

On Writing a Textbook on Natural Language Processing

Jacob Eisenstein

Google Research

jeisenstein@google.com

Abstract

There are thousands of papers about natural language processing and computational linguistics, but very few textbooks. I describe the motivation and process for writing a college textbook on natural language processing, and offer advice and encouragement for readers who may be interested in writing a textbook of their own.

1 Introduction

As natural language processing reaches ever-greater heights of popularity, its students can learn from blogs and tutorials, videos and online courses, podcasts, social media, open source software projects, competitions, and more. In this environment, is there still any room for textbooks? This paper describes why you might write a textbook about natural language processing, how to do it, and what I learned from writing one.

Summary of the book. This paper will not focus on the details of my textbook (Eisenstein, 2019), but I offer a brief summary for context. My main goal was to create a text with a formal and coherent mathematical foundation in machine learning, which would explain a broad range of techniques and applications in natural language processing. The first section of the book builds up the mathematical foundation from linear classification through neural networks and unsupervised learning. The second section extends this foundation to structure prediction, with classical algorithms for search and marginalization in sequences and trees, while also introducing some ideas from morphology and syntax. The third section treats the special problem of semantics, which distinguishes natural language processing from other applications of machine learning. This section is more methodologically diverse, ranging from logical to distributional semantics. The final section treats three

of the primary application areas: machine translation, information extraction, and text generation. Altogether this comprises nineteen chapters, which is more than could be taught in a single semester. Rather, the teacher or student can select subsets of chapters depending on whether they wish to emphasize machine learning, linguistics, or applications. The preface sketches out a few paths through the book for various types of courses.

2 Motivation and related work

In this section, I offer some reasons for writing a textbook, compare textbooks with alternative educational formats, and provide a few words of encouragement for prospective authors.

2.1 Why you might want to write a textbook

The first requirement is that you expect to enjoy the type of work involved: reading the most impactful papers in the field, synthesizing and curating the ideas these papers contain, and presenting them in a way that is accessible and engaging for students. One of the main contributions of a textbook over the original research material is the unification of terminology and mathematical notation, so it will help if you have strong opinions about this and an impulse toward consistency. Finally, writing a good textbook requires reading great textbooks to understand what makes them work, and I enjoyed having a reason to spend more time with some of my favorites (e.g., MacKay, 2003; Blackburn and Bos, 2005; Cover and Thomas, 2012; Sipser, 2012; Murphy, 2012).

A more respectable reason to write a textbook is to clarify and amplify your vision for the field. The writing process forces you to try to understand things from multiple perspectives and to identify connections across diverse methods, problems, and concepts. If you are an opinionated researcher or teacher, there are probably ideas that you think haven't gotten the credit they deserve or haven't

been presented in the right way. Maybe you think students should know more about some method or set of problems: for example, I felt that learning to think about NLP by doing paper-and-pencil exercises could help students avoid wasting a lot of time writing code that was conceptually flawed. A textbook is the perfect vehicle for grinding such axes, as long as you don't take it too far and you keep the focus on what will benefit the reader.

One more reason to write a textbook is that we really do need them: only a small handful of NLP textbooks have ever been written. It is true that the textbook market is somewhat “winner-take-all”: it is easiest to build a course around a textbook that is already widely in use, and hard to get teachers to change their materials. But different types of courses and students have different needs, and mature fields have dozens of books that target each of these audiences. Compared with the difficulty of finding a niche among the thousands of research papers written each year, a well-written NLP textbook is almost guaranteed to offer something valuable to a large number of readers.

2.2 Why I did it

Honesty requires some additional introspection about my real motivations. The project started because I felt unprepared to teach many topics in natural language processing, and could think of no better preparation than writing out some notes and derivations in my own words. I find it hard to focus on lectures that are based on slides, and I have noticed that many students seem to have the same difficulty. So I tried to write notes that would enable me to teach from a whiteboard.¹

A second motivation was to create a resource for my students. When I started teaching in 2012, there was really only one textbook that was sufficiently complete and contemporary to offer in a college-level NLP course: [Jurafsky and Martin \(2008, J&M\)](#).² But as an incoming faculty member, I was particularly eager to train graduate students as potential research assistants, and J&M was less mathematical than I would have liked for this purpose. My first approach was to have students read contemporary research papers and surveys,

¹A specific inspiration to my early notes were the teaching materials from Michael Collins, e.g., <http://www.cs.columbia.edu/~mcollins/loglinear.pdf>.

²There are also some outstanding books that were either too old ([Manning and Schütze, 1999](#)) or not quite aligned with my goals for the course (e.g., [Bender, 2013](#); [Bird et al., 2009](#); [Goldberg, 2017](#); [Smith, 2011](#)).

but this requires training, and students struggled with inconsistencies in notation and terminology across papers. I needed something that would give students a bridge to contemporary research, and decided I would have to write it myself.

