence and length: F = 4.34; p = 0.02. Whereas the mean length of the qualitative and practical type of concluding sentences was almost identical (23.36 vs. 24.23 words), the length of the combined group of obvious and unclear type sentences was 15.62, for which the analysis confirmed significant variation. Thus, Hypothesis 10 claiming that type of sentence affected length was verified.

The statistical finding may imply that students who wrote the type of concluding sentences that were categorized as either unclear or obvious themselves had difficulty ending their papers, and thus they opted to write much shorter sentences than others. This hypothesis, however, does not intend to suggest that there is correlation between quality of conclusion and quantity of concluding sentence. Also, factors such as grammatical accuracy of the sentences, the type of concluding sentence and the full concluding paragraph, and the appropriateness of the type of conclusion in relation to the body text of the research paper are to be investigated in the future.

4.4 Pedagogical exploitation of the corpus

4.4.1 Learning driven by data from the learner

The JPU Corpus has been conceived as a potentially useful basis for two major types of application: linguistic and pedagogical. We have seen some of the results of the linguistic analysis of the corpus, already noting pedagogical motivations and outcomes. But as producers of these text, students could also directly benefit from contributing to the collection: this use has been facilitated by worksheets in recent pre-service and in-service WRS courses. This section will present the rationale, design and use of such materials, after which I will suggest ways of incorporating the results of the present analysis in designing new worksheets.

As Chapter 2 demonstrated, DDL is often used for individual study. Applying the classroom online concordancing technique, the tutor and the student focus on relevant issues, arising from either the student's or the tutor's initiative. Parallel concordances are exploited, as in Johns's (1997b) kibbitzer technique. However, the corpus of students' texts facilitates pair and group work, too. In several WRS courses, students were provided with handouts that featured samples of their own writing, the purpose being that I aimed to draw attention to the importance of lexical and collocational choices. As authorship was hidden in these examples, the affective filter was lowered, yet the studying and discussing of the co-texts allowed for the effective use of the monitor (Krashen, 1985).

4.4.2 Exploiting for classroom work

I introduced off-line concordancing in university language education to add a dimension to the awareness raising activities conducted in the sessions. The first versions of students' scripts were submitted to KWIC concordancing. On several occasions, this technique served to highlight common features of students' writing, which appeared especially characteristic of Hungarian teachers' discourse. Here, I will present two such examples.

The first example posed the question of how appropriate it is to refer to students as "ours." Especially in the RRS and the PGS, authors seemed to prefer the use of the first person possessive pronoun as a collocate of "pupils" and "students." Example 1 aimed to raise the issue and allow for group discussion.

Example 1: Worksheet on possessives

In academic writing, participants in research and in the wider educational context should always be referred to as that: individuals. No matter how much we like them, students and pupils we teach should not become our property. In the following concordance lines, the authors have appropriated students. With a partner, discuss your views on this issue, and then rewrite the co-texts by replacing the possessives. In a number of instances, several alternatives are possible.

1 tions about the television and most of my pupils agree with her point of
2 cially on the introduction part. When my pupils had finished their works
3 scussion. It is fascinating for me that my pupils liked that Barbara - the
4 opic's historical background. Some of my pupils opted for this method.
5 irstly, I reply on the second question. My pupils were satisfied with their

1 and analyse it. The next step was that my students had to fill in a questi
2 of the original introduction and what my students have done. I wanted
3 ng the original introduction and using my students' opinions about this
4 the specific. My last question was for my students what they think, what
5 ussion. I am going to prove it through my students' works. There was a

The purpose of the second example was to present to students the task of reporting the author's aims in a research paper. I had sampled the introductions of their submissions and found a limited lexis of verbs that announced the purpose and method of the paper. Although most of this vocabulary appeared to be relevant to the main texts they were clipped from, I realized there was a need to raise students' consciousness of the importance of using more specific verbs in these sections. The following handout was produced.

Example 2: Worksheet on reporting verbs

When you read or write a paper, you often find that reporting what the researcher will do greatly facilitates the clarity and relevance of the results. With a partner, list ten verbs, appearing in introduction, that indicate what the paper will "do." After that, skim the worksheet and underline those you listed.

1 and distribution. In this paper I will address the latter of the issues,
2 links with the rest of the paper. I will also scan for the thesis sentenc
3 were written in 1996. I will analyse my essay's introductio
4 texts, conclusions and references. I will check whether there are
5 and their analyses. In my paper I will concentrate on semantic relati
6 are analysed in a text. I will concentrate on pronouns in t
7 a foreign language - writing skills. I will evaluate my essays in terms of
8 that makes a text coherent. I will examine repetition in the
9 and Oleanna - of the chosen essay. I will examine the text according to
10 making the writing more effective. I will introduce different revision
11 many hyponyms and antonyms, but I will introduce some here.
12 The hypothesis that I will present and discuss in some d
13 in terms of their structures; I will survey the introductions, the

After the task, students discussed the use of verbs they listed but did not find on the worksheet.

4.4.3 Guiding individual study

In writing courses, tutors aim to allow students to experiment with topics, text types and purposes so that what they learn in the sheltered environment may be applicable in future courses. The process approach to writing pedagogy emphasizes this need for sustainable improvement—but even if the curriculum facilitates cooperation between courses, in the framework known as writing across the curriculum, the role of the writing course has been fulfilled when the course ends. To provide for continuity after these classes are over, writing tutors can apply one task type based on DDL: the individual study guide based on each student's last submission to the course (Horváth, 1999b).

In recent JPU ED writing courses, undergraduate and postgraduate students have received such tasks. Combined with the tutor's assessment of their work, these guides aimed to raise students' awareness of discrete features of their writing, positive and negative qualities that I commented on in the final assessment but also regarded as suitable for further study. The use of the guides followed weeks of work on the text: the students and the teacher had consulted the merits of the submission and the latter suggested areas for thematic, structural, and grammatical improvement. It stands to reason that individual students' consciousness of their writing strategies and skills grew as a result—what the study guides added to this process was the opportunity to focus on one factor of their writing. Example 3 presents a study guide for a student who was asked to consider replacing the all-purpose noun "things" for more specific terms in the paper.

Example 3: Replacing things

1	or a comic strip. They are usually funny	things in some connection
2	to underline, to write in bold type and other	things. One of the six "La
3	language in a variety of forms (desribing	things, people, places ...; st
4	They should be able to inquire about these	things. They should be ab

Example 4 is similar to the previous one: it, too, is concerned with concrete vocabulary, this time challenging the writer to evaluate her data and identify more precise terminology instead of "good."

*Example 4: What makes a good ***?*

1 revises the essential rules of how to write a good composition, from a goo
2 composition, from a good introduction to a good conclusion. In this exer
3 from what you are trying to say. It's a good idea to check through y
4 feelings; word order; semantic markers; a good introduction and con
5 of how to write a good composition, from a good introduction to a good

Potentially the most intrinsically motivating of this type of study guides are those that invite the student to scan and reflect on the co-texts of the first person singular pronoun. When such use is frequent, the student can discover new contexts for the theme, enabling her to verify a focus.

Example 5: What I could and would

7 I could not cope with the problem of expressing my ideas in an exact w
2 I could not get rid of my second person sigular personal pronouns. I c
3 I could so as to fulfill the requirements of a good essay which is subj
4 I tried to be more careful and accurate as a whole. I managed to elimin
5 I tried to translate expressions word- by-word in lacking an up-to-date
6 I tried to use the language as creatively as I could so as to fulfill the req
7 I used a lot of abbreviations ("can't" or "isn't") and noteforms (under
8 I wanted a quick result, therefore the presentation of my work was simp
9 I wanted to be more wise than I really was. It is best represented by the
10 I wanted to have my own special style even if it was ridiculous someti
11 I would be still happy but then came learning to write in the Writing Ce
12 I would like to develop to be an academic English writer.
13 I would like to give a clear chart about the strong and weak points of m
14 I would like to point out my mistakes and to give suggestion how I can r

Both the classroom and the individual study guides aimed to raise students' awareness of their own writing, so they were in a better position to continue to improve editing and revising skills. By using students' original texts in the early stages of developing a research paper, I aimed to help students from a discourse community in a sheltered environment. Scaffolding and focusing on discrete elements of their writing was not employed to focus on error; rather, the objective was to highlight features that represented choices writers made in the process of exploring a field. The study guides also encouraged exploitation of students' texts after the course ended. The concordance revealed lexical choices that were often subconscious. Used in combination with more traditional task types, the concordance-based study guides can result in increasing levels of learner autonomy, an essential criterion for development in the long run.

4.4.4 Other applications

Besides the study guides prepared earlier, the analyses presented in this chapter lend themselves to practical applications. As noted in section 4.3.4, students used the modal auxiliary *will* more often in thesis and method statements than the *I would like to* construction, and they employed a wider array of verbs. This data can be adopted for WRS sessions that deal with the need for explicit and valid information on, for example, how the student will present various data types.

The verbs that were shown to collocate with *I will* can be listed and the following worksheet prepared for pair work:

Example 6: Recycling students' speech acts

The verbs listed below are clipped from previous students' research papers. They were used in the Introductory and Method sections. With your partner, discuss what these verbs indicate in a paper. Then, suggest which three of the verbs were most frequently used by the students.

address
analyse
analyze
argue
attempt
check
compare
concentrate on
deal with
delineate
demonstrate
discuss
evaluate
examine
focus on
give analysis
point out
present
summarize
survey

The JPU Corpus sample can facilitate the preparation of a large number of such authentic study guides.

4.5 Future directions

Since 1992, I have been collecting students' scripts for research and pedagogic purposes. The largest EFL written learner data in Hungary, the JPU Corpus has been instrumental in the description of learner lexis in written discourse. The gain this resource has offered has included linguistic and pedagogical applications.

There are limitations, however. The analysis of the corpus could not take advantage of tagging and the use of more sophisticated concordancing software. For future analytic studies, word-class and syntactic tagging has to be added. Another limitation is that the corpus contains no data from courses taught by other teachers at the department. For the corpus to represent such diversity, this avenue also has to be explored.

Yet even with these limitations, the corpus is representative enough for valid linguistic and pedagogical application. In the next phase of its development, I am planning to focus on incorporating first and last versions of personal narrative essays and research papers. A subcorpus will provide data for analyzing lexical and discourse changes a text undergoes during the process of revision. This parallel set of data will enable future research on vocabulary choice and size. Also, a growing corpus will continue to provide the raw material for classroom concordancing and study guides.

A second plan is to include the test essays written in the past six years as part of the proficiency tests. The current size of that handwritten data set is about half a million words. As the conditions of the essay writing test have differed greatly from those that gave rise to scripts currently incorporated in the JPU Corpus, a more refined view of learner written English may emerge. Together with the present structure of the corpus, these two sets of data can also facilitate diachronic studies of various features of language use under different circumstances.

Yet another vista of future work is the incorporation of students' theses in the corpus. The majority of writers who have contributed to the WRS and PGS subcorpora are still at JPU and will be submitting their dissertations in the next few years. Obtaining the electronic version of these texts would enable research to investigate the final outcome of university education.

Finally, to bring about an even more structured synthesis of corpus methods and writing pedagogy, a new type of annotation will be worked out: pedagogical corpus annotation (PCA). PCA is what teachers of writing already do all the time: they mark up text by students, who, in turn, attempt to understand, critique and apply some of the comments. This part of the pedagogical process, however, is often lost to research and pedagogy when the comments are shared. With PCA made part of the corpus, teachers' commentary can be incorporated with the student text, and fine-tuned analysis would be made possible. Applications of PCA could include the testing of the consistency and reliability of types of comments across comments, as well as the validation of the comments teachers make. Another use lies in the contrastive analysis of discourse and style in students' and teachers' texts (Horváth, 2001). Such an incorporation of teacher comments can be managed when learners submit scripts on disk, so that the reader can add comments via either a word processor's annotation or footnote module or a dedicated co-author program, such as *Prep* 1.0 (Chandhok, Kaufer, Morris, & Neuwirth, Miller, & Erion, 1993). Besides, students' own reflective notes about the purpose and evaluation of their own texts and those of their peers can enhance the data of present-day learner corpus projects.

CONCLUSION

This book has undertaken to probe into EFL writing pedagogy with advanced Hungarian university students. After reviewing relevant theory and empirical work in the related fields of writing pedagogy and corpus linguistics, it has presented a possible ethnography of advanced writing at university, and given details of the lexical and discourse features of the JPU Corpus. It has aimed to synthesize pedagogy and linguistics by bridging the gap between process and product approaches—an area that remained virtually unexplored before this endeavor. The course of investigation has been framed by current understanding of writing processes (Zamel, 1992; Zinsser, 1998) and by the increasing prevalence of developing and exploiting representative and specialized computer corpora (Sinclair, 1997; Kennedy, 1998).

A number of factors have remained beyond the scope of the analysis. For example, there has been no space to position the theory and practice of the institutional assessment and evaluation of writing skills, which represents one of the outcomes of the writing process. This field is well worth further investigation. Also missing from the evaluation of the writing process have been the wide range of classroom and study guide materials developed during the past years and the assessment of the practice in out-off-class meetings with students.

