INTRODUCTION

As the title of this volume implies, its contents focus on the use of statistical analysis in the study of language. None of the authors regard themselves as statisticians, and only one, Prof. Christopher Butler of the University of Nottingham, is generally regarded as a leading authority on the topic of statistics for language studies. Although the first few chapters do cover basic tutorial material, the primary thrust of Chapters Four to Thirteen is to present examples of statistical applications across a broad range of studies including both applied and theoretical topics. The purpose, then, is to focus primarily on some statistical techniques which average researchers in a variety of subfields of language studies in Poland, including advanced students taking their master's seminars, may be able to use to obtain more refined analyses of their data.

To this end, authors of the later chapters have been asked to focus on (a) the appropriateness of the statistical procedures they discuss (i.e., why they use the particular statistics they do) and (b) how the results of the statistical tests are to be interpreted. We agree with Christopher Butler, who states in Chapter One

> It cannot be too strongly emphasized that in order to use statistical techniques sensibly, linguists should make the effort to understand the reasons for the choice of particular methods, the assumptions being made, and the limitations of the techniques available.

We believe that this kind of understanding is more important for students of language studies than an understanding of the theory behind the statistics in terms of their mathematical basis. The latter is the kind of understanding one gains, for

example, by deriving the formulas for doing the actual computations of the statistics from the theoretical formulas that lie behind them. It is a kind of understanding that is of undeniable importance for students of statistics. In contrast, the former kind of understanding - that alluded to in the above quotation - is the kind possessed by the user, rather than the designer, of a tool.

As an analogy, we could compare the kind of understanding required for the design vs. the use of another complex tool, e.g., a car. The kinds of knowledge required of the automotive engineer and the typical driver are quite different. The driver does not need to understand how the brake and steering wheel work at a mechanical level (much less how they are manufactured) in order to coordinate their use appropriately under different driving conditions. For example, when skidding on ice, drivers must learn to turn in the direction of the skid and to brake firmly but not too sharply. Learning when and how to perform this action appropriately is clearly more important than having an understanding of the theoretical basis for the action. In fact, having the theoretical knowledge does not imply that the driver necessarily has the practical ability to correct for a skid on ice. In a similar vein, understanding the engineering principles behind the steering mechanism of a car does not prevent the driver from taking a wrong turn.

What our students need to know about statistics is analogous to knowing when to drive (vs. taking the bus or plane), how to operate a car, and, of course, how to avoid taking those wrong turns. Through this volume we hope to help our students gain this kind of knowledge.

On the other hand, should anybody be interested in the details of the 'engineering principles' underlying theoretical formulas in statistics, a number of books are available for reference [cf. e.g., Guirard 1966, Sambor 1972, Herdan 1960, Clark 1977, Welkowitz, Ewen and Cohen 1976, or Blalock 1976].

Most of the chapters of this book are based on workshops or papers given at the Seminar on Statistics for Language Studies held in Ustronie, Poland, February 28 - March 2, 1988. The Seminar was co-sponsored by the University of Lodz and the Press and Cul

ture Section of the U.S. Embassy, Warsaw. It involved the participation of workshop leaders and presenters from the University of Łódź (UŁ), Adam Mickiewicz University (UAM, Poznań), Jagiellonian University (UJ, Cracow), the University of Gdańsk (UG), and the University of Nottingham (Nottingham, UK), including three Fulbright Lecturers and one British Council Lecturer on assignment in Poland. Workshops have been written up as detailed tutorials or case studies, and all of the papers have been expanded and revised in the versions that follow.

Chapters One through Four represent an extended introduction of sorts. Butler's Chapter One *(A review of statistical techniques in the analysis of linguistic data)* provides a broad overview of the field and lays the primary groundwork for later chapters by introducing most of the basic concepts which form the core of the discussions in later chapters of the book. Coleman's (UŁ) Chapter Two *(Random sampling and the normal distribution)* continues the introduction of basic concepts, focusing especially on the issue of sampling. Weinstein's (UG), Chapter Three *(The use of the t-test in language studies)* is the first to focus in detail on a specific statistic - it does so in the form of an introductory tutorial.

Butler's Chapter Four *(Statistics for linguists: some case studies to illustrate techniques and their applicability)* is a transitional one. It continues the introductory slant of the earlier chapters but at the same time delves into some specific case studies. Coleman's Chapter Five *(IFASTATS: using a microcomputer for statistical analysis in language studies)* also does not fit clearly with the first chapters or with the later ones. The slant is introductory, but what Chapter Five introduces is not statistical concepts, but - as its title suggests - the use of the microcomputer for statistical analysis. (If a Commodore 64 computer is available, readers may prefer to obtain the IFASTATS programs and go through Chapter Five at the keyboard.)

Chapters Six through Thirteen discuss specific studies or ongoing projects in which statistical analyses have been applied. Jankowski's (UAM) *A software tool for the statistical analysis of language tests* (Chapter Six) describes the use of a microcomputer for the analysis of English language tests at the Institute of English Studies in Poznań. Jones's Chapter Seven presents a compu-

ter based test item bank using a Rasch model. Nizegorodcew and Krzanowska's (UJ) Chapter Eight describes a study on the reception of British and American English by Polish university students. Turewicz's (UŁ) *The sign-test for significance used in support of new concepts of meaning* (Chapter Nine) attempts to relate some empirical findings to the theoretical realm, i.e. to claims made by Langacker's Cognitive Linguistics. In Chapter Ten, Waniek-Klimczak (UŁ) tackles a sociolinguistic problem in her analysis of vowel length and word final consonant devoicing in the English utterances of native speakers of Polish. Lewandowska-Tomaszczyk's (UŁ) *A statistical analysis of lexis in conversational English* (Chapter Eleven) extends the analysis of data which she examined in her habilitacja work, *Conceptual Analysis, Linguistic Meaning, and Verbal Interaction*. Johnson's (UAM) Chapter Twelve, *The effects of cohesion and coherence in written discourse on comprehension*, presents an example of classroom-oriented research for applied linguistics. *The structure of FL competence* is discussed in Chapter Thirteen by Tomaszczyk (UŁ). Finally, in Chapter Fourteen, Kwiatkowska (UŁ) describes an experimental psycholinguistic study on discourse *(The definitiveness hierarchy and strength of anaphoric link in Polish)*.

For the reader's convenience, all references have been collected together at the back of the book, and an index has been provided to facilitate the use of the text as a reference source.

Douglas W. Coleman
Barbara Lewandowska-Tomaszczyk
Institute of English Studies
University of Łódź