Preface

Mathematical linguistics is rooted both in Euclid's (circa 325–265 BCE) axiomatic method and in Pāṇini's (circa 520–460 BCE) method of grammatical description. To be sure, both Euclid and Pāṇini built upon a considerable body of knowledge amassed by their precursors, but the systematicity, thoroughness, and sheer scope of the *Elements* and the *Aṣhṭādhyāyī* would place them among the greatest landmarks of all intellectual history even if we disregarded the key methodological advance they made.

As we shall see, the two methods are fundamentally very similar: the axiomatic method starts with a set of statements assumed to be true and transfers truth from the axioms to other statements by means of a fixed set of logical rules, while the method of grammar is to start with a set of expressions assumed to be grammatical both in form and meaning and to transfer grammaticality to other expressions by means of a fixed set of grammatical rules.

Perhaps because our subject matter has attracted the efforts of some of the most powerful minds (of whom we single out A. A. Markov here) from antiquity to the present day, there is no single easily accessible introductory text in mathematical linguistics. Indeed, to the mathematician the whole field of linguistics may appear to be hopelessly mired in controversy, and neither the formidable body of empirical knowledge about languages nor the standards of linguistic argumentation offer an easy entry point.

Those with a more postmodern bent may even go as far as to doubt the existence of a solid core of mathematical knowledge, often pointing at the false theorems and incomplete or downright wrong proofs that slip through the peer review process at a perhaps alarming rate. Rather than attempting to drown such doubts in rivers of philosophical ink, the present volume will simply proceed *more geometrico* in exhibiting this solid core of knowledge. In Chapters 3–6, a mathematical overview of the traditional main branches of linguistics, phonology, morphology, syntax, and semantics, is presented.

Who should read this book?

The book is accessible to anyone with sufficient general mathematical maturity (graduate or advanced undergraduate). No prior knowledge of linguistics or languages is assumed on the part of the reader. The book offers a single entry point to the central methods and concepts of linguistics that are made largely inaccessible to the mathematician, computer scientist, or engineer by the surprisingly adversarial style of argumentation (see Section 1.2), the apparent lack of adequate definitions (see Section 1.3), and the proliferation of unmotivated notation and formalism (see Section 1.4) all too often encountered in research papers and monographs in the humanities. Those interested in linguistics can learn a great deal more about the subject here than what is covered in introductory courses just from reading through the book and consulting the references cited. Those who plan to approach linguistics through this book should be warned in advance that many branches of linguistics, in particular psycholinguistics, child language acquisition, and the study of language pathology, are largely ignored here – not because they are viewed as inferior to other branches but simply because they do not offer enough grist for the mathematician's mill. Much of what the linguistically naive reader may find interesting about language turns out to be more pertinent to cognitive science, the philosophy of language, and sociolinguistics, than to linguistics proper, and the Introduction gives these issues the shortest possible shrift, discussing them only to the extent necessary for disentangling mathematical linguistics from other concerns.

Conversely, issues that linguists sometimes view as peripheral to their enterprise will get more discussion here simply because they offer such a rich variety of mathematical techniques and problems that no book on mathematical linguistics that ignored them could be considered complete. After a brief review of information theory in Chapter 7, we will devote Chapters 8 and 9 to phonetics, speech recognition, the recognition of handwriting and machine print, and in general to issues of linguistic signal processing and pattern matching, including information extraction, information retrieval, and statistical natural language processing. Our treatment assumes a bit more mathematical maturity than the excellent textbooks by Jelinek (1997) and Manning and Schütze (1999) and intends to complement them. Kracht (2003) conveniently summarizes and extends much of the discrete (algebraic and combinatorial) work on mathematical linguistics. It is only because of the timely appearance of this excellent reference work that the first six chapters could be kept to a manageable size and we could devote more space to the continuous (analytic and probabilistic) aspects of the subject. In particular, expository simplicity would often dictate that we keep the underlying parameter space discrete, but in the later chapters we will be concentrating more on the case of continuous parameters, and discuss the issue of quantization losses explicitly.