However, it seems that the original three-fold aim of the study, that is, to collect evidence of advanced students' language use, to apply this data to research, and to apply the results in writing pedagogy, has been met.

In terms of processes, we have seen the development of the writing course syllabus that gradually moved in the direction of focusing on the process of developing writing skills, according equal importance to the products of each step during that process. A number of pedagogical innovations and new task and text types have been introduced and evaluated. Their analysis has been special in that it has had to be predominantly self-reflective: as no concurrent outside observation took place, the study describing and evaluating EFL writing pedagogy at JPU has employed mainly qualitative data.

In terms of products, this book has been the first in Hungary to present the case for the need to collect data on written language performance by advanced students of EFL. It has described and explained the design and development of the JPU Corpus, and provided a sample of the lexical and discourse analysis made possible by the scripts. We have seen the results of writing pedagogy in the work of pre-service and in-service students of EFL. The study of the ten hypotheses has provided evidence of the uniqueness of learner scripts. One area where the investigation may have opened up a new perspective of corpus application has been its limited focus on error. Instead, it has attempted to capture some of the characteristic elements of student writing in a non-prescriptive manner, much in the tradition of how corpora of

L1 texts are dealt with. The pedagogical exploitation of the scripts has been shown to include the design and application of several types of tasks that are to guide students' acquisition of lexis and discourse patterns.

Further empirical work is also to be conducted. There are scores of areas that can be studied for a fuller understanding of learner writing, and, eventually, how new generations of EFL students acquire and reflect on written discourse, whether personal or academic. Of these, I will recommend what appear to be most needed and relevant, divided into three groups according to discipline: those that are primarily concerned with writing pedagogy, those that focus on the analysis of a corpus, and those that would aim to synthesize the two disciplines.

Implications for writing pedagogy research

- Of the most relevance would be cross-sectional studies, based on classroom observations, structured interviews, the comparative analysis of school curricula and think-aloud protocols, exploring writing pedagogy in Hungarian secondary-school programs in the native language and in FLs. This appears to be among the most pressing needs. What is currently known about this practice indicates that much is to be done for students to find writing a meaningful and empowering activity (cf. Nikolov, 1999). The qualitative and quantitative study of L1 writing processes and products by Hungarian secondary-school and university students could begin to explore contrastive rhetorical issues.
- To establish what factors contribute to success in writing in a FL, case studies are necessary. For researchers and teachers to validate theory, such investigations can supply the data on which to plan syllabus development. These studies would need to be conducted on intermediate and advanced levels of proficiency.
- Innovative designs, with students involved to enhance the validity of the research agenda, could result in an ethnography of writing, similar to the chronicle-based report of cross-cultural and pedagogical issues in writing instruction by Pally & Diallo (1995). A nationwide project of action research in writing pedagogy could facilitate the development of such projects and the publication of the results. The project sets the context of systematic work in studying academic writing, including such activities as sharing and evaluating syllabuses, evaluating the curricular status of writing courses, establishing writing centers

at participating institutions, developing new team-taught courses, and developing new teaching materials.

- The expertise of the writing tutor and the role of training and continuous development are two closely related concerns well worth further research. Case studies are necessary to establish the pedagogical variables that contribute to student achievement. This focus could help reveal how various learner styles match and mismatch with teaching preferences. The results may also have implications in developing pre-service and in-service teacher education.
- An exploration of current issues in testing writing skills in EFL at Hungarian secondary schools and universities would contribute to collecting data on the validity and reliability of various types of assessment instruments. More research is to be initiated on the writing component of the secondary school-leaving exam, on marker training, and on validation procedures. Teams and individual teachers could conduct action research and validate progress and proficiency tests. Experiments measuring the effects of item and task type, length, and time variables would require concerted effort and sustained funding.

Implications for corpus studies

- One of the areas of such further research may seek to study the correlation between performance in oral and written tasks. We still have little empirical evidence of Hungarian EFL students' performance in these areas. Especially promising would be the assessment of communicative skills in authentic academic oral and written tasks across several higher-education institutions. The development of a parallel oral and written learner corpus could provide the sample for piloting such an investigation.
- Also of potential interest would be the development and piloting of further sets of study guides based on parallel L1 and L2 corpora. This could be achieved by extending the framework of John's (1997b) remedial data-driven tuition by supplementing a course with these types of materials and tasks, assisting individual students and groups with lexis, collocational use, grammar, discourse, and even punctuation.
- Clearly, the international extension of the use of the JPU Corpus would play a key role in the future. This would

mean the contribution of a 200.000-word subset of the JPUC in the ICLE and the LonLC. Comparative and contrastive studies of lexis and discourse could be developed.

Implications for a synthesis of pedagogy and corpus studies

- To analyze, using both qualitative and quantitative approaches, the processes observed in writing conferences and to inspect the text of teachers' written feedback we will need cross-institutional cross-cultural studies of response to student scripts. One could investigate the factors underlying feedback by native and non-native English speaker teachers. This would necessitate piloting the procedures for developing a learner corpus annotated by three types of comments: the writer's commentary on purpose and content; peers' comments; and the teacher's response. By incorporating these comments in the corpus, further research could investigate specific features of L2 writing processes, which, in turn, could be applicable in writing pedagogy. As suggested in the last chapter, the development of a corpus with PCA added to it could reveal patterns of discourse as writers and readers negotiate meaning (Horváth, 1999b).
- Finally, the study of the writing of the tutor would provide a new perspective on writing pedagogy. This research domain may include the investigation of how teachers' preferences in their own writing transfer to syllabus design and classroom procedures. This factor represents an entirely new vista in the ethnographic study of writing. Data used for the analysis could comprise authentic scripts by teachers and a sample of syllabuses.

It remains to be seen how many of these suggestions for further inquiries into the three areas will meet with support. Clearly, awareness of, and interest in, the need to raise standards in Hungarian writing pedagogy and the potential outcome of improved levels of student performance are among the motives that determine the sustainability of any educational proposal. In submitting my work on the processes and products of advanced writing in EFL, I hope to have laid the necessary basis and shown some direction for these and other studies that aim to achieve those ultimate goals.

Furthermore, the JPU collection of scripts can become the basis of establishing a Hungarian EFL learner corpus, so that college and university students' scripts may be collected and studied by a national team. One outcome of this study may be that writing teachers with a similar concern may cooperate in syllabus development and corpus analysis. We need such a collection to capture the essence of what goes into the writing process and how its products can be appreciated.

REFERENCES

Alderson, C. (1997). Do corpora have a role in language assessment? In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 248-259). London: Longman.

Allwright, J. (1988). Don't correct—Reformulate! *ELT Documents, 129,* 109-116.

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Washington: Author.

Andersen, R. (1988). Overwriting and other techniques for success with academic articles. *ELT Documents, 129,* 151-158.

Andor, J. (1985). On the psychological relevance of frames. *Quadreni di Semantica, 6* (2), 212-221.

Andor, J. (1998). On the lexical bases of ellipsis in English. In J. Andor, B. Hollósy, T. Laczkó, & P. Pelyvás (Eds.), *The diversity of linguistic description: Studies in the honour of Béla Korponay—A special issue* (pp. 39-54). Debrecen: Kossuth Lajos Tudományegyetem Angol-Amerikai Intézete.

Arnaudet, M. L., & Mary, E. B. (1990). *Paragraph development* (2nd ed.). Englewood Cliffs: Prentice Hall Regents.

Arnold, J., & Harmer, J. (1978). *Advanced writing skills.* London: Longman.

Aston, G. (1995). Corpora in language pedagogy: Matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Theory and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 257-270). Oxford: Oxford University Press.

Aston, G. (1996). The British National Corpus as a language learner resource. Paper presented at TALC 96 [Online]. Available http://info.ox.ac.uk/bnc/using/papers/aston96a.html

Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 51-64). Applied linguistics and language study. London: Longman.

Aston, G. (1998, April). A very large corpus, in or out of the classroom. Paper presented at the 32nd International IATEFL Conference, Manchester.

Atkinson, D., & Ramanathan, V. (1995). Cultures of writing: An ethnographic comparison of L1 and L2 university writing/language programs. *TESOL Quarterly, 29*, 539-568.

Babarci, B. (1998). In the cocoon of my room. *Novelty, 5* (2), 50-51.

Bacskay, K. (1998). *Portfolio* [Typescript].

Bárdos, J. (2000). *Az idegen nyelvek tanításának elméleti alapjai és gyakorlata* [Theory and practice of teaching foreign languages]. Budapest: Nemzeti Tankönyvkiadó.

Bartram, M., & Walton, R. (1991). *Correction.* Hove: Language Teaching Publications.

Bello, T. (1997). Improving ESL learners' writing skills. *ERIC Digest,* ED409746.

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition.* Hillsdale: Lawrence Erlbaum.

Berry, R. (1991). Re-articulating the articles. *ELT Journal, 45,* 252-259.

Berry, R. (1994). The *research project: How to write it.* New York: Routledge.

Biber, D. (1994). Representativeness in corpus design. *Linguistica Computazionale, 9-10,* 377-407.

Biber, D. (1996). Investigating language use through corpus-based analyses of association patterns. *International Journal of Corpus Linguistics, 1* (2), 171-197.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge approaches to linguistics. Cambridge: Cambridge University Press.

Bloor, M., & St. John, M. J. (1988). Project writing: The marriage of process and product. *ELT Documents, 129,* 85-94.

Blue, G. (1988). Individualising academic writing instruction. *ELT Documents, 129,* 85-94.

Boughey, C. (1997). Learning to write by writing to learn. *ELT Journal, 51*, 126-134.

Brumfit, C. (1995). Teacher professionalism and research. In G. Cook & B. Seidlhofer (Eds.), *Theory and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 27-41). Oxford: Oxford University Press.

Bullon, S. (1988). The treatment of connotation in learners' dictionaries. In T. Magay & J. Zigány (Eds.), *BudaLEX '88 proceedings: Papers from the 3rd International EURALEX Congress, Budapest, 4-9 September, 1988* (pp. 27-33). Budapest: Akadémiai.

Bullon, S. (1997, January). The evidence is mounting: Why corpora are growing. Plenary paper presented at HUSSE 3, Pécs.

Burnard, L. (1996). Introducing SARA, an SGML-aware retrieval application for the British National Corpus. Paper presented at TALC 96 [Online]. Available http://info.ox.ac.uk/bnc/using/papers/burnard96a.html

Caudrey, T. (1996). What the "process approach" means to practising teachers of second language writing skills. *TESL-EJ, 1* (2), A-3.

Caudrey, T. (1998, April). Assisted writing in the computer lab. Paper presented at the 32nd International IATEFL Conference, Manchester.

Chandhok, R., Kaufer, D., Morris, J., Neuwirth, C., Miller, D., & Erion, P. (1993). *Prep 1.0* [Computer program]. Carnegie Mellon University.

Chapelle, C., & Jamieson, J. (1989). Research trends in computer-assisted language learning. In M. C. Pennington (Ed.), *Teaching languages with computers: The state of the art* (pp. 47-59). La Jolla: Athelstan.

Chen, J. F. (1997). Computer generated error feedback and writing process: A link. *TESL-EJ, 2* (3), A-1.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambrigde, MA: MIT Press.

Clear, J. (1993). From Firth principles: Computational tools for the study of collocation. In M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 271-292). Philadelphia: Benjamins.

Clear, J., Fox, G., Francis, G., Krishnamurthy, R., & Moon, R. (1996). COBUILD: The state of the art. *International Journal of Corpus Linguistics, 1* (2), 303-314.

Clouse, B. F. (1986). *The student writer: Editor and critic.* New York: McGraw.

Collins COBUILD English Dictionary. (1995). London: HarperCollins.

Cohen, A. D., & Cavalcanti, M. C. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155-177). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Composition of the BNC. (1997, October 22). [Online]. Available http://info.ox.ac.uk/what/balance.html

Cook, L. (Ed.). (1996). *Voices in the English classroom: Honoring diversity and change.* Classroom practices in teaching English. Urbana: National Council of Teachers of English. (ERIC Document Reproduction Service No. ED 386 756)

Crews, F. (1987). *The Random House handbook* (5th ed.). New York: McGraw.

Csapó, J. (1997). Corpus linguistics and English language teaching. In M. Kurdi & J. Horváth (Eds.), *HUSSE papers 1997: Proceedings of the Third Biennial Conference* (pp. 304-313). Pécs: University Press Pécs.

Csomay, E., & Szerdahelyi, J. (1997). *Writing: A process approach composition book for secondary school students.* Skills. Budapest: Nemzeti Tankönyvkiadó.

Davies, F. (1988). Designing a writing syllabus in English for academic purposes: Process and product. *ELT Documents, 129,* 130-142.

Demeter, A. (1998). *Portfolio* [Typescript].

Dheram, P. K. (1995). Feedback as a two-bullock cart: A case study of teaching writing. *ELT Journal, 49,* 160-168.

Dickson, M. (1995). *It's not like that here: Teaching academic writing and reading to novice writers.* Portsmouth: Heinemann.

Dörnyei, Z. (1997). What can researchers offer teachers? The case of motivation. *Novelty, 4* (4), 6-21.