These reasons added up to a set of course notes that I posted on Github, but not a textbook. After periodic nudges from editors over a period of several years (see [Table 1](#)), and some experience reviewing books and book proposals, I finally decided to submit a proposal of my own in 2017. At this time I was close to submitting my tenure materials, and writing a book seemed like a welcome change of pace. I had become friends with a group of professors in the humanities and social sciences who were sweating over their own book projects at the time, and I envied their focus on solo long-term work, which seemed so different from my life of bouncing from one student-led project to the next. And finally, I flattered myself to think that I would be able to write the book quickly from the material that I had amassed in five years of teaching — read on to learn whether this prediction was accurate. Overall, the book arose from a combination of impostor syndrome and irrational optimism, a recipe that may be at the heart of many writing projects.

2.3 Why not do something else?

When people find out that you are writing a textbook, you may receive suggestions for all sorts of better ways to communicate the same information. In the 2010s, there was great interest in online courses — particularly at Georgia Tech, which was then my home university — and I was urged to produce videos for such a course on natural language processing. Another possibility would have been to write a blog, which would be easier to keep current than a textbook, and would permit readers to post comments and questions (e.g., [Ruder, 2021](#)). Going further, tools like Jupyter notebooks ([Kluyver et al., 2016](#)) offer exciting new ways to combine writing, math, and code. Some intrepid authors have even written entire textbooks as collections of these interactive documents (e.g., [VanderPlas, 2016](#)). With all these alternatives, why write a traditional textbook on “dead trees,” (as one of my students put it)? Some reasons are more personal and others are practical. Here are three:

Longevity. Although much of the textbook will be obsolete in a few years, some parts may stand the test of time; there are topics for which I still

turn to my copy of [Manning and Schütze \(1999\)](#). Even if my book does not offer the best explanation of anything that anyone cares about in twenty years, I am glad to know that people will probably be able to read it if they want to. With more innovative online media, there is no such guarantee. Course videos may be available far into the future, but they are difficult to produce well, requiring an entirely different set of skills than the amateur typesetting capabilities that most academics acquire in the course of their studies.

Quality. The publication process brings in several people who help you write the best possible book: an editor who helps you choose the material and the high-level approach, reviewers who make sure the presentation is clear and correct, and a copy editor who finds writing errors. Perhaps because textbooks are rare, I also found that colleagues were very generous when asked to lend their expertise.

Finality. The field of natural language processing will surely continue to grow and evolve, and online media offers the temptation to try to keep pace with these changes. But if you agree to be bound by the conventional publishing process, there will come a day where you send a file to the publisher and are unable to make any further changes. While some authors seem to be happy (or at least willing) to continually revise through many editions over several decades (e.g., [Russell and Norvig, 2020](#)), I wanted the option to move on to other things.

While textbooks can be expensive, open access online editions are increasingly typical. In my case, I was able to negotiate a free online edition in exchange for a small portion of the royalties.

2.4 Yes, you

Before committing, I confessed to my prospective editor one of my deepest fears about the project: the best-known textbooks on natural language processing ([Manning and Schütze, 1999](#); [Jurafsky and Martin, 2008](#)) were written by true luminaries. Who was I to try to compete with them? Being a crafty and experienced editor, she replied that perhaps those authors were not so luminous before their textbooks, and wouldn't I like to write one and join them in the firmament? Although I am not so crafty, even I could see through this ploy. What ultimately gave me the courage to proceed was the realization that if I didn't write this *particular* textbook, then no one else would. AI summer was then

coming into full bloom, and the true luminaries had plenty of other things to keep them occupied. In any case, there is no minimum amount of luminosity required for writing a textbook: publishers will ask that you give some evidence that you know what you're talking about, but the main criterion is to have a compelling vision for a book that hasn't been written yet.

3 Methodology

Publishers seem keenly aware of the need for more textbooks in natural language processing and in AI more generally, and I found several editors that were eager to talk at conferences. I selected MIT Press because of their track record in publishing some of my favorite computer science textbooks. Other factors that you may wish to consider are the length of the review and production process, and the publisher's position towards open access. I was lucky to get feedback on the contract from another editor and from colleagues who have written books in other fields, but I did not think of negotiating with regard to electronic editions and translations. Fortunately the publisher was generous on these points, as they turned out to be a significant fraction of the revenue for the book. In any case, in the current environment of high demand for AI expertise, the financial compensation is not competitive with other uses of the same amount of time. You may find that it makes more sense to negotiate on aspects of the book and publishing process, such as length, open access, and support.