In the early days of computers, there was a great deal of overlap between the concerns of mathematical linguistics and computer science, and a surprising amount of work that began in one field ended up in the other, sometimes explicitly as part of computational linguistics, but often as general theory with its roots in linguistics largely forgotten. In particular, the basic techniques of syntactic analysis are now

firmly embedded in the computer science curriculum, and the student can already choose from a large variety of textbooks that cover parsing, automata, and formal language theory. Here we single out the classic monograph by Salomaa (1973), which shows the connection to formal syntax in a way readily accessible to the mathematically minded reader. We will selectively cover only those aspects of this field that address specifically linguistic concerns, and again our guiding principle will be mathematical content, as opposed to algorithmic detail. Readers interested in the algorithms should consult the many excellent natural language processing textbooks now available, of which we single out Jurafsky and Martin (2000, with a new edition planned in 2007).

How is the book organized?

To the extent feasible we follow the structure of the standard introductory courses to linguistics, but the emphasis will often be on points only covered in more advanced courses. The book contains many exercises. These are, for the most part, rather hard (over level 30 in the system of Knuth 1971) but extremely rewarding. Especially in the later chapters, the exercises are often based on classical and still widely cited theorems, so the solutions can usually be found on the web quite easily simply by consulting the references cited in the text. However, readers are strongly advised not to follow this route before spending at least a few days attacking the problem. Unsolved problems presented as exercises are marked by an asterisk, a symbol that we also use when presenting examples and counterexamples that native speakers would generally consider wrong (ungrammatical): *Scorsese is a great director* is a positive (grammatical) example while **Scorsese a great director is* is a negative (ungrammatical) example. Some exercises, marked by a dagger †, require the ability to manipulate sizeable data sets, but no in-depth knowledge of programming, data structures, or algorithms is presumed. Readers who write code effortlessly will find these exercises easy, as they rarely require more than a few simple scripts. Those who find such exercises problematic can omit them entirely. They may fail to gain direct appreciation of some empirical properties of language that drive much of the research in mathematical linguistics, but the research itself remains perfectly understandable even if the motivation is taken on faith. A few exercises are marked by a raised M – these are major research projects the reader is not expected to see to completion, but spending a few days on them is still valuable.

Because from time to time it will be necessary to give examples from languages that are unlikely to be familiar to the average undergraduate or graduate student of mathematics, we decided, somewhat arbitrarily, to split languages into two groups. *Major* languages are those that have a chapter in Comrie's (1990) *The World's Major Languages* – these will be familiar to most people and are left unspecified in the text. *Minor* languages usually require some documentation, both because language names are subject to a great deal of spelling variation and because different groups of people may use very different names for one and the same language. Minor languages are therefore identified here by their three-letter Ethnologue code (15th edition, 2005) given in square brackets [].

Each chapter ends with a section on further reading. We have endeavored to make the central ideas of linguistics accessible to those new to the field, but the discussion offered in the book is often skeletal, and readers are urged to probe further. Generally, we recommend those papers and books that presented the idea for the first time, not just to give proper credit but also because these often provide perspective and insight that later discussions take for granted. Readers who industriously follow the recommendations made here should do so for the benefit of learning the basic vocabulary of the field rather than in the belief that such reading will immediately place them at the forefront of research.

The best way to read this book is to start at the beginning and to progress linearly to the end, but the reader who is interested only in a particular area should not find it too hard to jump in at the start of any chapter. To facilitate skimming and alternative reading plans, a generous amount of forward and backward pointers are provided – in a hypertext edition these would be replaced by clickable links. The material is suitable for an aggressively paced one-semester course or a more leisurely paced two-semester course.

Acknowledgments

Many typos and stylistic infelicities were caught, and excellent references were suggested, by Márton Makrai (Budapest Institute of Technology), Dániel Margócsy (Harvard), Doug Merritt (San Jose), Reinhard Muskens (Tilburg), Almerindo Ojeda (University of California, Davis), Bálint Sass (Budapest), Madeleine Thompson (University of Toronto), and Gabriel Wyler (San Jose). The painstaking work of the Springer editors and proofreaders, Catherine Brett, Frank Ganz, Hal Henglein, and Jeffrey Taub, is gratefully acknowledged.

The comments of Tibor Beke (University of Massachusetts, Lowell), Michael Bukatin (MetaCarta), Anssi Yli-Jyrä (University of Helsinki), Péter Gács (Boston University), Marcus Kracht (UCLA), András Serény (CEU), Péter Siptár (Hungarian Academy of Sciences), Anna Szabolcsi (NYU), Péter Vámos (Budapest Institute of Technology), Károly Varasdi (Hungarian Academy of Sciences), and Dániel Varga (Budapest Institute of Technology) resulted in substantive improvements.