Dörnyei, Z., & Malderez, A. (1997). Group dynamics and foreign language teaching. *System, 25* (1), 65-81.

Eco, U. (1994a). How to eat in flight. In *How to travel with a salmon and other essays* (pp. 13-16). London: Minerva. (Translated by W. Weaver)

Eco, U. (1994b). How to write an introduction. In *How to travel with a salmon and other essays* (pp. 172-175). London: Minerva. (Translated by W. Weaver)

Ellis, R. (1995). Appraising second language acquisition theory in relation to language pedagogy. In G. Cook & B. Seidlhofer (Eds.), *Theory and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 73-89). Oxford: Oxford University Press.

Ellis, R. (1998, April). Second language acquisition research: What's in it for teachers? Plenary paper presented at 32nd IATEFL Conference, Manchester.

Elsbree, L., Bracher, F., & Alitzer, N. (1977). *Heath's brief handbook of usage.* Lexington: Heath.

Evans, V. (1998). *Successful writing proficiency.* Student's book, teacher's book and cassette. Swansea: Express.

Farrell, P. B. (1989). Computers interact with writers and tutors. In P. B. Farrell (Ed.), *The high school writing center: Establishing and maintaining one* (pp. 107-110). Urbana: National Council of Teachers of English.

Farrington, B. (1996). Data-driven learning: A badly needed new horizon for CALL. Paper presented at EuroCALL 96, Szombathely.

Fathman, A. K., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 178-190). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Fekete, H., Major, É., & Nikolov, M. (Eds.). (1999). *English language education in Hungary: A baseline study.* Budapest: British Council Hungary.

Fillmore, C. (1992). "Corpus linguistics" or "computer-aided armchair linguistics." In J. Svartvik (Ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991* (pp. 35-60). Berlin: Mouton de Gruyter.

Flower, L., & Hayes, J. (1981). A cognitive process theory of writing. *College Composition and Communication, 32,* 365-387.

Fox, G. (1987). The case for examples. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 137-149). London: Collins.

Földesi, V. (1998a). A piece of Italy. *Novelty, 5* (2), 48-49.

Földesi, V. (1998b). *Portfolio* [Typescript].

Frankenberg-Garcia, A. (1999). Providing student writers with pre-text feedback. *ELT Journal, 53,* 100-106.

Frodesen, J. (1991). Grammar in writing. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 264-276). Boston: Heinle.

Gadó, L. (1998). Computer at school—Sulinet. Unpublished thesis, Pécs, Janus Pannonius University.

Garside, R. (1997). The robust tagging of unrestricted text: The BNC experience. In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 167-180). London: Longman.

Garside, R., Fligelstone, S., & Botley, S. (1997). Discourse annotation: Anaphoric relations in corpora. In R. Garside, G. Leech, & T. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 66-84). London: Longman.

Garside, R., Leech, G., & McEnery, A. (Eds.). (1997). *Corpus annotation: Linguistic information from computer text corpora.* London: Longman.

Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 83-99). Applied linguistics and language study. London: Longman.

Gere, A. R. (1985). *Writing and learning* (2nd ed.). New York: Macmillan.

Geresdi, A. (1997). 537893, hallo? *Novelty, 4* (2), 65-67.

Gibaldi, J. (1995). *MLA handbook for writers of research papers.* New York: MLA.

Gillard, P., & Gadsby, A. (1998). Using a learners' corpus in compiling ELT dictionaries. In S. Granger (Ed.), *Learner English on computer* (pp. 159-171). Studies in language and linguistics. London: Longman.

Gould, E., DiYanni, R., & Smith, W. (1989). *The act of writing.* New York: Random.

Gowers, E. (1953). *The complete plain words.* London: Pelican.

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective.* Applied linguistics and language study. London: Longman.

Granger, S. (1993). The International Corpus of Learner English. In J. Aarts, P. de Haan & N. Oostdijk (Eds.), *English language corpora: Design, analysis and exploitation* (pp. 57-71). Amsterdam: Rodopi.

Granger, S. (1994). Learner corpus: A revolution in applied linguistics. English Today, 10 (3), 25-29. (ERIC Document Reproduction Service No. EJ 514 886)

Granger, S. (1996). Learner English around the world. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 13-24). Oxford: Clarendon Press.

Granger, S. (1998a). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). London: Longman.

Granger, S. (Ed.). (1998b). *Learner English on computer.* Studies in language and linguistics. London: Longman.

Granger, S. (in press). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology.* Oxford: Oxford University Press.

Granger, S., & Tribble, C. (1998). Learner corpus data in the foreign language classroom: Form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on computer* (pp. 199-208). Studies in language and linguistics. London: Longman.

Gray, J., & Melis, I. (1996). *Little red writing book.* Budapest: Nemzeti Tankönyvkiadó.

Greenbaum, S. (1996a). Introducing ICE. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 3-12). Oxford: Clarendon Press.

Greenbaum, S. (Ed.). (1996b). *Comparing English worldwide: The International Corpus of English.* Oxford: Clarendon Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3., pp. 41-58). New York: Academic Press.

Grundy, P., & Li, V. (1998). Responding to writing: Credible alternatives to the "you write: I correct" syndrome. *Novelty, 5* (3), 7-13.

Hadley, G. (1997). Sensing the winds of change: An introduction to data-driven learning. [Online.] Available http://web.bham.ac.uk/johnstf/winds.htm

Hall, D. (1988). *Writing well* (6th ed.). Glenview: Scott.

Hamp-Lyons, L. (1986). Writing in a foreign language and rhetorical transfer: Influences on raters' evaluations. (ERIC Document Reproduction Service No. ED 348 847)

Hamp-Lyons, L. (1989). The product before: Task-related influences on the writer. *ELT Documents, 129,* 35-46.

Hamp-Lyons, L. (1990). Second language writing: Assessment issues. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 69-87). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Hamp-Lyons, L., & Condon, W. (1993). Questioning assumptions about portfolio-based assessment. (ERIC Document Reproduction Service No. EJ 462 284)

Hamp-Lyons, L., & Heasley, B. (1987). *Study writing: A course in written English for academic and professional purposes.* Cambridge: Cambridge University Press.

Hansen, J. (1987). *When writers read.* Portsmouth: Heinemann.

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. In L. W. Gregg & F. R. Steinberg (Eds.), *Cognitive processes in writing.* Hillsdale: Lawrence Erlbaum.

Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *Modern Language Journal, 80*, 287-308.

Higgins, J. (1991). Looking for patterns. *ELR Journal, 4,* 63-70.

Higgins, J., & Johns, T. (1984). *Computers in language teaching.* London: Collins.

Hoey, M. (1998, April). New perspectives on vocabulary. Paper presented at the 32nd International IATEFL Conference, Manchester.

Hollósy, B. (1996). Compiling a dictionary of academic English: Progress report. In *English studies and the curriculum: Proceedings of the First TEMPUS-JEN Mini Conference* (pp. 33-45). Debrecen: Institute of English and American Studies, Kossuth Lajos University.

Hollósy, B. (1998). Database management of corpora for a dictionary of academic English. In J. Andor, B. Hollósy, T. Laczkó, & P. Pelyvás (Eds.), *The diversity of linguistic description: Studies in the honour of Béla Korponay—A special issue* (pp. 89-111). Debrecen: Kossuth Lajos Tudományegyetem Angol-Amerikai Intézete.

Hoppert, L. (1997). Tutoring non-native writers. *Connections, 1* (7). [Online]. Available http://owl.ccd.cccoes.edu/asc/connections/spring97/onep.htm (The Community College of Denver Writing Center Journal)

Horváth, J. (1994a, October). EFL in hypertext. Paper presented at EuroCALL 94, Kalsruhe.

Horváth, J. (1994b). The Pécs corpus. Paper presented at the 4th IATEFL-Hungary Conference, Gödöllő.

Horváth, J. (1995a). EFL corpus online. Paper presented at the 5th IATEFL-Hungary Conference, Szombathely.

Horváth, J. (1995b, October). Focusing on university essay assignments: The Pécs experience. Paper presented at the 5th IATEFL-Hungary Conference, Szombathely.

Horváth, J. (1995c). JPU essays 2.1 [Computer program]. Available ftp://sumex-aim.stanford.edu/info-mac/applications

Horváth, J. (1996a, December). Angol szakos hallgatók esszéírásának értékelése [The assessment of essay writing by English majors]. Paper presented at FEFA "Megmérettetések" Conference, Pécs.

Horváth, J. (1996b). The assessment of essay-writing skills in the first-year proficiency test. In M. Nikolov & J. Horváth (Eds.), *Learning lessons: Innovations in teacher education and assessment* (pp. 88-106). Pécs: Lingua Franca Csoport.

Horváth, J. (1997a). Komputeres nyelvtanulási projekt. *Szaknyelvoktatás, 2,* 19-31.

Horváth, J. (1997b). Mark our word: Applying grader comments in writing and research courses. In M. Kurdi & J. Horváth (Eds.), *HUSSE papers 1997: Proceedings of the third biennial conference* (pp. 379-384). Pécs: University Press Pécs.

Horváth, J. (1998a). The key word in context: Corpus [Review]. *Novelty, 5* (1), 67-70.

Horváth, J. (1998b). Port Folio: Writing tips for advanced students. *Novelty, 5* (3), 14-18.

Horváth, J. (1998c). Processing corpus data on the internet. *Nyelvvizsga Fórum, 1* (2), 3-6.

Horváth, J. (1998d). Razvijane vjestine pisanja esaja u nastavi engleskoga kao stranoga jezika: Mapa radova [Developing and assessing essay writing skills in EFL: The portfolio approach]. In A. Peko & I. Vodopija (Eds.), *Vrjedovanje obrazobanja: Zbornik radova s medunarodnoga znanstvenoga skupa* [Proceedings of the International Conference on Assessment in Education] (pp. 190-196). Osijek: Sveuciliste Josipa Juirja Strossmayera u Osijeku, Pedagoski Fakultet.

Horváth, J. (1998e). Take-off [Online]. Available http://ipisun.jpte.hu/~joe/take-off

Horváth, J. (1998f, October). Take-off: Writing tips for intermediate and advanced students. Paper presented at the 8th IATEFL-Hungary Conference, Szeged.

Horváth, J. (1999a). Collins COBUILD home page. *Modern Nyelvoktatás, 5* (1), 83-84.

Horváth, J. (1999b, October). Commenting on student writing: A source of corpus annotation. Paper presented at the 9th IATEFL-Hungary conference, Győr.

Horváth, J. (1999c, January). Concordance-based study guides in university writing courses. Paper presented at HUSSE 4 Conference, Budapest.

Horváth, J. (1999d). EZ APA 2.0 [Shareware computer program].

Horváth, J. (1999e). EZ MLA 2.0 [Shareware computer program].

Horváth, J. (1999f). [Review of the book *Language learning online*]. *Language Learning and Technology, 2* (2), 28-30.

Horváth, J. (2001, June). Pedagogical annotation of learner corpora. Paper presented at the Empirically Based Approaches to Linguistic Description Thematic HUSSE Conference, Debrecen.

Horváth, J. (in preparation). *Take-off: A writing resource book for Hungarian university students of English.*

Horváth, J., Nikolov, M., & Turner, S. (1997). *Guidelines for writing theses in the English department of JPU.* Pécs: University Press Pécs.

Huang, S. (1995). The efficacy of using writing groups to help students generate ideas for writing and revise drafts on an EFL university writing class. (ERIC Document Reproduction Service No. ED 397 654)

Hubbard, F. A. (1988). *How writing works: Learning and using the processes.* New York: St. Martin's.

Hughes, G. (1997). Developing a computing infrastructure for corpus-based teaching. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 292-307). Applied linguistics and language study. London: Longman.

Hult, C. A. (1986). *Researching and writing: An interdisciplinary approach.* Belmont: Wadsworth.

Hunston, S., & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistics, 19* (1), 45-72.

Hurtt, S. D., & Boylan, B. (Eds.). (1989). *Seventy-five readings: An anthology* (2nd ed.). New York: McGraw Hill.

Hymes, D. H. (1974). *Foundations in sociolinguistics: An ethnographic approach.* Philadelphia: University of Philadelphia Press.

Inkster, G. (1997). First catch your corpus: Building a French undergraduate corpus from readily available textual resources. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 267-276). Applied linguistics and language study. London: Longman

The Internet grammar of English. (1997, October 22). [Online]. Available http://www.ucl.ac.uk:80/english-usage/grammar.htm

Johansson, S., Leech, G. N., & Goodluck, H. (1978). Manual information to accompany the Lancaster—Oslo/Bergen Corpus of British English, for use with digital computers [Online]. Available http://www.hd.uib.no/lob-www.htm

Johns, T. (1991a). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal, 4,* 27-37.

Johns, T. (1991b). Should you be persuaded: Two examples of data-driven learning. *ELR Journal, 4,* 1-16.

Johns, T. (1997a). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). Applied linguistics and language study. London: Longman.

Johns, T. (1997b). Kibbitzer 1: Connotation incessant v. steadfast. [Online]. Available http://web.bham.ac.uk/johnstf/revis001.htm

Jones, C. (1986). It's not so much the program, more what you do with it: The importance of methodology in CALL. *System, 14,* 171-178.

