

# Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications

By James Pustejovsky, Amber Stubbs

 Download

 Read Online

## Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs

Create your own natural language training corpus for machine learning. Whether you're working with English, Chinese, or any other natural language, this hands-on book guides you through a proven annotation development cycle—the process of adding metadata to your training corpus to help ML algorithms work more efficiently. You don't need any programming or linguistics experience to get started.

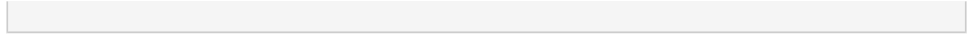
Using detailed examples at every step, you'll learn how the *MATTER Annotation Development Process* helps you **Model, Annotate, Train, Test, Evaluate, and Revise** your training corpus. You also get a complete walkthrough of a real-world annotation project.

- Define a clear annotation goal before collecting your dataset (corpus)
- Learn tools for analyzing the linguistic content of your corpus
- Build a model and specification for your annotation project
- Examine the different annotation formats, from basic XML to the Linguistic Annotation Framework
- Create a gold standard corpus that can be used to train and test ML algorithms
- Select the ML algorithms that will process your annotated data
- Evaluate the test results and revise your annotation task
- Learn how to use lightweight software for annotating texts and adjudicating the annotations

This book is a perfect companion to O'Reilly's *Natural Language Processing with Python*.

 [Download Natural Language Annotation for Machine Learning: ...pdf](#)

 [Read Online Natural Language Annotation for Machine Learning ...pdf](#)



# Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications

By James Pustejovsky, Amber Stubbs

## Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications

By James Pustejovsky, Amber Stubbs

Create your own natural language training corpus for machine learning. Whether you're working with English, Chinese, or any other natural language, this hands-on book guides you through a proven annotation development cycle—the process of adding metadata to your training corpus to help ML algorithms work more efficiently. You don't need any programming or linguistics experience to get started.

Using detailed examples at every step, you'll learn how the *MATTER Annotation Development Process* helps you **Model**, **Annotate**, **Train**, **Test**, **Evaluate**, and **Revise** your training corpus. You also get a complete walkthrough of a real-world annotation project.

- Define a clear annotation goal before collecting your dataset (corpus)
- Learn tools for analyzing the linguistic content of your corpus
- Build a model and specification for your annotation project
- Examine the different annotation formats, from basic XML to the Linguistic Annotation Framework
- Create a gold standard corpus that can be used to train and test ML algorithms
- Select the ML algorithms that will process your annotated data
- Evaluate the test results and revise your annotation task
- Learn how to use lightweight software for annotating texts and adjudicating the annotations

This book is a perfect companion to O'Reilly's *Natural Language Processing with Python*.

## Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications

By James Pustejovsky, Amber Stubbs **Bibliography**

- Sales Rank: #400941 in Books
- Brand: Brand: O'Reilly Media
- Published on: 2012-11-04
- Released on: 2012-11-01
- Original language: English
- Number of items: 1
- Dimensions: 9.19" h x .73" w x 7.00" l, 1.21 pounds
- Binding: Paperback
- 342 pages

 [Download Natural Language Annotation for Machine Learning: ...pdf](#)

 [Read Online Natural Language Annotation for Machine Learning ...pdf](#)

## Download and Read Free Online Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs

---

### Editorial Review

#### About the Author

James Pustejovsky teaches and does research in Artificial Intelligence and Computational Linguistics in the Computer Science Department at Brandeis University. His main areas of interest include: lexical meaning, computational semantics, temporal and spatial reasoning, and corpus linguistics. He is active in the development of standards for interoperability between language processing applications, and lead the creation of the recently adopted ISO standard for time annotation, ISO-TimeML. He is currently heading the development of a standard for annotating spatial information in language. More information on publications and research activities can be found at his webpage: [pusto.com](http://pusto.com).

Amber Stubbs recently completed her Ph.D. in Computer Science at Brandeis University, and is currently a Postdoctoral Associate at SUNY Albany. Her dissertation focused on creating an annotation methodology to aid in extracting high-level information from natural language files, particularly biomedical texts. Her website can be found at <http://pages.cs.brandeis.edu/~astubbs/>

### Users Review

#### From reader reviews:

##### Jeff Sanchez:

Do you considered one of people who can't read satisfying if the sentence chained inside the straightway, hold on guys this specific aren't like that. This Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications book is readable by simply you who hate the straight word style. You will find the information here are arrange for enjoyable studying experience without leaving also decrease the knowledge that want to supply to you. The writer regarding Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications content conveys thinking easily to understand by most people. The printed and e-book are not different in the written content but it just different in the form of it. So , do you continue to thinking Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications is not loveable to be your top listing reading book?

##### Dustin Davis:

The reserve untitled Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications is the e-book that recommended to you to learn. You can see the quality of the book content that will be shown to you. The language that publisher use to explained their ideas are easily to understand. The writer was did a lot of investigation when write the book, hence the information that they share to you personally is absolutely accurate. You also might get the e-book of Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications from the publisher to make you a lot more enjoy free time.

**Nick Gulbranson:**

Do you like reading a book? Confuse to looking for your chosen book? Or your book seemed to be rare? Why so many question for the book? But virtually any people feel that they enjoy to get reading. Some people likes reading through, not only science book but also novel and Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications or even others sources were given know-how for you. After you know how the good a book, you feel wish to read more and more. Science guide was created for teacher or even students especially. Those publications are helping them to put their knowledge. In different case, beside science guide, any other book likes Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications to make your spare time a lot more colorful. Many types of book like here.

**Brant Castillo:**

Guide is one of source of knowledge. We can add our know-how from it. Not only for students but additionally native or citizen have to have book to know the update information of year to be able to year. As we know those ebooks have many advantages. Beside all of us add our knowledge, could also bring us to around the world. By book Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications we can acquire more advantage. Don't you to definitely be creative people? To get creative person must prefer to read a book. Merely choose the best book that acceptable with your aim. Don't always be doubt to change your life with that book Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications. You can more attractive than now.

**Download and Read Online Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs #0T8Q54O9FCA**

## **Read Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs for online ebook**

Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs Free PDF d0wnl0ad, audio books, books to read, good books to read, cheap books, good books, online books, books online, book reviews epub, read books online, books to read online, online library, greatbooks to read, PDF best books to read, top books to read Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs books to read online.

### **Online Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs ebook PDF download**

**Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs Doc**

**Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs Mobipocket**

**Natural Language Annotation for Machine Learning: A Guide to Corpus-Building for Applications By James Pustejovsky, Amber Stubbs EPub**