Writing this book would not have been possible without the generous support of MetaCarta Inc. (Cambridge, MA), the MOKK Media Research center at the Budapest Institute of Technology Department of Sociology, the Farkas Heller Foundation, and the Hungarian Telekom Foundation for Higher Education – their help and care is gratefully acknowledged.

Contents

1

Introduction

1.1 The subject matter

What is *mathematical linguistics*? A classic book on the subject, (Jakobson 1961), contains papers on a variety of subjects, including a categorial grammar (Lambek 1961), formal syntax (Chomsky 1961, Hiż 1961), logical semantics (Quine 1961, Curry 1961), phonetics and phonology (Peterson and Harary 1961, Halle 1961), Markov models (Mandelbrot 1961b), handwriting (Chao 1961, Eden 1961), parsing (Oettinger 1961, Yngve 1961), glottochronology (Gleason 1961), and the philosophy of language (Putnam 1961), as well as a number of papers that are harder to fit into our current system of scientific subfields, perhaps because there is a void now where once there was cybernetics and systems theory (see Heims 1991).

A good way to understand how these seemingly so disparate fields cohere is to proceed by analogy to mathematical physics. Hamiltonians receive a great deal more mathematical attention than, say, the study of generalized incomplete Gamma functions, because of their relevance to mechanics, not because the subject is, from a purely mathematical perspective, necessarily more interesting. Many parts of mathematical physics find a natural home in the study of differential equations, but other parts fit much better in algebra, statistics, and elsewhere. As we shall see, the situation in mathematical linguistics is quite similar: many parts of the subject would fit nicely in algebra and logic, but there are many others for which methods belonging to other fields of mathematics are more appropriate. Ultimately the coherence of the field, such as it is, depends on the coherence of linguistics.

Because of the enormous impact that the works of Noam Chomsky and Richard Montague had on the postwar development of the discipline, there is a strong tendency, observable both in introductory texts such as Partee et al. (1990) and in research monographs such as Kracht (2003), to simply equate mathematical linguistics with formal syntax and semantics. Here we take a broader view, assigning syntax (Chapter 5) and semantics (Chapter 6) no greater scope than they would receive in any book that covers linguistics as a whole, and devoting a considerable amount of space to phonology (Chapter 2), morphology (Chapter 3), phonetics (Chapters 8 and 9), and other areas of traditional linguistics. In particular, we make sure that the

reader will learn (in Chapter 7) the central mathematical ideas of information theory and algorithmic complexity that provide the foundations of much of the contemporary work in mathematical linguistics.

This does not mean, of course, that mathematical linguistics is a discipline entirely without boundaries. Since almost all social activity ultimately rests on linguistic communication, there is a great deal of temptation to reduce problems from other fields of inquiry to purely linguistic problems. Instead of understanding schizoid behavior, perhaps we should first ponder what the phrase *multiple personality* means. Mathematics already provides a reasonable notion of 'multiple', but what is 'personality', and how can there be more than one per person? Can a proper understanding of the suffixes *-al* and *-ity* be the key? This line of inquiry, predating the Schoolmen and going back at least to the *cheng ming* (rectification of names) doctrine of Confucius, has a clear and convincing rationale (*The Analects* 13.3, D.C. Lau transl.):

> When names are not correct, what is said will not sound reasonable; when what is said does not sound reasonable, affairs will not culminate in success; when affairs do not culminate in success, rites and music will not flourish; when rites and music do not flourish, punishments will not fit the crimes; when punishments do not fit the crimes, the common people will not know where to put hand and foot. Thus when the gentleman names something, the name is sure to be usable in speech, and when he says something this is sure to be practicable. The thing about the gentleman is that he is anything but casual where speech is concerned.

In reality, linguistics lacks the resolving power to serve as the ultimate arbiter of truth in the social sciences, just as physics lacks the resolving power to explain the accidents of biological evolution that made us human. By applying mathematical techniques we can at least gain some understanding of the limitations of the enterprise, and this is what this book sets out to do.