Kail, H. (1988). Narratives of knowledge: Story and pedagogy in four composition texts. *Rhetoric Review, 6,* 178-189.

Kaplan, R. B. (1983). Contrastive rhetoric: Some implications for the writing process. In A. Freedman, I. Pringle, & J. Yalden (Eds.), *Learning to write: First language/second language* (pp. 139-161). London: Longman.

Kaszubski, P. (1997). Writing courses at Adam Mickiewicz University [Online]. Available http://main.amu.edu.pl/~przemka/ifawrit.html

Kaszubski, P. (1998). Enhancing a writing textbook: A national perspective. In S. Granger (Ed.), *Learner English on computer* (pp. 172-185). Studies in language and linguistics. London: Longman.

Kennedy, G. (1998). *An introduction to corpus linguistics.* Studies in language and linguistics. London: Longman.

Kerka, S. (1996). Journal writing and adult learning. *ERIC Digest,* ED399413 96.

Kirk, J. M. (1996). ICE and teaching. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 227-238). Oxford: Clarendon Press.

Kirschenbaum, M. (1998). Intellectual property online: The case of student writing. *Kairos, 3* (1). [Online]. Available http://english.ttu.edu/kairos/3.1/coverweb/mk.html (A Journal for Teachers of Writing in Webbed Environments)

Kiszely, Z. (Ed.). (1998). *A writing handbook with readings.* Székesfehérvár: Kodolányi János Főiskola.

Koster, C. J. (1996). *Statisztika nyelvtanárok és nyelvszakos diákok számára* [Statistics for language teachers and students]. Pécs: JPTE TK Kiadói Iroda.

Kowitz, J., & Carroll, D. (1991). Using computer concordances for literary analysis in the classroom. *ELR Journal, 4,* 135-149.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37-56). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Krashen, S. (1984). *Writing: Research, theory and applications.* Torrance: Laredo.

Krashen, S. (1985). *The input hypothesis: Issues and implications.* Torrance: Laredo.

Krashen, S. (1993). *The power of reading: Insights from the research.* Englewood: Libraries Unlimited.

Krishnamurthy, R. (1987). The process of composition. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 41-61). London: Collins.

Kroll, B. (1991). Teaching writing in the ESL context. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 245-263). Boston: Heinle.

Kurdi, M., & Horváth, J. (Eds.). (1997). *HUSSE papers 1997: Proceedings of the Third Biennial Conference.* Pécs: University Press Pécs.

Labov, W. (1996). When intuitions fail. In L. McNair, K. Singer, L. M. Dobrin, & M. M. Aucoin (Eds.), *CLS 32: Papers from the Parasession on Theory and Data in Linguistics* (pp. 77-105). Chicago: Chicago Linguistic Society.

Lane, B. (1993). *After the end: Teaching and learning creative revision.* Portsmouth: Heinemann.

Larsen-Freeman, D. (1991). Second language acquisition research: Staking out the territory. *TESOL Quarterly, 25,* 315-350.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research.* Applied linguistics and language study. London: Longman.

Lázár A., Péter. (1997). *Wordsmith tools. Modern Nyelvoktatás, 3* (4), 88-90.

Leech, G. (1997a). Introducing corpus annotation. In R. Garside, G. Leech, & T. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 1-18). London: Longman.

Leech, G. (1997b). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1-23). Applied linguistics and language study. London: Longman.

Legett, G., Mead, C. D., Kramer, M. G., & Beal, R. S. (1988). *Prentice Hall handbook for writers.* Englewood Cliffs: Prentice Hall.

Leki, I. (1989). *Academic writing: Techniques and tasks.* New York: St. Martin's.

Leki, I. (1990). Coaching from the margins: Issues in written response. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 57-68). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Leki, I. (1991). Twenty-five years of contrastive rhetoric: Text analysis and writing pedagogies. *TESOL Quarterly, 25,* 123-143.

Leki, I. (1995). Coping strategies of ESL students in writing tasks across the curriculum. *TESOL Quarterly, 29,* 235-260.

Leki, I., & Carson, J. (1997). "Completely different worlds": EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31* , 39-69.

Longman essential activator. (1997). Harlow: Longman.

Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (pp. 53-66). London: Longman.

Madden, C. G., & Rholck, T. N. (1997). *Discussion and interaction in the academic community.* Michigan series in English for academic and professional purposes. Ann Arbor: University of Michigan Press.

Manglesdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *ELT Journal, 46,* 274-284.

Marius, R., & Weiner, H. S. (1985). *The McGraw-Hill college handbook.* New York: McGraw-Hill.

Mark, K. (1997a). Autonomy, multimedia, and university foreign language education. *Journal of Humanities, 4,* 91-104.

Mark, K. (1997b). Curriculum development issues and the need for a learner corpus. In V. Berry, B. Adamson, & W. Littlewood (Eds.), *Applying linguistics: Insights into language education* (pp. 139-150). Hong Kong: University of Hong Kong.

Mark, K. (1997c). A Japanese learner corpus [Manuscript].

Mark, K. (1998, April). A Japanese learner corpus and its uses. Paper presented at the 32nd IATEFL Conference, Manchester.

McCrimmon, J. M. (1984). *Writing with a purpose* (8th. ed.). Boston: Houghton Mifflin.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics.* Edinburgh: Edinburgh University Press.

McMahan, E., & Day, S. (1984). *The writer's rhetoric and handbook.* New York: McGraw-Hill.

Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal, 46,* 340-349.

Medgyes, P. (1994). *The non-native teacher.* MEP monographs. Np.: Macmillan.

Meyers, G. (1997). What do we do here: The multiple literacies of the writing center. *Connections, 1* (7) [Online]. Available http://owl.ccd.cccoes.edu/asc/connections/spring97/fourp.htm

Mindt, D. (1996). English corpus linguistics and the foreign language teaching syllabus. In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey* Leech (pp. 232-247). London: Longman.

Mindt, D. (1997). Corpora and the teaching of English in Germany. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 40-50). Applied linguistics and language study. London: Longman.

Minugh, D. (1997). All the language that's fit to print: Using British and American newspaper CD-ROMs as corpora. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 67-82). Applied linguistics and language study. London: Longman.

Mosher, J. (1998). Responding to student papers: Responses to avoid and productive advice to give. *Word Works, 90-91* [Online]. Available http://www.idbsu.edu/wcenter/ww9091.html

Myers, S. (1997). Teaching writing as a process and teaching sentence-level syntax: Reformulation as ESL composition feedback. *TESL-EJ, 2* (4), A-2.

Nikolov, M. (1999). Classroom observation project. In H. Fekete, É. Major & M. Nikolov (Eds.), *English language education in Hungary: A baseline study* (pp. 221-245). Budapest: British Council Hungary.

Nikolov, M., & Turner, S. (1997). *Guidelines for writing theses in the English department.* Pécs: English Department of JPU.

Novelty (1997). Volume 4, Number 3.

Owen. C. (1996). Do concordances require to be consulted? *ELT Journal, 50,* 219-224.

Pally, M., & Diallo, A. (1995). The man who mistook "wet paint" for a verb: A chronicle for thinking about language, culture and writing. *TESL-EJ, 2* (2), A-2.

Partington, A. (1996). *Patterns and meaning: Using corpora for English language research and teaching.* Studies on corpus linguistics. Amsterdam: John Benjamins.

Pennington, M. C. (Ed.). (1989). *Teaching languages with computers: The state of the art.* La Jolla: Athelstan.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107-118). London: Longman.

Phillipson, R. (1992a). ELT: The native speaker's burden? *ELT Journal, 46,* 12-18.

Phillipson, R. (1992b). *Linguistic imperialism.* Oxford: Oxford University Press.

Pinker, S. (1994). *The language instinct.* London: Penguin.

Prabhu, N. S. (1995). Concept and conduct in language pedagogy. In G. Cook & B. Seidlhofer (Eds.), *Theory and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 57-71). Oxford: Oxford University Press.

Quirk, R., & Widdowson, H. (Eds.). (1985). *English in the world: Teaching and learning the language and literatures.* Cambridge: Cambridge University Press.

Rackham, J., & Bertagnolli, O. (1988). *From sight to insight: States in the writing process* (3rd ed.). New York: Holt.

Rácz, E. (1998). Reforming university. *Novelty, 5* (2), 52-54.

Raimes, A. (1983a). Anguish as a second language? Remedies for composition teachers. In A. Freedman, I. Pingle, & J. Yalden (Eds.), *Learning to write: First language/second language* (pp. 258-272). London: Longman.

Raimes, A. (1983b). *Techniques in teaching writing.* Teaching techniques in English as a second language. New York: Oxford University Press.

Raimes, A. (1991). Out of the woods: Emerging traditions in the teaching of writing. *TESOL Quarterly, 25,* 407-430.

Raimes, A. (1992). *Exploring through writing* (2nd ed.) New York: St. Martin's.

Raimes, A. (1996). *Keys for writers: A brief handbook.* Boston: Houghton Mifflin.

Rankin, D. (1972). *Style and structure.* New York: Harcourt.

Renouf, A. (1987a). Corpus development. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 1-40). London: Collins.

Renouf, A. (1987b). Moving on. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 167-178). London: Collins.

Renouf, A. (1997). Teaching corpus linguistics to teachers of English. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 255-266). Applied linguistics and language study. London: Longman.

Research papers. (1998). Online. Available http://ipisun.jpte.hu/~joe/papers

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed.), *Learner English on computer* (pp. 41-52). London: Longman.

Ronesi, L. (1996). Meeting in the writing center: The field of ESL. *TESL-EJ, 1* (2), A-1.

Rósa, G. (1995). Computer assisted language learning (CALL) in EFL teaching. Unpublished MA thesis, Janus Pannonius University, Pécs.

Salamon, G., & Zalotay, M. (Eds., Trans.). (1996). *Hány óra van? Úgy érted, most? What time is it? You mean now?* (5th ed.). Budapest: Biográf.

Sankó, Gy. (1997). Problems and possibilities of integrating information technology into TEFL. *Novelty, 4* (1), 32-42.

Sato, T. (1990). Revising strategies in Japanese students' writing in English as a foreign language. Unpublished Ph.D. dissertation, Indiana University of Pennsylvania.

Schenk, M. J. (1988). *Read, write, revise: A guide to academic writing.* New York: St. Martin's.

Schmied, J. (1996). Second-language corpora. In S. Greenbaum (Ed.), *Comparing English worldwide: The International Corpus of English* (pp. 182-196). Oxford: Clarendon Press.

Schubert, G. (1996). An introduction to xyllabology. *Novelty, 3* (1), 65-67.

Schultz, J. M. (1994). Stylistic reformulation: Theoretical premises and practical applications. *Modern Language Journal, 78*, 169-178.

Scott, M. (1996). *Wordsmith tools* [Computer program]. Oxford: Oxford University Press.

Scott, M., & Johns, T. (1993). *Microconcord* [Computer program]. Oxford: Oxford University Press.

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford: Oxford University Press.

Selinker, L. (1992). *Rediscovering interlanguage.* Oxford: Longman.

Sengupta, S. (1998). Peer evaluation: "I am not the teacher." *ELT Journal, 52*, 19-28.

Silva, T. (1990). Second language composition instruction: Developments, issues, and directions in ESL. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 11023). Cambridge applied linguistics. Cambridge: Cambridge University Press.

Silva, T. (1993). Towards an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*, 657-677.

Sinclair, J. (1987a). The nature of the evidence. In J. Sinclair (Ed.), *Looking up: An account of the COBUILD project in lexical computing* (pp. 150-159). London: Collins.

Sinclair, J. (Ed.). (1987b). *Looking up: An account of the COBUILD project in lexical computing.* London: Collins.

Sinclair, J. (1991). *Corpus, concordance, collocation.* Describing English language. Oxford: Oxford University Press.

Sinclair, J. (1996, October). The challenge of corpora in CALL. Paper presented at EuroCALL 96, Szombathely.

Sinclair, J. (1997). Corpus evidence in language description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 27-39). Applied linguistics and language study. London: Longman.

Singh, M. (1992). A student's guide to process writing. *Forum, 30* (2), 44-45.

Smalzer, W. (1996). *Write to be read: Reading, reflection, and writing.* Cambridge: Cambridge University Press.

Smith, M., & Smith, G. (1990). *A study skills handbook.* Oxford: Oxford University Press.

Sotiriou, P. E. (1984). *Integrating college study skills: Reasoning in reading, listening and writing.* Belmont: Wadsworth.

Stevens, V. (1989). A direction for CALL: From behavioristic to humanistic courseware. In M. C. Pennington (Ed.), *Teaching languages with computers: The state of the art* (pp. 29-43). La Jolla: Athelstan.

Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. *ELR Journal, 4,* 47-62.

Strunk, W., & White, E. B. (1979). *The elements of style.* New York: Macmillan.

Stubbs, M. (1995). Corpus evidence and norms of lexical collocation. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 245-256). Oxford: Oxford University Press.

Stubbs, M. (1996). *Text and corpus analysis.* Oxford: Blackwell.