The publisher requires four main inputs from the author: a proposal, a complete draft for review, a "finished version" for copy editing and composition, and markup of page proofs. In the rest of the section, I'll describe how I approached each of these inputs. A timeline is given in [Table 1](#).

3.1 Proposal

The publisher required a proposal with two complete chapters (which were entirely rewritten later), a detailed table-of-contents for the rest of book, and a discussion of the imagined readership and the books that readers currently have to choose from. You will also give an estimate for some factors that affect the price: how long the book will be, how many figures to include, and whether color is required; and you will be asked to provide a time-

Fall 2012	Started teaching natural language processing and writing lecture notes.
July 2014	First contact with an editor.
2014-2017	Periodic nudges from the editor to please finish my book proposal someday.
March 2017	Book proposal done and sent out for review.
May 2017	Book proposal reviewed and accepted.
June 2017	Signed agreement with publisher.
Summer 2017-2018	Did most of the writing.
Early summer 2018	Solicited informal reviews of chapters from subject experts.
June 2018	Manuscript draft sent out for formal reviews.
Summer 2018	Wrote most of the exercises while awaiting reviews.
July 2018	Received reviews, started revisions.
November 2018	Revised manuscript sent out for production.
Winter 2019	Received and reviewed copy edits.
May 2019	Received and reviewed page proofs.
Summer 2019	I was supposed to make slide decks while waiting for the book to come out.
October 2019	Book is published.

Table 1: A rough timeline. Key inputs from section 3 are highlighted in bold.

line, which no one takes too seriously.³ As with anything else, it helps to see other proposals that have been successful, and you may ask your editor for positive examples. I spent a significant amount of time on the example chapters, and relatively little on the proposal itself, although it did help me to identify the overall structure of the book.

3.2 Draft

If the proposal is accepted, it's time to start writing. The purpose of this stage is to produce something that can be sent to the reviewers. In my case, the editor did not require the exercises or figures to be done at this stage; I have heard that other presses will solicit reviews on a chapter-by-chapter basis. After getting to a complete draft of each chapter, I also solicited informal reviews from friends and colleagues, which both improved the content and gave me far more confidence about the chapters that did not align with my expertise.

At first I tried to schedule the writing to align with teaching — for example, writing the chapter on parsing while teaching the same unit — but I wasn't able to keep up, and several chapters had to be left to the following summer. I hesitate to offer much writing advice to this audience, but I will pass along one thing I learned from Mark Liberman, when I asked how he was such a prolific blogger:⁴ it's possible to learn to write well if you

constrain yourself to write quickly, but it's much more difficult to learn to write quickly while constraining yourself to write well. So write quickly, and eventually the quality will catch up.

One regret about this stage is that I did not adopt the publisher's formatting templates. I had already written many pages of course notes, and when I couldn't immediately get them to compile against the publisher's format, I decided to put this off until later. Naturally that only made things much more difficult in the end, and I didn't use all that much of my original material anyway.

There are several reasons why my estimate of the completeness of the original course notes was too optimistic. While teaching, you are likely to emphasize the aspects of the subject that you know best. This means that the remaining parts to write are exactly those that are most difficult for you. In the classroom, you can rely on interactive techniques such as dialog and demonstrations to overcome weaknesses in the exposition of technically-challenging material, but the textbook must stand alone. Finally, the requirements for consistency, clarity, and accuracy of attribution in a textbook are much higher than the standard that I had reached in my course notes, and although the difference may seem small to many readers, it represents quite a lot of work for the writer. In total, I kept hardly any of the original text, although I was able to reuse the high-level structure of roughly half of the chapters.

³As my editor put it, "if missing deadlines was a crime, the prisons would be full of authors."

⁴<https://linguagelog.ldc.upenn.edu>

3.3 Revision(s)

The reviews were generally positive, but one reviewer was quite critical of the early chapters; although the publisher didn't require it, I made substantial changes based on this feedback. At this point I also tried to add a few notes about very recent work, such as BERT (Devlin et al., 2019), which appeared on arXiv while I was doing the revisions. Had the original reviews been more negative, the publisher might have required another round before accepting my revisions, but luckily this wasn't required in my case. The reviewers were very helpful, but I am skeptical that any of them read the whole thing, and I recommend seeking external reviews, especially for the later chapters that the reviewers are likely to skip or skim.