1.2 Cumulative knowledge

It is hard to find any aspect of linguistics that is entirely uncontroversial, and to the mathematician less steeped in the broad tradition of the humanities it may appear that linguistic controversies are often settled on purely rhetorical grounds. Thus it may seem advisable, and only fair, to give both sides the full opportunity to express their views and let the reader be the judge. But such a book would run to thousands of pages and would be of far more interest to historians of science than to those actually intending to learn mathematical linguistics. Therefore we will not necessarily accord equal space to both sides of such controversies; indeed often we will present a single view and will proceed without even attempting to discuss alternative ways of looking at the matter.

Since part of our goal is to orient the reader not familiar with linguistics, typically we will present the majority view in detail and describe the minority view only

tersely. For example, Chapter 4 introduces the reader to morphology and will rely heavily on the notion of the morpheme – the excellent book by Anderson (1992) denying the utility, if not the very existence, of morphemes, will be relegated to footnotes. In some cases, when we feel that the minority view is the correct one, the emphasis will be inverted: for example, Chapter 6, dealing with semantics, is more informed by the 'surface compositional' than the 'logical form' view. In other cases, particularly in Chapter 5, dealing with syntax, we felt that such a bewildering variety of frameworks is available that the reader is better served by an impartial analysis that tries to bring out the common core than by in-depth formalization of any particular strand of research.

In general, our goal is to present linguistics as a cumulative body of knowledge. In order to find a consistent set of definitions that offer a rational reconstruction of the main ideas and techniques developed over the course of millennia, it will often be necessary to take sides in various controversies. There is no pretense here that mathematical formulation will necessarily endow a particular set of ideas with greater verity, and often the opposing view could be formalized just as well. This is particularly evident in those cases where theories diametrically opposed in their means actually share a common goal such as describing all and only the well-formed structures (e.g. syllables, words, or sentences) of languages. As a result, we will see discussions of many 'minority' theories, such as case grammar or generative semantics, which are generally believed to have less formal content than their 'majority' counterparts.

1.3 Definitions

For the mathematician, definitions are nearly synonymous with abbreviations: we say 'triangle' instead of describing the peculiar arrangement of points and lines that define it, 'polynomial' instead of going into a long discussion about terms, addition, monomials, multiplication, or the underlying ring of coefficients, and so forth. The only sanity check required is to exhibit an instance, typically an explicit set-theoretic construction, to demonstrate that the defined object indeed exists. Quite often, counterfactual objects such as the smallest group K not meeting some description, or objects whose existence is not known, such as the smallest nontrivial root of ζ not on the critical line, will play an important role in (indirect) proofs, and occasionally we find cases, such as *motivic cohomology*, where the whole conceptual apparatus is in doubt. In linguistics, there is rarely any serious doubt about the existence of the objects of inquiry. When we strive to define 'word', we give a mathematical formulation not so much to demonstrate that words exist, for we know perfectly well that we use words both in spoken and written language, but rather to handle the odd and unexpected cases. The reader is invited to construct a definition now and to write it down for comparison with the eventual definition that will emerge only after a rather complex discussion in Chapter 4.

In this respect, mathematical linguistics is very much like the empirical sciences, where formulating a definition involves at least three distinct steps: an *ostensive* def-

inition based on positive and sometimes negative examples (vitriol is an acid, lye is not), followed by an *extensive* definition delineating the intended scope of the notion (every chemical that forms a salt with a base is an acid), and the *intensive* definition that exposes the underlying mechanism (in this case, covalent bonds) emerging rather late as a result of a long process of abstraction and analysis.

Throughout the book, the first significant instance of key notions will appear in *italics*, usually followed by ostensive examples and counterexamples in the next few paragraphs. (Italics will also be used for emphasis and for typesetting linguistic examples.) The empirical observables associated with these notions are always discussed, but textbook definitions of an extensive sort are rarely given. Rather, a mathematical notion that serves as a stand-in will be defined in a rigorous fashion: in the defining phrase, the same notion is given in **boldface**. Where an adequate mathematical formulation is lacking and we proceed by sheer analogy, the key terms will be *slanted* – such cases are best thought of as open problems in mathematical linguistics.