Sullivan, N. (1998). Developing critical reading and writing skills: Empowering minority students in a networked computer classroom. In J. Swaffar, S. Romano, P. Markley, & K. Arens (Eds.), *Language learning online: Theory and practice in the ESL and L2 computer classroom* (pp. 41-55). Austin: Labyrinth Publications.

Sullivan, N., & Pratt, E. (1996). A comparative study of two ESL writing environments: A computer-assisted classroom and a traditional oral classroom. *System, 29,* 491-501.

Summer Institute of Linguistics. (1992). *Concordance generating program.* [Shareware]. Dallas: Authors.

Summers, D. (1996). Computer lexicography—The importance of representativeness in relation to frequency. In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 260-266). London: Longman.

Summers, D. (1998, October). How your students can benefit. Paper presented at the 8th IATEFL-Hungary Conference, Szeged.

Svartvik, J. (1996). Corpora are becoming mainstream. In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 3-13). London: Longman.

Szabó, G. (1996). Effective filter? Proficiency testing of first-year English majors. In M. Nikolov & J. Horváth (Eds.), *Learning lessons: Innovations in teacher education and assessment* (pp. 76-87) Pécs: Lingua Franca Csoport.

Szirmai, M. (2000). The theory and practice of corpus linguistics. Unpublished Ph. D. dissertation, University of Debrecen.

Tantervek [Curricula]. (1998). Pécs: JPTE Bölcsészettudományi Kar.

Tono, Y. (1999, May 1). Using learner corpora for L2 lexicography: Information of collocational errors for EFL learners [Online]. Available http://www.lancs.ac.uk/postgrad/tono/userstudy/LEXICOS6.html

Tribble, C. (1997). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk, & J. Melia (Eds.), *Proceedings of the First International Conference on Practical Applications in Language Corpora.* [Online]. Available http://web.bham.ac.uk/ johnstf/palc.htm

Tribble, C., & Johns, G. (1990). *Concordances in the classroom.* London: Longman.

Tsui, A. B. M. (1996). Learning how to teach ESL writing. In D. Freeman & J. C. Richards (Eds.), *Teacher learning and language teaching* (pp. 97-119). Cambridge language teaching library. Cambridge: Cambridge University Press.

Turi, Sz. (1997). Hypertext in literature studies at university. *Novelty, 4* (1), 43-47.

Vadon, B. (1998). *Portfolio* [Typescript].

Villamil, O. S., & de Guerro, M. C. M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics, 19,* 491-514.

Virtanen, t. (1998). Direct questions in argumentative student writing. In S. Granger (Ed.), *Learner English on computer* (pp. 94-106). London: Longman.

Vlaskovits, D. (1996). *Contour* [Computer program]. Pécs: Author.

Wardhaugh, R. (1995). *Szociolingvisztika* [Sociolinguistics]. Budapest: Osiris.

Warschauer, M. (1997). Computer-mediated collaborative learning: Theory and practice. *Modern Language Journal, 81,* 470-481.

Weiner, H. S. (1973). *Creating compositions* (4th ed.). New York: McGraw-Hill.

White, R. V. (1988). Academic writing: Process and product. *ELT Documents, 129,* 4-16.

Widdowson, H. (1979). *Explorations in applied linguistics.* Oxford: Oxford University Press.

Widdowson, H. (1991). The description and prescription of language. In J. E. Alatis (Ed.), *Linguistics and language pedagogy: The state of the art* (pp. 11-24). Washington, D.C.: Georgetown University Press.

Wilson, A. (1997). The automatic generation of CALL exercises from general corpora. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 116-130). Applied linguistics and language study. London: Longman.

Wolff, D. (1993). State of the art: New technologies for foreign language teaching. In *Lingua Delta: Foreign language teaching and the use of new technologies* (pp. 17-27). Brussels: Bureau Lingua.

Young-Scholten, M. (1994). Having it both ways: Individualized reading and writing in large classes. In J. Kohn & D. Wolff (Eds.), *New methodologies in foreign language learning and teaching* (pp. 247-255). Szombathely: Berzsenyi Dániel Tanárképző Főiskola.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly, 19,* 79-101.

Zamel, V. (1992). Writing one's way into reading. *TESOL Quarterly, 26,* 463-485.

Zinsser, W. (1988a). *On writing well* (3rd ed.). New York: Harper.

Zinsser, W. (1988b). *Writing to learn.* New York: Harper.

Zinsser, W. (1998). *On writing well* (6th. ed.). New York: Harper.

Zirinsky, D. (1995). "How long does it have to be?" Helping students understand research writing. *Word Works, 72* [Online]. Available http://www.idbsu.edu/wcenter/ww72.html

APPENDIXES

APPENDIX A

Essay titles and themes suggested by the ICLE developers (courtesy of Sylviane Granger)

Crime Does Not Pay

The prison system is outdated. No civilised society should punish its criminals: It should rehabilitate them.

Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.

A man/woman's financial reward should be commensurate with their contribution to the society they live in.

The Role of Censorship in Western Society

Marx once said that religion was the opium of the masses. If he [were] alive at the end of the 20th century, he would replace religion with television.

All armies should consist entirely of professional soldiers: There is no value in a system of military service.

The Gulf War has shown us that it is still a great thing to fight for one's country.

Feminists have done more harm to the cause of women than good.

In his novel Animal Farm, George Orwell wrote, "All men are equal: but some are more equal than others." How true is this today?

In the words of the old song, "Money is the root of all evil."

Europe

In the 19th century, Victor Hugo said: "How sad it is to think that nature is calling out but humanity refuses to pay heed." Do you think it is still true nowadays?

Some people say that in our modern world, dominated by science technology and industrialisation, there is no longer a place for dreaming and imagination. What is your opinion?

APPENDIX B

Janus Pannonius University, English Department
Fall 1996

ANG 1601 Formal Writing

Tutor Horváth József
2 credits
Prerequisites: none

This course is a prerequisite for the
First English Linguistics Examination.

Classes: Monday, 8-9:30, D 545
Office hours: Thursday, 8-9:30 and Friday, TBA, D 546
Students are encouraged to make an appointment with the tutor in advance.
Office hours will be available for private tutorials and book loans. Office phone: 314-
714. E-mail: hjoe@btkstud.jpte.hu. Course materials on the World Wide Web: http://www.jpte.hu/~joe/fall96/welcome.html

Course Description

This Essay Writing course offers students the opportunity to further develop skills required in university formal writing assignments: short in-class expository writing,
longer take-home assignments, and the thesis. The course will enable students to

understand the nature of various assignments and take appropriate writer's decisions based on this understanding
relate to different types of audience and draft essays for these audiences
develop personal strategies that will carry essay theses
review and consolidate spelling, punctuation and formatting styles
appreciate, analyze and comment on other students' writing in a professional manner

Course Themes

Reading writing
The photography of writing
Types of academic writing
Style manuals
Understanding the assignment
Developing a research plan
Subjectivity vs. objectivity
Focus, Accuracy, Vocabulary, Paragraph and Essay Organization

Responding to commentary
Drafting and revising
Presenting the essay
Plagiarism: what it is and how to avoid it
The mechanics: punctuation, dangling modifiers, sentence fragments, run-ons and comma splices

Requirements

Class attendance 30 marks
Take-Home Draft 1 (500 words) 10 marks. Due: October 14
Take-Home Draft 2 (1,000 words) 20 marks. Due: November 11
Take-Home Draft 3 (2,000 words) 30 marks. Due: December 2.
In-class essay (500 words) 20 marks. May be taken on November 25 or December 9.
Essay-writing notebook 10 marks. Continuous; will be collected on December 2.

A grade (1 to 5) will be awarded to each student, based on the quality of the assessed
work. A top grade of 5 will be awarded to students collecting over 100 marks. The course may be dropped before September 30.

Submission guidelines

Each take-home assignment will be submitted printed and on disk in one of the following formats: WordPerfect 5 or 6, or Word 2, 5, or 5. For the printed version, follow the master essay for this course. Students are well advised to join one of the word processing sessions offered on campus.

Required Readings

Horváth, J. (ed) 1995. *Janus Pannonius University Essays.* University STUD network.
j:\taninfo\angol\joe\96fall\1601\
Hubbard, F.A. 1988. *How Writing Works. Learning and Using the Processes.* New York: St. Martin's Press
Marius, R. & Weiner, H.S. 1985. *The McGraw-Hill College Handbook.* New York:
McGraw-Hill Book Company
McGrimmon, J.M. 1984. *Writing With a Purpose. Eighth Edition.* Boston: Houghton
Mifflin Company
Nikolov, M & Turner, S. 1996. *Guidelines for Writing Theses in the English Department.* Pécs: Janus Pannonius University
Schubert, G. 1996. 'Introduction to Xyllabology.' *NovELTy.* Volume 3:1 pp 65-7.

Additionally, students will read the drafts of other participants. These drafts will be available in print and electronic form.

APPENDIX C

ANG 1601 Writing and Research Skills

Tutor: Horváth József
Credits: 2
Meetings: Monday 0900 to 1300
Venue: Arizona Room (JPU Central Library, 3rd floor, Szepesy Ignác utca)

Aims

This course aims to enable students to develop writing and research skills and strategies that will empower them to achieve success in academic discourse. Central to the course is the communication of what constitutes academic writing and how such writing is conceived, structured, edited, drafted and presented. The course will assist students in communicating relevant ideas and research findings in various types of academic discourse, such as descriptive essays, review essays and research papers.

Class procedures

Students will be invited to participate in individual, pair and group activities that will be structured so that they can build on what skills they have already acquired. Activities and tasks will include predicting. drafting, interviewing, observing, completing, editing and reflecting in various oral and written formats.

Course themes

Reading styles
Writing styles
Processes, purposes and preferences
Plain English
Content and form
Readers and editors
Academic assignments
Peer editing
Group writing
The miniature essay
The research paper
The mechanics
Plagiarism and how to attack it
Writing for an audience
Packaging and selling

Required Readings

Hubbard, F. A. (1988). *How writing works. Learning and using the processes.* New York: St. Martin's.

Horváth, J. (1996). The assessment of essay writing skills in the first-year proficiency test. In M. Nikolov & J. Horváth (ed.), *Learning lessons: Innovations in teacher education and assessment* (pp. 88-106). Pécs: Lingua Franca Csoport.

Hurtt, S. D. & Boylan, B. (Eds.). (1989). *Seventy-five readings: An anthology.* (2nd ed.). New York: McGraw-Hill Book Company.

Marius, R. & Weiner, H.S. (1985). *The McGraw-Hill college handbook.* New York: McGraw-Hill. 598-620.

Nikolov, M. & Turner, S. (1997). *Guidelines for writing theses in the English Department.* (Ver. 1.4). Pécs: Janus Pannonius University.

Pinker, S. (1994). *The language instinct.* London: Penguin. 370-403.

Schubert, G. (1996). Introduction to xyllabology. *NovELTy*, 3 (1), 86-88.

Recommended Readings

Arnaudet, M.L. & Mary E. B. (1990). Paragraph development. (2nd ed.). Englewood Cliffs: Prentice Hall Regents.

Berry, A. (1994). *The research project. How to write it.* New York: Routledge.

Gray, J. and Melis, I. (1996). *Little red writing book.* Budapest: Nemzeti Tankönyvkiadó.

Assessment

Workshop Participation (discussing, brainstorming, group writing, presenting and commenting) 20 marks

Writing Portfolio
- drafts 10 marks
- peer editing 10 marks
- essay on tape 10 marks

Research Paper 20 marks

Writing Terms and Processes Quiz 10 marks

Assignments

Students will receive a calendar of dates when assignments are due. The dates will be negotiated in the second class and the calendar signed by each student.

Copyright

All student material generated in this course will be the copyright of the individual authors. The tutor will, however, solicit contributions to collections to be published during and after the course.

APPENDIX D

Janus Pannonius University, English Department
Fall Semester 1997

ANG 1602 Writing and Research Skills:
The Fitness Center

Classes: Monday
11-1230
Place: D 555

Tutor: Horváth József
hjoe@btkstud.jpte.hu
Phone: 314-714

Library call number of references and manuals: 802
Internet access to course materials:
http://ipisun.jpte.hu/~joe/welcome.html

Welcome to the Fitness Center. As you might know, this is the place where I will take care of putting your writing skills into good shape by allowing you to work out and get the right amount of nutrition and protein. To become a lean, strong, and energetic sportsperson in academic writing, you will do some physical, mental, grammar and interpersonal activities.

Workout schedule

Course themes
Introductions
Writing quizzes
Five basics
Simplicity vs. clutter
Layers of text—The verbal approach
The glamour of grammar
Plain English
Me and my reader
The secret of introductions
Body text
Effective conclusions
Revision techniques
Writing for the mind, eye, and ear
The active voice
Punctuation
Discovery and research
Planning research
Methods, Results and Discussion
Sources and References; MLA and APA
Rhetoric
Fallacies
Packaging the Paper

Writing Portfolio

Keep a record of your writing in a dedicated notebook. I will ask you to consider sharing this portfolio with other students and with me. Choose at least eight from the following options but feel free to add pieces of your own devising.

