

INTRODUCTION

Even though linguistics departments are, by tradition, usually located in colleges of humanities, linguistics itself aspires to be—and at its best, manages to be—a science. That is, linguists aim to do much the same thing that scientists in general (such as physicists, geologists, biologists, and chemists) do: to make observations of certain kinds of natural phenomena, and then state **empirical hypotheses** about them. The only difference is that the phenomena linguists study have to do not with swinging pendulums, tectonic plates, zebra mussels, or hydrocarbons, but with human language: how it sounds, what it means, how it varies across space and time, how it is learned, used, and understood.

Roughly speaking, an empirical hypothesis is just a well-informed and careful guess about what certain kinds of events will be like, based on past observations of events of that kind. To put it a bit more precisely, an empirical hypothesis is a general statement about a class of phenomena that has the following properties: (1) It is *clear and unambiguous*, that is, there is no question what it asserts (how things would have to be in order for it to be true). (2) It is *general*, in the sense that even though it is based only on a finite number of observations, it makes predictions about how other phenomena of the same kind will unfold. (3) There is a way to tell whether or not a given observation of the kind of phenomenon in question is consistent with it, so that if the hypothesis is wrong, there is some hope of finding out that it is wrong. This third property of empirical hypotheses is called **falsifiability**. Especially valued are empirical hypotheses with the additional property of being *illuminating*, in the sense of being sufficiently simple and comprehensible to help us grasp some of the hidden orderliness or systematicity in seemingly random or chaotic phenomena.

A **linguistic theory** is just a set of empirical hypotheses about a class of natural-language phenomena. Linguists often refer to the process of formulating empirical hypotheses about human languages as ‘capturing linguistic generalizations’. This is just a fancy name for linguistic theorizing. The purpose of this book is to introduce some techniques for doing just that.

The techniques introduced in this book are drawn from areas of mathematics (such as set theory, logic, algebra, and formal language theory) that are usually described as **discrete**, as opposed to **continuous** (such as calculus, differential equations, Fourier analysis, or probability). The natural numbers are discrete; the real numbers are continuous. The subdisciplines of linguistics that most readily lend themselves to analysis by discrete methods include (but are not limited to) the following: (1) **morphology** (how

words are built up from their meaningful subparts); (2) **syntax** (how words combine into successively larger phrases, including sentences); and (3) **semantics** (how linguistic expressions manage to refer to things in the world and express propositions about them, and how it is that some propositions follow from (or are **entailed** by) other propositions). There are also parts of **phonology** (how human languages structure spoken sounds) and **computational linguistics** (the analysis and manipulation of human language using computational concepts or computer programs) that yield to such methods. But other linguistic disciplines, such as phonetics, psycholinguistics, sociolinguistics, and historical linguistics in general call for continuous methods. Interestingly, many of the discrete mathematical techniques that come into play in the analysis of human language are the same ones used in analyzing the artificial languages employed in logic and computer science.

This is an applied mathematics book, not a linguistics book, and so the emphasis is primarily on the mathematical concepts and techniques themselves, not on the phenomena to which they are applied. In fact, most of these are of inherent interest independent of the linguistic applications, and it is entirely possible to master them without knowing or caring about linguistics at all! But the book is written primarily with the needs of linguistics students in mind.