1.4 Formalization

In mathematical linguistics, as in any branch of applied mathematics, the issue of formalizing semiformally or informally stated theories comes up quite often. A prime example is the study of phrase structure, where Chomsky (1956) took the critical step of replacing the informally developed system of immediate constituent analysis (ICA, see Section 5.1) by the rigorously defined context-free grammar (CFG, see Section 2.3) formalism. Besides improving our understanding of natural language, a worthy goal in itself, the formalization opened the door to the modern theory of computer languages and their compilers. This is not to say that every advance in formalizing linguistic theory is likely to have a similarly spectacular payoff, but clearly the informal theory remains a treasure-house inasmuch as it captures important insights about natural language. While not entirely comparable to biological systems in age and depth, natural language embodies a significant amount of evolutionary optimization, and artificial communication systems can benefit from these developments only to the extent that the informal insights are captured by formal methods.

The quality of formalization depends both on the degree of faithfulness to the original ideas and on the mathematical elegance of the resulting system. Because the proper choice of formal apparatus is often a complex matter, linguists, even those as evidently mathematical-minded as Chomsky, rarely describe their models with full formal rigor, preferring to leave the job to the mathematicians, computer scientists, and engineers who wish to work with their theories. Choosing the right formalism for linguistic rules is often very hard. There is hardly any doubt that linguistic behavior is governed by rather abstract rules or constraints that go well beyond what systems limited to memorizing previously encountered examples could explain. Whether these rules have a stochastic aspect is far from settled: engineering applications are dominated by models that crucially rely on probabilities, while theoretical models, with the notable exception of the *variable rules* used in sociolin-

guistics (see Section 5.4.3), rarely include considerations relating to the frequency of various phenomena. The only way to shed light on such issues is to develop alternative formalizations and compare their mathematical properties.

The tension between faithfulness to the empirical details and the elegance of the formal system has long been familiar to linguists: Sapir (1921) already noted that "all grammars leak". One significant advantage that probabilistic methods have over purely symbolic techniques is that they come with their own built-in measure of leakiness (see Section 5.4). It is never a trivial matter to find the appropriate degree of idealization in pursuit of theoretical elegance, and all we can do here is to offer a couple of convenient stand-ins for the very real but still somewhat elusive notion of elegance.

The first stand-in, held in particularly high regard in linguistics, is *brevity*. The contemporary slogan of algorithmic complexity (see Section 7.2), that the best theory is the shortest theory, could have been invented by Pāṇini. The only concession most linguists are willing to make is that some of the complexity should be ascribed to principles of *universal grammar* (UG) rather than to the *parochial* rules specific to a given language, and since the universal component can be amortized over many languages, we should maximize its explanatory burden at the expense of the parochial component.

The second stand-in is *stability* in the sense that minor perturbations of the definition lead to essentially the same system. Stability has always been highly regarded in mathematics: for example, Birkhoff (1940) spent significant effort on establishing the value of lattices as legitimate objects of algebraic inquiry by investigating alternative definitions that ultimately lead to the same class of structures. There are many ways to formalize an idea, and when small changes in emphasis have a very significant impact on the formal properties of the resulting system, its mathematical value is in doubt. Conversely, when variants of formalisms as different as indexed grammars (Aho 1968), combinatory categorial grammar (Steedman 2001), head grammar (Pollard 1984), and tree adjoining grammar (Joshi 2003) define the same class of languages, the value of each is significantly enhanced.

One word of caution is in order: the fact that some idea is hard to formalize, or even seems so contradictory that a coherent mathematical formulation appears impossible, can be a reflection on the state of the art just as well as on the idea itself. Starting with Berkeley (1734), the intuitive notion of infinitesimals was subjected to all kinds of criticism, and it took over two centuries for mathematics to catch up and provide an adequate foundation in Robinson (1966). It is quite conceivable that equally intuitive notions, such as a *semantic theory of information*, which currently elude our mathematical grasp, will be put on firm foundations by later generations. In such cases, we content ourselves with explaining the idea informally, describing the main intuitions and pointing at possible avenues of formalization only programmatically.

1.5 Foundations

For the purposes of mathematical linguistics, the classical foundations of mathematics are quite satisfactory: all objects of interest are sets, typically finite or, rarely, denumerably infinite. This is not to say that nonclassical metamathematical tools such as Heyting algebras find no use in mathematical linguistics but simply to assert that the fundamental issues of this field are not foundational but definitional.