* The lyrics of a rap song
* An introduction to a favorite novel, poem, short story, play or other text
* An afterword for the same text

* Three in one
* A high school graduation ceremony speech
* A travel guide to your home town
* A travel guide to JPU campus
* A travel guide to a Pécs site
* A description of how a house, a bridge, a television set, a book, a pen, a pullover, or some other object or construction is made
* The revised news story
* A travel guide to any place you've been to
* Profile of a person, based on interview
* A description of how you make coffee, shave, or apply make-up

* Truncate it: make each word longer than two syllables one syllable shorter in any text you choose
* A learning experience: personal essay on tape

Required Readings

Besides reading and using class handouts, a thesaurus, and former students' papers, you will need to read, evaluate and use the following resources:

Hubbard, F. (1988). *How writing works: Learning and using the processes.* New York: St. Martin's.
Hurtt, S. D., & Boylan, B. (Eds.). (1989). *Seventy-five readings: An anthology* (2nd ed.). New York: McGraw-Hill.
Kurdi, M., & Horváth, J. (Eds.). (1997). *HUSSE papers 1997: Proceedings of the third biennial conference.* Pécs: University Press Pécs.
Novelty. (1997). Volume 4, Number 3.
Strunk, W., & White, E. B. (1979). *The elements of style.* New York: Macmillan.
Zinsser, W. (1988). *On writing well.* New York: Harper.

Recommended Readings

American Psychological Association. (1994). *The publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
Gibaldi, J. (1995). *MLA handbook for writers of research papers.* New York: MLA.
Gray, J., & Melis, I. (1996). *Little red writing book.* Budapest: Tankönyvkiadó.
A writer's handbook or manual published after 1985 (for instance by St. Martin's or McGraw-Hill.)

Assessment

You will receive a grade at the end of this course. The grade will be a composite of the marks I have awarded to you:

Attendance and participation: 10 marks
Portfolio: 20 marks
Writing Quiz: 5 marks
Research paper: 10 marks

Note that you will have the option of rewriting the paper. If you choose to do so, the mark you receive on the last version will be incorporated in your final grade.

Research paper submission format

Printed and on computer disk, in WordPerfect or Word 2, 5, 6, or 7 format. For more details on the paper, please refer to the Submission Guidelines.

I wish you a useful and memorable experience in the center.

APPENDIX E

Janus Pannonius University
English Department
Spring 1998

Writing and research skills: The script center

Code: ANG 1601
14-1530
Venue: D 555
Credits: 2
Office hours: TBA

☎ 314-714
Email: hjoe@btkstud.jpte.hu or joe@btk.jpte.hu
Web: http://ipisun.jpte.hu/~joe/98spring/welcome.html

Welcome to the Script Center. In this course, I will aim to enable you to write in fluent, accurate, and plain English. I encourage experimentation with topics, genres, audiences, and purposes. During the sessions and out-of-class meetings, you will practice and improve your planning, sequencing, presenting, packaging, and editing skills. By the end of the course you can expect to have become a more proficient writer of the following types of texts: descriptive essays, reviews, timed exam essays, and research papers.

❶ *I'm not young enough to know everything.* (Sir James Barrie) Introduction: Writers' workshop
❷ *If we had a little ham we could have some ham and eggs if we has dome eggs.* (Carl Sandburg) The English sentence: Balance and grace
❸ *I'm an instant star. Just add water and stir.* (David Bowie) How to write a unified paragraph: Five easy steps ®
❹ *Siberia with palms.* (Unknown, about Hollywood) My dictionary
❺ *The cube* (Ilona Leki) Concrete language ®
❻ *I'm writing this letter slowly because I know you cannot read fast.* (Anthony Denny) Purpose and audience: My voice and my reader √
❼ Take-off and touch-down: Coherence and cohesion ®
❽ Ketchup and mayonnaise
❾ *If you don't know, why ask?* (John Cage) Summary of essay writing: The road ahead ®
❿ *I don't know who's ahead—it's either Oxford or Cambridge.* (John Snagge) Field work; The research paper
❶❶ *Two psychiatrists meet. One says: You're feeling fine, how am I?* (Unknown) The what, the why, and the how ®
❶❷ *I'm not a snob. Ask anybody. Well, anybody who matters.*

(Simon Le Bon) Rhetoric and referencing √
❶❸ *The auk flies backward so as to see where it's been.* (Carl Sandburg) Presentation of first-draft of papers; Assessment

Required Readings ❐

Three or more other students' portfolios.

Your favorite English thesaurus.

Your favorite monolingual dictionary.

A college writing handbook.

American Psychological Association. (1994). *The publication manual of the American Psychological Association* (4th ed.). Washington, DC: Author.
Eco, U. (1994a). How to eat in flight. In *How to travel with a salmon and other essays* (pp. 13-16). London: Minerva. (Transl. W. Weaver).
Eco, U. (1994b). How to write an introduction. In *How to travel with a salmon and other essays* (pp. 172-175). London: Minerva. (Transl. W. Weaver).
Geresdi, A. (1997). 538893, hallo? *Novelty, 4* (2), 65-67.
Gibaldi, J. (1995). *Handbook for writers of research papers* (4th ed.). New York: Modern Language Association.
Horváth, J., Nikolov, M., & Turner, S. (1997). *Guidelines for writing theses in the English Department of JPU.* Pécs: University Press Pécs.
Salamon, G., & Zalotay, M. (Eds.). (1996). *Hány óra van most? Úgy érted, most? What time is it? You mean now?* (5th ed.). Budapest: Biográf.
Schubert, G. (1996). An introduction to xyllabology. *Novelty, 3* (1), 65-67.
Smalzer, W. (1996). *Write to be read: Reading, reflection, and writing.* Cambridge: Cambridge University Press.
Zinsser, W. (1988). *On writing well: An informal guide to writing nonfiction.* New York: Harper.

Assessment ✍

Attendance and participation	10 points
Essays on readings ®	2 points each (see separate info sheet)
Portfolio √	10 points (see separate info sheet)
Research paper	10 points (see separate info sheet)

APPENDIX F

Janus Pannonius University
Department of English Applied Linguistics
Fall 1998

ANG 1601 Writing and Research Skills

Tutor: Horváth József
Phone: 314-714
Email: joe@btk.jpte.hu or
hjoe@btkstud.jpte.hu

Credits: 2
Classes: Wednesday, 10-1130
Office hours: Monday, 11-12;
Tuesday, 9-10; Thursday, 9-11

Description

Welcome to the writing center. This course will aim to enable you to present your ideas, opinions, and observations in personal narrative and descriptive essays, as well as to formulate a manageable research question and develop a small-scale project. You will read and learn about the content, structure, and presentation requirements of academic writing.

I encourage experimentation with topics, genres, audiences, and purposes. During sessions and office hour meetings, you will practice to improve your planning, sequencing, presenting, packaging and editing skills. On a number of occasions, guest students will co-facilitate activities. By the end of the course you can expect to have become a more proficient writer of crisp, fluent, and accurate texts.

For each class, there will be at least one non-graded written assignment to be developed later in the portfolio. Please make photocopies of these scripts so that you can show them to your classmates, other students at JPU, and possibly to international students.

Assessment

Participation	10 points
Test	5 points (to be given on November 4)
Portfolio	10 points (to be finalized by November 11)
Research paper	10 points (first draft to be submitted by December 2)

For the test, the portfolio, and the research paper, I will distribute separate information sheets.

Readings

Babarci Bulcsú. "In the Cocoon of My Room." *Novelty* 5.3 (1998): 50-51.
Bacskay Katalin. *Portfolio.* 1998.
Demeter Andrea. *Portfolio.* 1998.
Földesi Virág. "A Piece of Italy." *Novelty* 5.2 (1998): 48-49.
———. Portfolio. 1998.
Gibaldi, John. *MLA Handbook for Writers of Research Papers.* 4th ed. New York: MLA, 1995.
Grundy, Peter and Vivian Li. "Responding to Writing: Credible Alternatives to the 'You Write: I Correct' Syndrome." *Novelty* 5.3 (1998): 7-13.
Horváth József. "Port Folio: Writing Tips for Advanced Students." *Novelty* 5.3 (1998): 14-18.
———, Nikolov Marianne and Sarah Turner. *Guidelines for Writing Theses in the English Department of JPU.* Rev. ed. Pécs: UP Pécs, 1998.
Rácz Emese. "Reforming University." *Novelty* 5.2 (1998): 52-54.
Research Papers. Online. Available http://ipisun.jpte.hu/~joe/papers. 1998.
75 Readings: An anthology. 2nd ed. New York: McGraw, 1989.
Smalzer, William. *Write to Be Read: Reading, Reflection, and Writing.* Cambridge: Cambridge UP.
Vadon Balázs. *Portfolio.* 1998.
Zinsser, William. *On Writing Well: An Informal Guide to Writing Nonfiction.* New York: Harper, 1988.

Note that the libraries in my office, at the department and faculty abound in resource books for writers. Also, if you do not yet have one, you will do well to check out a recent edition of a thesaurus.

Course Themes (write tasks in the spaces)

1 You as the reader	
2 You as the writer	
3 Concrete language and how to cultivate it	
4 Vocabulary work: A moratorium on "very"	
5 The five T-tips	
6 Your voice and your reader	

7 Processes and skills: Identifying, restricting, drafting, revising, and editing	
8 Ketchup and mayonnaise: The intro and the close	
9 Observation and opinion: Preparing for the research paper	
10 Dissecting a paper	
11 Assembling a paper	
12 Referencing and pragmatics	
13 Assessment and further opportunities	

I look forward to the time we will be spending together, and to your ideas and texts.

I wish you a memorable and exciting time in the writing center.

APPENDIX G

Metaphors used in the WRS courses

The photographer

To help students visualize writing processes, I introduced the familiar activities of taking a snapshot. The photographer chooses a subject just like the writer does, and, having the right camera and film, sets exposure time, decides on distance, and clicks. These four phases correspond to narrowing down a subject, exposing it, and deciding on the approach, and making the final focus decision. A number of frames can be taken—as in writing, when false starts may first be seen as bad choices, but which eventually may lead to approximating the image we had in mind when setting out to capture a subject on film.

A photographer need not be a professional one—amateur practice can lead to improvement in the activity, or in one or two skills involved. This meant emphasizing the importance of setting realistic goals: one could shoot a number of badly exposed pictures, but the process can lead to the improvement of distance setting skills.

Subsequent processes of film development and album design correspond to revision and editing. Students were encouraged to set up their own teams of working on their personal essay photography—asking for and giving, in classes and out of class, advice on some of these issues.

The plain flight

The flight metaphor was similar to photography: it, too, described processes, but the main purpose this time was to illustrate the relationship between writer and reader. In a plane flight, the metaphor suggested, the pilot (the writer) took off (introduced a subject) to passengers (readers) who trusted the professionalism and skills of the pilot and who had a definite destination (purpose). The flight is the pilot's responsibility: the route has to be followed (coherent discussion) at the right altitude (sufficient distance from subject), with the plane loaded with the right type of fuel (mechanics, syntax, vocabulary). This was no Boeing jet: passengers at the windows had to have a fine view of the countryside (theme) below. At arrival, the pilot touches down safely (concludes), as the crew (the group) were assisted by ground control (academic requirements) and the co-pilot (peers and the tutor).

A variation of this technique was the train ride: a more mundane experience, this involves the passenger's personal desire to reach a destination after checking timetables, arriving at the station in time, buying a ticket, locating the right car, and enjoying the ride. As opposed to the flight metaphor, the train ride example aimed to communicate the writer's own purposes in discovering a territory, a route, and reaching a relevant goal.

Slim plain English

The slim plain English metaphor aimed to illustrate the need for energetic and strong sentence-level writing. Beginning with the Spring 1997 WRS course, some form of plain English content was always present in the sessions. By the following semester, it had become the central identity of the whole course. As the section on Objectives has shown, goals included putting students' writing skills in "good shape," and helping them become "lean, strong, energetic" sports people by allowing them to "work out." While it can be problematic that the descriptive part of the syllabus contains no concrete academic writing terms, the Course themes section does identify related tasks and concepts: simplicity versus clutter, based on Zinsser (1988, 1998), concern the choices writers make about what degree of formality they envision for their writing. Specifically, the emphasis on short Anglo-Saxon verbs and concrete nouns, features of clutter-free texts, as opposed to a preference of long Latinate expressions and redundancies, qualities of cluttered writing, a symptom of fuzzy thinking.

Arguably, much contemporary academic writing is loaded with clutter, with several authors placing a premium on the long as opposed to the short, the passive rather than the active. I aimed to introduce students to these choices and let them decide which major direction they wished to go. Slim and plain English inhaled and exhaled a lot of oxygen and let the reader breathe, too. By contrast, cluttered text suffocated the reader by a hazardous concentration of cholesterol contained in its fat expressions and constructions.

APPENDIX H

An example of a co-authored essay: We and Our Names

My name is Polgár Judit. Before you ask whether I am the famous world champion of chess, I have to disappoint you. I am not.

My father gave me this name by the name of a Jewish girl. Maybe he didn't think of being another girl who has the same name as his "little angel."

Sometimes I don't like my name because of the silly jokes in the University and in the offices, but other times I like belonging to the group of famous people.

Am I famous? No, I don't think so. I have a funny name, but it hasn't changed me. I am not more valuable than other students, I am not more intelligent than someone else, and I can't give check to the computer in five moves. I am only a person in society.

