# Mathematical Linguistics - A. Kornai

Alok Debnath

September 2019

**Abstract**

This is a summary of the textbook "Mathematical Linguistics" by A. Kornai. The book may be found on `http://www.helsinki.fi/esslli/courses/readers/K54.pdf`. Some parts of the book require some prerequisites which are non-crucial. I have attempted to capture those in the footnotes. For those which are crucial a relevant section has been added, which can be used. This is only a summary of the book, and I have skimmed over a lot of the parts, some of which may be of interest to the reader. If this is so, I recommend reading the textbook.

# Chapter 1

# Introduction

The aim of this text is to analyze the notions which are of linguistic interest, such as phonology, morphology, syntax and semantics by using "mathematical techniques". The author compares this to a study of mathematical physics, where some branches or subfields are studied far more than others, not because they are not interesting, more so beacuse there is no direct utility in studying them over mathematical explanations of observable phenomenon.

**Definitions**: Definitions in mathematical linguistics follow three basic steps, an *ostensive* definition based on examples, an *extensive* definition which dealiantes the intended scope of the notion and an *intensive* definition, which exposes the underlying mechanism.

**Formalization**: Much like other branches of applied mathematics, the problem of formalization of semi-formally or informally stated theories. However, the choice of a formal definitions and structures is often hard. The stochastic nature of language and linguistic rules is still a debate.

**Foundations**: Mathemaical linguistics has sets as the object of interest, usually finite but sometimes denumberably infinite (see: set of words). Due to emperical restrictions, however, denumerable generailzations of finite objects such as $\omega$-words are rarely used. [1] For the purpose of this series, such constructions are not taken into account. The study of language and mathematics have evolved rather independently, the primary exception being Indian traditions of logic.

**Mesoscopy**: A mesoscopic system is one which is neither too small to be appropriately studied by microscopic methods and tools, and yet is not large enough for statistical approximations of macroscopic systems are necessarily accurate. Natural language may be viewed as a mesoscopic systems, which is supposedly governed by a thousands of rules, wheras the statistical methods of macroscopic levels (Markov assumption, for example) is a reasonable approximation of the behavior of Natural Language. Macroscopic techniques yeild only generalizations of mesoscopic systems. Further, statistical quantities of interest

[1] $\omega$- words are words that can be written on an $\omega$-automaton, a finite automaton that runs on infinite strings as inputs. Buchi automata are examples of the same.

are linked "only very indirectly" to the object of study, which requires very specific techniques in order to decide which parameters are of interest and which are to be left unmodeled.