Given the finitistic nature of the subject matter, we will in general use the terms set, class, and collection interchangeably, drawing explicit cardinality distinctions only in the rare cases where we step out of the finite domain. Much of the classical linguistic literature of course predates Cantor, and even the modern literature typically conceives of infinity in the Gaussian manner of a potential, as opposed to actual, Cantorian infinity. Because of immediate empirical concerns, denumerable generalizations of finite objects such as ω-words and Büchi automata are rarely used,[1] and in fact even the trivial step of generalizing from a fixed constant to arbitrary n is often viewed with great suspicion.

Aside from the tradition of Indian logic, the study of languages had very little impact on the foundations of mathematics. Rather, mathematicians realized early on that natural language is a complex and in many ways unreliable construct and created their own simplified language of formulas and the mathematical techniques to investigate it. As we shall see, some of these techniques are general enough to cover essential facets of natural languages, while others scale much more poorly.

There is an interesting residue of foundational work in the Berry, Richard, Liar, and other paradoxes, which are often viewed as diagnostic of the vagueness, ambiguity, or even 'paradoxical nature' of natural language. Since the goal is to develop a mathematical theory of language, sooner or later we must define English in a formal system. Once this is done, the buck stops there, and questions like "what is the smallest integer not nameable in ten words?" need to be addressed anew.

We shall begin with the seemingly simpler issue of the first number not nameable in *one* word. Since it appears to be one hundred and one, a number already requiring *four* words to name, we should systematically investigate the number of words in number names. There are two main issues to consider: what is a word? (see Chapter 4); and what is a name? (see Chapter 6). Another formulation of the Berry paradox invokes the notion of syllables; these are also discussed in Chapter 4. Eventually we will deal with the paradoxes in Chapter 6, but our treatment concentrates on the linguistic, rather than the foundational, issues.

1.6 Mesoscopy

Physicists speak of mesoscopic systems when these contain, say, fifty atoms, too large to be given a microscopic quantum-mechanical description but too small for the classical macroscopic properties to dominate the behavior of the system. Linguistic

[1] For a contrary view, see Langendoen and Postal (1984).

systems are mesoscopic in the same broad sense: they have thousands of rules and axioms compared with the handful of axioms used in most branches of mathematics. Group theory explores the implications of five axioms, arithmetic and set theory get along with five and twelve axioms respectively (not counting members of axiom schemes separately), and the most complex axiom system in common use, that of geometry, has less than thirty axioms.

It comes as no surprise that with such a large number of axioms, linguistic systems are never pursued microscopically to yield implications in the same depth as group theory or even less well-developed branches of mathematics. What is perhaps more surprising is that we can get reasonable approximations of the behavior at the macroscopic level using the statistical techniques pioneered by A. A. Markov (see Chapters 7 and 8).

Statistical mechanics owes its success largely to the fact that in thermodynamics only a handful of phenomenological parameters are of interest, and these are relatively easy to link to averages of mechanical quantities. In mathematical linguistics the averages that matter (e.g. the percentage of words correctly recognized or correctly translated) are linked only very indirectly to the measurable parameters, of which there is such a bewildering variety that it requires special techniques to decide which ones to employ and which ones to leave unmodeled.

Macroscopic techniques, by their very nature, can yield only approximations for mesoscopic systems. Microscopic techniques, though in principle easy to extend to the mesoscopic domain, are in practice also prone to all kinds of bugs, ranging from plain errors of fact (which are hard to avoid once we deal with thousands of axioms) to more subtle, and often systematic, errors and omissions. Readers may at this point feel very uncomfortable with the idea that a given system is only 70%, 95%, or even 99.99% correct. After all, isn't a single contradiction or empirically false prediction enough to render a theory invalid? Since we need a whole book to develop the tools needed to address this question, the full answer will have to wait until Chapter 10.

What is clear from the outset is that natural languages offer an unparalleled variety of complex algebraic structures. The closest examples we can think of are in crystallographic topology, but the internal complexity of the groups studied there is a product of pure mathematics, while the internal complexity of the syntactic semigroups associated to natural languages is more attractive to the applied mathematician, as it is something found in vivo. Perhaps the most captivating aspect of mathematical linguistics is not just the existence of discrete mesoscopic structures but the fact that these come embedded, in ways we do not fully understand, in continuous signals (see Chapter 9).