In the past twenty-one years I have met a lot of people with common names who belonged to the group of Kovács Józsefs, Kiss Katalins, and Horváth Károlys.

Are they worthless? No, they aren't, but many of them run to and fro in the world.

If someone introduces himself to me like "Antal Imrének hivnak," I break into a little smile and memorize his face. He is my friend, he is a member of famous persons' group.

Names!

Your name is you, and my name is me. You have the same name as me and he has the same one as you. We are different, but we are equal.

I am my name, and my name is me. I cannot live without it, and it's not a name without the person. Yet the problem is: there is only one me, but thousands with the same name. Horváth József is among the most common names in Hungary.

Not that it's a big deal. Living, as I have, 39 years does reconcile one to knowing one's name is anything but special.

It does get special when I meet a stranger and he happens to be called—me. Two or three years ago, for example, when I first heard of the email service the White House had introduced, I sent a message to Bill Clinton. I knew one of his secretaries would answer, not him.

And so it was. I got an answer in a couple of days saying my email had been received and that the President was regularly updated on all the mail he was getting.

At the end of the message, there was a request that I should check whether I was on the list attached. Some technical problem occurred as a result of the thousands of messages to the White House, and they were asking me to check whether the right person got the right message.

And so I began scanning the list of names and email addresses for my name. As I was doing this, I noticed dozens of Horváth Józsefs. "Hűha," I said to myself. "Írni kéne nekik—talán ôket is érdekli, ha már így összejöttünk, ki kicsoda."

I sent the same message to all these people in the world, twenty-odd of them. Quite a few answered. There we were, in the US, Hungary, New Zealand, and Poland, all with the same names, all with similar email addresses, all writing to the President of the US.

It was one of the times that made my name special.

And the only time Washington made a notice of me. Me and my name.

APPENDIX I

The theme selection table of the writing test of the Filter Test used in 1999

Title 1	Title 2	Reader's goal
Me And The	Word Processor	to read about a new idea
The Dream And The Reality: The	English Department	to understand information
The Mystery Of The	Soap Opera	to learn about an opinion
The Smart	Library	to apply tips
Designing A Perfect	* *	to read a personal account

First-year students take the proficiency test, which includes a writing component. They are instructed to select one element from each of the three columns, thus making up their titles and the purpose with which a reader will evaluate the text. The last rubric in the second column is left blank so that students can add their own selections to a title. (On the development and validation of this type of test, see Horváth, 1996b; Szabó, 1996.)

APPENDIX J

A copy of the Permission form

Permission

I hereby authorize Horváth József to incorporate my research paper in his collection of students' scripts. I understand that Horváth will use may paper for corpus linguistic study only. I retain the copyright to my work.

Horváth will have the following rights:

1 To list my name and the title of my work in the References section of his Ph. D. dissertation.

2 To quote short passages from my work as illustration. Whenever this is done, my identity shall not be revealed.

3 To list my name in the Acknowledgment section of his Ph. D. dissertation as a contributor to the corpus.

Signed by................................

Signed on (date)..........................

APPENDIX K

The most frequent word forms in the JPU Corpus

the	(32198)
of	(14756)
to	(11597)
and	(10834)
in	(9071)
a	(8526)
is	(6408)
it	(4149)
that	(4123)
i	(3695)
are	(3265)
they	(3195)
not	(3041)
for	(2981)
be	(2916)
this	(2759)
with	(2755)
as	(2732)
was	(2566)
on	(2521)
can	(2214)
or	(2190)
students	(2164)
their	(2081)
have	(1986)
but	(1784)
which	(1754)
one	(1750)
from	(1634)
about	(1612)
an	(1571)
writing	(1552)
by	(1513)
at	(1412)
them	(1349)
there	(1340)
my	(1316)
more	(1312)
were	(1286)
he	(1258)
these	(1250)
we	(1127)
only	(1085)
his	(1061)
all	(1053)
other	(1033)
had	(1015)
some	(990)
what	(959)
essay	(945)
first	(937)
also	(930)
how	(920)
when	(913)
two	(909)
language	(898)
most	(858)
you	(844)
do	(839)
if	(839)
has	(822)
who	(818)
because	(805)
will	(796)
so	(794)
people	(773)
english	(747)
different	(746)
time	(729)
she	(705)
out	(694)
use	(680)
words	(660)
would	(660)
like	(651)
well	(634)
could	(627)
its	(624)
paper	(606)
introduction	(587)
up	(572)
should	(559)
very	(556)
no	(555)
make	(554)
write	(553)
between	(550)
work	(549)
than	(546)
way	(539)
used	(531)
her	(524)
text	(524)
reading	(506)
into	(505)
did	(502)
course	(499)
essays	(499)
student	(488)
each	(486)
group	(484)
three	(479)
many	(478)
any	(477)
new	(477)
same	(471)
life	(466)
teacher	(465)
part	(461)
find	(460)
book	(449)
after	(441)
even	(440)
topic	(434)
word	(428)
been	(425)
teachers	(425)
information	(424)
important	(418)
our	(412)
may	(406)
research	(396)
sentences	(396)
such	(393)
made	(386)
get	(385)
good	(384)
too	(384)
question	(379)

another	(374)
found	(374)
news	(372)
where	(362)
both	(361)
number	(359)
tasks	(357)
know	(354)
sentence	(352)
task	(350)
hungarian	(349)
much	(349)
style	(348)
written	(348)
second	(345)
skills	(345)
help	(343)
those	(340)
reader	(338)
world	(337)
story	(335)
does	(329)
me	(325)
main	(322)
own	(318)
questions	(318)
paragraph	(316)
then	(315)
writer	(314)
your	(312)
according	(311)
give	(310)
conclusion	(303)
four	(299)
ideas	(297)
read	(296)
see	(295)
always	(293)
every	(292)
given	(292)
ones	(290)
point	(288)
example	(285)
think	(285)
papers	(284)
results	(284)
texts	(282)
year	(282)
readers	(280)
articles	(277)
while	(275)
order	(273)
during	(272)
just	(271)
being	(270)
why	(269)
form	(268)
vocabulary	(268)
teaching	(265)
last	(264)
since	(262)
through	(262)
grammar	(261)
attention	(259)
learners	(259)
present	(259)
him	(257)
learning	(256)
must	(252)
general	(249)
verbs	(248)
without	(248)
lot	(247)
computer	(245)
long	(245)
whole	(245)
author	(244)
day	(244)
type	(244)
university	(244)
however	(243)
short	(243)
got	(242)
items	(239)
often	(239)
take	(239)
content	(237)
five	(236)
others	(236)
knowledge	(235)
activities	(233)
aim	(233)
class	(233)
school	(233)
exercises	(232)
place	(231)
us	(230)
kind	(227)
title	(226)
fact	(225)
foreign	(225)
means	(225)
able	(224)
great	(224)
article	(223)
method	(223)
following	(222)
show	(221)
using	(220)
children	(219)
parts	(218)
topics	(218)
times	(216)
mistakes	(215)
case	(214)
practice	(214)
among	(213)
still	(213)
events	(212)
need	(212)
test	(212)
listening	(210)
useful	(210)
end	(209)
process	(209)
several	(209)
analysis	(208)
problems	(208)
personal	(207)
something	(207)
want	(207)
person	(206)
types	(205)
years	(204)
am	(202)
cannot	(202)
learn	(202)
third	(202)
before	(201)
writers	(201)
enough	(199)
certain	(198)
put	(198)
table	(198)
british	(196)
role	(196)
wrote	(196)

although	(193)
asked	(193)
paragraphs	(193)
problem	(192)
rather	(191)
usually	(191)
best	(190)
here	(190)
difficult	(189)
page	(189)
subject	(189)
level	(187)
based	(186)
discussion	(186)
opinion	(185)
system	(185)
view	(185)
whether	(185)
having	(182)
stories	(182)
possible	(181)
might	(180)
category	(179)
points	(179)
next	(178)
answers	(177)
unit	(177)
answer	(176)
groups	(175)
real	(175)
books	(173)
introductions	(173)
over	(173)
thesis	(173)
idea	(172)
together	(172)
few	(169)
coffee	(168)
said	(167)
speaking	(166)
though	(166)
clear	(165)
gave	(165)
become	(164)
feel	(164)
things	(164)
dallas	(163)
less	(163)
similar	(163)
quite	(162)
family	(161)
go	(161)
thoughts	(161)
again	(158)
computers	(158)
now	(158)
say	(158)
sometimes	(158)
culture	(157)
categories	(156)
down	(156)
structure	(156)
back	(155)
letter	(155)
either	(154)
especially	(154)
exercise	(154)
gives	(154)
really	(154)
interesting	(151)
simple	(151)
thought	(151)
data	(150)
makes	(150)
never	(150)
participants	(150)
american	(149)
beginning	(149)
far	(149)
better	(148)
focus	(148)
home	(148)
right	(147)
study	(147)
experience	(146)
little	(146)
result	(146)
hungary	(144)
shows	(144)
meaning	(142)
works	(142)
authors	(141)
britain	(141)
six	(140)
ways	(140)
seems	(139)
started	(139)
characters	(138)
follow	(137)
connection	(136)
section	(136)
situation	(136)
war	(136)
called	(135)
common	(135)
forms	(135)
members	(135)
further	(134)
itself	(134)
material	(134)
thus	(134)
already	(133)
power	(133)
special	(133)
examples	(132)
grammatical	(132)
name	(132)
reason	(132)
going	(131)
hand	(130)
later	(130)
look	(130)
picture	(130)
cases	(129)
develop	(129)
differences	(129)
wanted	(129)
intermediate	(128)
old	(128)
political	(128)
towards	(128)
importance	(127)
relationship	(127)
effective	(125)
themselves	(125)
understand	(125)
within	(124)
almost	(123)
behaviour	(123)
interest	(123)
mentioned	(123)
activity	(122)
come	(122)
length	(122)
choice	(121)
correct	(121)
tutor	(121)

classroom	(120)
expressions	(119)
purpose	(119)
cards	(118)
least	(118)
became	(117)
various	(117)
countries	(116)
easy	(116)
friends	(116)
letters	(116)
child	(115)
choose	(115)
chosen	(115)
development	(115)
future	(115)
human	(115)
ten	(115)
thing	(115)
change	(114)
considered	(114)
everything	(114)
units	(114)
giving	(113)
set	(113)
social	(113)
talk	(113)
writings	(113)
exam	(112)
specific	(112)
therefore	(112)
values	(112)
chose	(110)
features	(110)
passive	(110)
state	(110)
try	(110)
communication	(109)
description	(109)
examined	(109)
major	(109)
money	(109)
provide	(109)
against	(108)
composition	(108)
man	(108)
took	(108)
under	(108)
contains	(107)
cultural	(107)
high	(107)
previous	(107)
small	(107)
eight	(106)
facts	(106)
film	(106)
history	(106)
item	(106)
past	(106)
around	(105)
body	(105)
techniques	(105)
besides	(104)
front	(104)
india	(104)
negative	(104)
states	(104)
verb	(104)
elements	(103)
influence	(103)
instead	(103)
issue	(103)
proficiency	(103)
sections	(103)
seen	(103)
theme	(103)
yes	(103)
hard	(102)
interested	(102)
making	(101)
program	(101)
related	(101)
age	(100)
country	(100)
finally	(100)
free	(100)
getting	(100)
keep	(100)
kinds	(100)
sense	(100)
tried	(100)

INDEX

A
Alderson, C. 63
Alitzer, N. 87
Allwright, J. 29
Andersen, R. 17
Andor, J. 63, 132
Arnaudet, M. L. 87
Arnold, J. 87
Aston, G. 38, 43
Atkinson, D. 10

B
Bárdos, J. 8
Bartram, M. 27
Beal, R. S. 87
Bello, T.16
Bereiter, C. 9
Berry, R.38, 87
Bertagnolli, O. 88
Biber, D. 35, 37, 45, 53-55
Bloor, M. 13
Blue, G. 18
Bocz, A. 63,100
Botley, S. 47
Boughey, C. 14
Boylan, B. 87
Bracher, F. 87
Brumfit, C. 6
Bullon, S. 36, 38, 41
Burnard, L. 42, 43

C
Carroll, D. 58
Carson, J. 16
Caudrey, T. 12, 13, 19,20
Cavalcanti, M. C. 29
Chandhok, R. 144
Chapelle, C. 56
Chen, J. F. 29
Chomsky, N. 34
Clear, J. 41, 55
Clouse, B. F. 87
Cohen, A. D. 29
Conrad, S. 53-55
Cook, L. 16
Crews, F. 87

Cs
Csapó, J. 63
Csomay, E. 88

D
Davies, F. 13, 14
Day, S. 87, 129
Dheram, P. K. 28
Diallo, A. 146
Dickson, M. 16
DiYanni, R. 87
Dörnyei, Z. 7, 25, 60

E
Eco, U. 87, 132
Ellis, R. 6, 7, 60
Elsbree, L. 87
Erion, P. 144
Evans, V. 87

F
Farrell, P. B. 19
Farrington, B. 57
Fathman, A. K. 28
Fekete, H. 20
Fillmore, C. 35, 37, 102
Fligelstone, S. 47
Flower, L. S. 8
Fox, J. 41
Francis, G. 41, 59
Frankenberg-Garcia, A. 28
Frodesen, J. 21, 22