3.4 Proofs

The remaining steps involve details of the writing style and typesetting. At this stage I handed over the source documents to the production team, and could only communicate by adding notes to a PDF. This may be less of a technical requirement and more an incentive to prevent authors from introducing significant new content. The copy editing stage identified many writing problems, but the copy editor was unable to check the math. The publisher offered to pay for a math editor, but I was unable to find someone willing to do it. Fortunately, many of the mathematical errors had already been identified by students of my course. This stage also involved a bit of haggling about minor issues like whether it would be necessary for citations to include page numbers from conference proceedings, and when it was appropriate to use a term of art that violated the house style, such as “coreference” instead of “co-reference.”

Once the copy edits are complete, a LaTeX professional was able to compile the document using the publisher's format. This created a number of problems with the typesetting of the math, which were somewhat painstaking to check and resolve, and which could have been avoided if I had used the publisher's templates from the beginning. At this point the publisher is highly resistant to any changes to the content, but I did get them to fix some glaring errors that I found at the last minute.

4 Evaluation

What worked. It is difficult to be objective about a project of this scope. I am always happy to learn

that the book is being used in a course or for self-study, and was thrilled about translations into Chinese and Korean. The text seems best suited to classes that are similar to mine, where the primary goal is to train future researchers, who need a mathematical foundation in the discipline. I have been told that the exercises are particularly helpful, and I have received many requests for solutions, which I am happy to provide to teachers. There have been fewer reports of errors than I had expected, which I attribute to the careful reading of several classes of students while I was teaching from the unpublished notes.⁵ Offering the PDF online seems to have been an essential factor in the adoption of the textbook, especially given that the most popular alternative (J&M) is also freely available.

What could have been better. By the time the book appeared in print, there had been a number of significant changes in both the theory and practice of natural language processing. While this was expected, it is nonetheless hard not to be disappointed not to have put more emphasis on the topics that increased in importance — I'll say a bit more on this in the final section. Some readers feel that the term “introduction” in the title is misleading with regard to the amount of mathematical background that is expected. While the text assumes only multivariate calculus (and attempts to be clear about this expectation), the pace of the opening chapters is difficult for students who are out of practice. My editor was probably correct that adoption would be greater if I provided slides that professors could teach from, but I couldn't bring myself to make time for this tedious task after finishing the book.

5 Future work

As the field of natural language processing continues to progress, it is tempting to update the textbook with the latest research developments. For example, multilinguality and multimodality would deserve significantly more emphasis in a second edition, and a revision would have to reflect the maturation of applications such as question answering and dialog. But while some changes could be addressed by adding or modifying a few chapters, others — particularly the shift from conventional supervised learning to more complex methodologies like pretraining, multi-task learning, distilla-

⁵I maintain an errata page at <https://github.com/jacobeisenstein/gt-nlp-class/blob/master/notes/errata.md>

tion, and prompt-based learning — seem to require a more fundamental rethinking of the book’s underlying structure, particularly in a textbook that emphasizes a coherent mathematical foundation. Any such revisions would have to grow out of classroom teaching experience, which did so much to determine the shape of the first edition.

Acknowledgments. Thanks to William Cohen, Slav Petrov, and the anonymous reviewers for feedback.

References

- Emily M Bender. 2013. *Linguistic fundamentals for natural language processing: 100 essentials from morphology and syntax*, volume 6 of *Synthesis lectures on human language technologies*. Morgan & Claypool Publishers.
- Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O’Reilly Media, Inc.
- Patrick Blackburn and Johan Bos. 2005. *Representation and inference for natural language: A first course in computational semantics*. CSLI Publications.
- Thomas M Cover and Joy A Thomas. 2012. *Elements of Information Theory*. John Wiley & Sons.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. **BERT: Pre-training of deep bidirectional transformers for language understanding**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jacob Eisenstein. 2019. *Introduction to natural language processing*. MIT press.
- Yoav Goldberg. 2017. Neural network methods for natural language processing. *Synthesis lectures on human language technologies*, 10(1):1–309.
- Dan Jurafsky and James H Martin. 2008. *Speech and language processing*, 2nd edition. Prentice Hall.
- Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, and Jupyter development team. 2016. **Jupyter notebooks — a publishing format for reproducible computational workflows**. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press.
- David J. C. MacKay. 2003. *Information theory, inference and learning algorithms*. Cambridge university press.
- Christopher Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.
- Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.
- Sebastian Ruder. 2021. **Recent advances in language model fine-tuning**. <https://ruder.io/recent-advances-lm-fine-tuning/>.
- Stuart Russell and Peter Norvig. 2020. *Artificial intelligence: a modern approach*, 4th edition. Pearson.
- Michael Sipser. 2012. *Introduction to the Theory of Computation*. Cengage learning.
- Noah A Smith. 2011. Linguistic structure prediction. *Synthesis lectures on human language technologies*, 4(2):1–274.
- Jake VanderPlas. 2016. *Python data science handbook: Essential tools for working with data*. O’Reilly Media, Inc.