1.7 Further reading

The first works that can, from a modern standpoint, be called mathematical linguistics are Markov's (1912) extension of the weak law of large numbers (see Theorem 8.2.2) and Thue's (1914) introduction of string manipulation (see Chapter 2), but pride of place must go to Pāṇini, whose inventions include not just grammatical rules

but also a formal metalanguage to describe the rules and a set of principles governing their interaction. For modern accounts of various aspects of the system see Staal (1962, 1967) Cardona (1965, 1969, 1970, 1976, 1988), and Kiparsky (1979, 1982b, 2002). Needless to say, Pāṇini did not work in isolation. Much like Euclid, he built on the inventions of his predecessors, but his work was so comprehensive that it effectively drove the earlier material out of circulation. While much of linguistics has aspired to formal rigor throughout the ages (for the Masoretic tradition, see Aronoff 1985, for medieval syntax see Covington 1984), the continuous line of development that culminates in contemporary formal grammar begins with Bloomfield's (1926) Postulates (see Section 3.1), with the most important milestones being Harris (1951) and Chomsky (1956, 1959).

Another important line of research, only briefly alluded to above, could be called mathematical antilinguistics, its goal being the elimination, rather than the explanation, of the peculiarities of natural language from the system. The early history of the subject is discussed in depth in Eco (1995); the modern mathematical developments begin with Frege's (1879) system of *Concept Writing* (Begriffsschrift), generally considered the founding paper of mathematical logic. There is no doubt that many great mathematicians from Leibniz to Russell were extremely critical of natural language, using it more for counterexamples and cautionary tales than as a part of objective reality worthy of formal study, but this critical attitude has all but disappeared with the work of Montague (1970a, 1970b, 1973). Contemporary developments in model-theoretic semantics or 'Montague grammar' are discussed in Chapter 6.

Major summaries of the state of the art in mathematical linguistics include Jakobson (1961), Levelt (1974), Manaster-Ramer (1987), and the subsequent Mathematics of Language (MOL) conference volumes. We will have many occasions to cite Kracht's (2003) indispensable monograph *The Mathematics of Language*.

The volumes above are generally more suitable for the researcher or advanced graduate student than for those approaching the subject as undergraduates. To some extent, the mathematical prerequisites can be learned from the ground up from classic introductory textbooks such as Gross (1972) or Salomaa (1973). Gruska (1997) offers a more modern and, from the theoretical computer science perspective, far more comprehensive introduction. The best elementary introduction to the logical prerequisites is Gamut (1991). The discrete side of the standard "mathematics for linguists" curriculum is conveniently summarized by Partee et al. (1990), and the statistical approach is clearly introduced by Manning and Schütze (1999). The standard introduction to pattern recognition is Duda et al. (2000). Variable rules were introduced in Cedergren and Sankoff (1974) and soon became the standard modeling method in sociolinguistics – we shall discuss them in Chapter 5.

2

The elements

A primary concern of mathematical linguistics is to effectively enumerate those sets of words, sentences, etc., that play some important linguistic role. Typically, this is done by means of *generating* the set in question, a definitional method that we introduce in Section 2.1 by means of examples and counterexamples that show the similarities and the differences between the standard mathematical use of the term 'generate' and the way it is employed in linguistics.

Because the techniques used in defining sets, functions, relations, etc., are not always directly useful for evaluating them at a given point, an equally important concern is to solve the membership problem for the sets, functions, relations, and other structures of interest. In Section 2.2 we therefore introduce a variety of *grammars* that can be used to, among other things, create *certificates* that a particular element is indeed a member of the set, gets mapped to a particular value, stands in a prescribed relation to other elements and so on, and compare generative systems to logical calculi.

Since *generative grammar* is most familiar to mathematicians and computer scientists as a set of rather loosely collected string-rewriting techniques, in Section 2.3 we give a brief overview of this domain. We put the emphasis on context-sensitive grammars both because they play an important role in phonology (see Chapter 3) and morphology (see Chapter 4) and because they provide an essential line of defense against undecidability in syntax (see Chapter 5).

2.1 Generation

To define a collection of objects, it is often expedient to begin with a fixed set of primitive elements E and a fixed collection of *rules* (we use this term in a broad sense that does not imply strict procedurality) R that describe permissible arrangements of the primitive elements as well as of more complex objects. If x, y, z are objects *satisfying* a (binary) rule $z = r(x, y)$, we say that z **directly generates** x and y (in this order) and use the notation $z \rightarrow_r xy$. The smallest collection of objects closed