

How to read this book

Since you're reading this, I assume that your mind already has some method of transferring its written symbols into meaningful sentences. But sentence-level comprehension is not quite enough. This book doesn't follow a familiar linear structure, it is non-linear by design. Before venturing forth, you should get yourself acquainted with the overall structure.

1 Parts, chapters, units

The book is split into multiple **parts**: *Main units*, *Background*, and *Additional Practice Exercises*. I will have more to say on each part below.

Each part consists of multiple **chapters**. Chapters correspond to broad topics such as n-gram models or finite-state automata. Each chapter is broken down into **units**, with each unit illustrating a core aspect of the chapter topic or one of its linguistic applications. The main body of each unit is interspersed with offset blocks for examples, exercises, or advice. The end of a unit usually contains a brief recap with bullet points.

EXAMPLE 1.

This, for example, is an example block giving you an example of what an example block looks like.

EXERCISE 1.

Look at the table of contents for this book and identify parts, chapters, and units.

ADVICE

Take a brief glimpse at some of the units in this book to get a feel for their length and flow.

I tried my best to keep each unit short enough that it could make for a comfy bedtime read. While I didn't always succeed, I think I did so most of the time.

2 Prerequisites

The real meat of this book is in the *Main units* part, which will introduce you to how mathematical concepts can be applied to language and what we can learn from doing so. But this part does not actually teach you said math. Instead, all information on mathematical topics is kept separate, for easy reference, in the *Background* part.

EXAMPLE 2.

The chapter on n-gram models in *Main units* uses sets and functions, but those concepts aren't fully explained there. Instead you have to read specific units in the *Background* part.

You are not supposed to read the *Background* part in a linear fashion. Instead, the units in *Main units* will tell you when it is time to consult a unit in *Background*. If there is a specific mathematical concept that you need to be familiar with in order to fully appreciate a *Main* unit, that unit will start with a list of prerequisites, e.g.:

PREREQUISITES

- sets (operations, set-relations)

This means that you have to know about set operations like union and intersection, and the typical relations that can hold between sets (e.g. subset and superset). If you don't have the required background, you should go to the *Background* chapter about sets and read the units on operations and set-relations. These units may themselves have prerequisites. But the book is structured in such a manner that as long you follow the main units from start to finish without skipping anything, you will also move through the background units in a linear fashion so that you should never have to read a background unit for which you don't meet the prerequisites yet.

EXAMPLE 3.

Georg Ferdinand Ludwig Philipp Cantor wants to read up on n-grams models. He picks the first unit and is happy to see that it has no prerequisites. But as the second unit is much more formal, it requires some knowledge of set theory and functions:

- set theory (operations, set-relations)
- functions (notation)

While Cantor got the first one down, he's a little rusty on his function notation. So he goes to *Background* and checks out the unit on function notation. The unit also requires some set theory, but since Cantor is on top of that he can read the unit right away. Once he's done, he continues with the second unit on n-gram models.

3 Optional chapters

Some chapters in *Main units* are marked as optional. This means that they can be skipped without missing any prerequisites of later chapters.

EXAMPLE 4.

The chapter on n-grams in language technology introduces the concept of multisets. But since no other *Main* chapters use multisets, you can skip this optional chapter without missing any background material that is needed for subsequent *Main* chapters.

4 Exercises

The exercises in the *Main* and *Background* units are created manually and don't come with solutions (although your instructor may distribute official solutions in class). The part *Additional practice exercises* contains automatically generated exercises, and those do come with solutions. These mostly test your command of very mechanical concepts, e.g. computing the union of two sets. The idea is to give you ample opportunities to practice these calculations until they are second nature to you. Since they are automatically generated, these exercises aren't exactly fun or stimulating. To be honest, they are quite the grind. But that's exactly the point: if you struggle with a basic concept, you can keep grinding it out one problem set at a time until it finally clicks for you.

5 Recap

- The book consists of multiple parts, each part consists of chapters, and chapters consist of units.
- Don't read the *Background* chapters from start to finish, it will be incredibly dull. Instead, work your way through the *Main* units and only check out a *Background* unit when it is listed as a prerequisite.
- Each unit comes has exercises interspersed with the main text. But if you want to practice more, check out the automatically generated exercises in *Additional practice exercises*.