G
Gadsby, A. 62
Garside, R. 46, 47
Gavioli, L. 59
Gere, A. R. 87
Gibaldi, J. 87
Gillard, P. 62
Goodluck, H. 40
Gould, E. 87
Gowers, E. 87
Grabe, W. 9; 24
Granger, S. 60-62, 101, 102, 105, 127, 128
Gray, J. 87
Greenbaum, S. 38, 43
Grice, H. P. 128
Grundy, P. 29, 84, 87, 88
de Guerro, M. C. M. 25

H
Hadley, G. 59
Hall, D. 87
Hamp-Lyons, L. 10, 87
Hansen, J. 12, 87
Harmer, J. 87
Hayes, J. R. 8
Heasley, B. 87
Hedgcock, J. 30
Higgins, J. 57, 58
Hoey, M. 38
Hollósy, B. 63
Hoppert, L. 16
Huang, S. 26
Hubbard, F. A. 87
Hult, C. A. 87
Hunston, S. 59
Hurtt, S. D. 87
Hymes, D. H. 35

I
Inkster; 46

J
Jamieson, J. 56
Johansson, S. 40
Johns, T. 36, 37, 48, 57-59, 101, 139
Jones, C. 56, 58

K
Kail, H. 16
Kaplan, R. B. 8-10;,24
Kaszubski, P. 62, 102
Kaufer, D. 144
Kennedy, G. 34, 38-41, 44, 55, 59, 99, 112, 114-116
Kerka, S. 16
Kirk, J. M. 59
Kirschenbaum, M. 16
Kiszely, Z. 88
Koster, C. J. 55, 108
Kowitz, J. 58
Kramer, M. G. 87
Krapels, A. R. 10, 11
Krashen, S. 7, 8, 12, 56, 86, 139
Krishnamurthy, R. 41, 46
Kroll, B. 17, 18
Kurdi, M. 87

L
Labov, W. 35, 103
Lane, B. 22-24
Larsen-Freeman, D. 6
Lázár, A. P. 48
Leech, G. 33,40, 46, 47, 59
Lefkowitz, N. 30
Legett, G. 87
Leki, I. 9, 10, 16, 29, 30, 87, 129
Li, V. 29, 84, 87, 88
Long, M. 7

M
Madden, C. G. 87
Major, É. 20
Malderez, A. 25
Mangelsdorf, K. 26
Marius, R. 87

Mark, K. 62, 63, 102
Mary, E. B. 87
McCrimmon, 87
McEnery, T. 35, 38, 44, 46, 55, 108
McMahan, E. 87, 129
Mead, C. D. 87
Medgyes, P. 14
Melis, I. 87
Meyers, G. 16
Miller, D. 144
Mindt, D. 59
Minugh, D. 63
Moon, R. 41
Mosher, J. 29
Morris, J. 144
Myers, S. 29

N

Neuwirth, C. 144
Nikolov, M. 20, 56, 87

O

Owen, C. 37, 59

P

Pally, M. 146
Pennington, M. 57
Phillipson, R. 107
Pinker, S. 87
Prabhu, N. S. 7
Pratt, E. 19

Q

Quirk, R. 107

R

Rackham, J. 87
Raimes, A. 8, 9, 12, 16, 17, 84, 88, 129, 130
Ramanathan, V. 10
Rankin, D. 88
Renouf, A. 42, 46, 63
Reppen, R. 53-55
Rholck, T. N. 87
Ronesi, L. 16

S

Salamon, G. 81, 87
Sankó, Gy. 56
Sato, T. 26
Scardamalia, M. 9
Schenk, M. J. 88
Schmied, J. 37
Scott, M. 48
Seliger, H. W. 34
Selinker, L. 61, 107
Sengupta, S. 25
Shohamy, E. 34
Silva, T. 9-12
Sinclair, J. 33, 35, 36, 38, 41, 42, 46, 57, 59
Singh, 15
Smalzer, W. 84, 87
Smith, M. 87
Smith, G. 87
Sotiriou, P. E. 87
St. John, M. J. 13
Stevens, V. 56-58
Strunk, W. 87, 123, 130
Stubbs, M. 35, 36
Sullivan, N. 15, 19
Summers, D. 36, 37, 45, 46
Svartvik, J. 59

Sz

Szabó, G. 82
Szerdahelyi, J. 88
Szirmai, M. 63

T

Tono, Y. 62
Tribble, C. 58, 60, 62
Tsui, A. B. M. 13, 14
Turner, S. 87

V
Villamil, O. S. 25

W
Walton, R. 27
Wardhaugh, R. 34
Warschauer, M. 15
Weiner, H. S. 87, 88
Whalley, E. 28
White, E. B. 13, 87, 123, 130
Widdowson, H. 36, 37, 107
Wilson, A. 35, 38, 44, 55, 63, 108
Wolff, D. 56

Y
Young-Scholten, M. 18

Z
Zalotay, M. 81, 87
Zamel, V. 5, 12, 28
Zinsser, W. 16, 17, 29, 65, 84, 87, 123, 128,-130, 132
Zirinsky, D. 18

Adonyi Nóra, Andorkó Réka, Babarci Bulcsú, Bacskay Katalin, Baditz Mihály, Bagi Ádám, Bajomi Lázár Péter, Bajusz Judit, Bakonyi Berta, Baksa Gábor, Balassa Gábor, Balázsi Anikó, Balázsik Eszter, Bálint Anita, Balonyi Beáta, Barabás Mariann, Bán Helga, Bárdi Piroska, Barkaszi Orsolya, Barna Eszter, Barsi Gizella, Bauer Katalin, Bátor Mónika, Belső Adrienn, Bencze Krisztina, Benkő Ágnes, Berecz Mónika, Bereczki Veronika, Berke Ildikó, Bernáth Orsolya, Berta Anikó, Béresné Timár Andrea, Bieber László, Billege Virág, Bizse Ferenc, Blasszauer János, Bleyer Júlia, Bodroginé Farkas Piroska, Bogdán László, Bokodi Judit, Bokori Mónika, Bordás Pálné, Borsányi Valéria, Bradák Anna, Brunner Mónika, Bucsányi Erika, Buzási Balázs, Czeglédi Edina, Csapó Angéla, Csákó Fruzsina, Cseke Ildikó, Cserna György, Csiba Eszter, Csikós Erzsébet, Csipkéné Kocsis Éva, Csongor Csilla, Csonka Gábor, Csontos Zsuzsanna, Csordás Dávid, Daróczi Helga, Demeter Andrea, Dékány Renáta, Dévai Dóra, Dittrich Melinda, Ditz Éva, Dombóváriné Csillag Mária, Dózsa Éva, Döbrösy Ildkó, Dunai Péter, Egri Szabolcs, Elek Attila, Engert Ágnes, Enyedi Szilvia, Erdélyi Dóra, Erős Anikó, Ébel Emese, Élő Veronika, Faragó Edit, Farkas Ágnes, Farkas Gabriella, Farkas Roland, Fehér Fatime, Fehér Györgyi, Fejes István, Fekete Bálint, Fekete Tünde, Fodor Márton, Fodorné Horváth Anett, Fónai Annamária, Földesi Virág, Füzér Balázs, Gaálné Vörös Judit, Gadó László, Gallai Ditta, Gallen Mónika, Gavlikné Robotka Orsolya, Gazdag Zsuzsa, Gál Beáta, Gelencsér Péter, Gellénné Káldi Szilvia, Geresdi Annamária, Gergely Anikó, Gerhát Zsófia, Gertner Brigitta, Gombás Éva, Gregóné Szili Katalin, Grosch Andrea, Gulyás Judit, Gyallai Krisztina, Gyenisné Oszbach Mónika, Györök Tímea, Gyulavics Ildikó, Haas Klára, Hadnagyné Kovács Judit, Hajdú Anett, Harmat Ágota, Harmath Virág, Hartmann Edina, Hatta Diána, Hazafy Krisztina, Hegedűs Anita, Hegedűs Éva, Hencsei Vilmos, Herkovits Margit, Hetesi Sándor, Hirth Márkus, Horony Bernadett, Horvai Krisztina, Horváth Eszter, Horváth Gergely, Horváth János, Horváth Judit, Horváth Miléna, Horváth Nóra, Horváth Sándor, Hrivnák Orsolya, Huber Vera, Hubert Andrea, Huszanagics Melinda, Huszár Zsuzsanna, Huszárné Szabó Mária, Illésné Gesztesi Piroska, Inczédy Lászlóné, Iván Lucia, Jaskó Veronika, Jávorszky Mária, Jónás Ildikó, Józsa Anita, Juhász Hedvig, Kalmárné Máriás Ildikó, Kamper Gergely, Karádi Orsolya, Karagits Menyhért, Kaszás Henrietta, Károly Tünde, Kárpáti Ildikó, Kecse-Nagy Katalin, Kertai Hajnalka, Keszthelyi Erika, Kisgyörgy Péter, Kiss Ágnes, Kiss Gabriella, Kiss Judit, Kiss Eszter, Kohajda Edina, Kokas Marianna, Komáromi Csilla, Kondor Edith Mária, Kopjár Ivett, Kovács Attila, Kovács Zsolt, Kópis Andrea, Kövesdi Veronika, Kőberl Kinga, Krammerné Gehring Éva, Krigler Gábor, Kripner Mónika, Láng Katalin, Láng Rita, Lászlai Noémi, Lehmann Magdolna, Leksz András, Lengvári Ágnes, Lénárt Henriett, Luspai Gabriella, Madarász Mónika, Magné Turcsics Tímea, Magony Tímea, Magyar Zsuzsanna, Magyarfalvi Dóra, Makra Melinda, Marosváriné Deák Zsuzsa, Matheisz Éva, Matolcsi László, Mádai Gellért, Márk Mónika, Márki Miklós, Márkus Melinda, Meggyesi Gábor, Merényi Adrienn, Mészáros Krisztián, Mikola Erzsébet, Mojzes Kamilla, Molenaar P. Gábor, Molnár Krisztián, Mondok Ágnes, Muth Imre, Nagy Ágnes, Nagy Borbála, Nagy György, Nagy Elvira, Nagy-Mélykuti Luca, Nemes Dóra, Németh Erika, Németh Piroska, Nógrádi László, Oláh Márta, Öllős Anikó, Paczolay Éva, Palla Lívia, Pandur Zsanett, Papósné Horváth Judit, Papp Árpád, Papp Eszter, Pádár Dénes, Páli Noémi, Perjés Beatrix, Pete István, Pető Orsolya, Petz János, Pércsich Richárd, Pétersz Tamás, Pfeifer Bernadett, Pintér Katalin, Piskó Beáta, Poturak Laura, Povázsai Zita, Pónya Klára, Radnai Tamás, Rácz Emese, Rátkay Orsolya, Rekettye Gábor, Renkó Zsuzsanna, Ribiczei Zsuzsa, Rohonczi Krisztina, Rohonyi Borbála, Roll Eszter, Roza Szilvia, Rozgonyiné Inke Krisztina, Rósa Gézáné, Saskői Kinga, Sata Előd, Sárfy Mária, Schlögl Krisztián, Schönberger Tamás, Schram Ágnes, Schrottiné Tamási Monika, Schubert Gábor, Scsaurszki Rita, Simon Györgyné, Simó Boglárka, Sinka Hilda, Skultéty Judit, Som Katalin, Somorjainé Orosz Andrea, Soós Zoltán, Srankó Nóra, Stemler Kata, Süle Brigitta, Szabados Péter, Szabó Gabriella, Szabó Ildikó, Szabó Imre, Szabó Soma, Szakály Zsófia, Szalontai Mária, Szarka Brigitta, Szántó Györgyi, Szathmári Timea, Szegedi Csilla Eszter, Szécsi Tünde, Sziklai Gábor, Szmolenszky Tünde, Szokol Nóra, Sztanics Lidia, Szűcs Mariann, Szűcs Rita, Takács Anita, Takács Zsolt, Tánczikné Varga Szilvia, Tánczos Balázs, Tarnai Dóra, Tarnai Gábor, Tihanyi László, Tímár Zsuzsa, Tokorcsi Éva, Torma Szilvia, Tóth Andrea, Tóth Ágnes, Tóth Boglárka, Tóth Mónika, Tóth Tamás, Tövisháti Lilla, Tőke Orsolya, Trapl Szilvia, Triznay Zsuzsa, Trócsányi Balázs, Turi Szabolcs, Turiák Tibor, Túri Andrea, Tüű Orsolya, Urbán Miklós, Úz Márk, Vadon Balázs, Vajda Violetta, Váciné Nyári Eszter, Vámosi Gyula, Varga Bernadett, Varga Tamás, Vargáné Uzsoki Márta, Végh Orsolya, Viszlai Eszter, Vízhányó Barbara, Vlaskovits Dóra, Wagner Zita, Wittinger Katalin, Wollent Ida, Záborszky Zoltánné, Zajzon Tamás, Zongor Judit, Zsiborás Eszter, Zsigovicsné Kovács Margit