

ANLP 2023-24 Course Syllabus

Adam Lopez, Edoardo Ponti, and Shay Cohen; School of Informatics, University of Edinburgh

Welcome to Accelerated Natural Language Processing!

This syllabus provides an overview of the course content, what is expected of you, how to make the most of this course, and the tools and structure we plan to use to deliver the course this year. You should read through the whole syllabus, as you are expected to be familiar with the information in it.

Contents

Course overview	2
Our partnership with you.....	3
A typical week’s schedule and deadlines.....	4
Course materials, activities, and assessment.....	4
Content units.....	4
Reading.....	4
Self-test quizzes.....	4
Tutorial exercises	5
Labs	5
Tutorial group meetings (discussion groups)	5
Assessment	5
How to get the most out of this course.....	5
Communication and where to get help	6
How we will communicate with you	6
Where to get help	6
A final word.....	7

Course overview

The course will run for 10 weeks, covering the topics as listed below. The precise order is subject to change.

1. Introduction, words, and morphology
 - Including: ambiguity, words, and morphology
2. Basic algorithms
 - Including: finite state models, dynamic programming, and edit distance
3. Language models
 - Including: probability, N-gram models, smoothing, perplexity
4. Text classification
 - Including: Naïve Bayes, logistic regression, basic neural networks, evaluating classifiers
5. Tagging and hidden Markov models
 - Including: parts of speech, HMMs, Viterbi algorithm
6. Phrase-structure syntax and parsing
 - Including: syntactic structure, context free grammars, and the CKY algorithm
7. Beyond basic grammars
 - Including: probabilistic grammars and parsing, dependency parsing, evaluating parsers
8. Lexical semantics and word embeddings
 - Including: word senses, distributional semantics, mutual information, word embeddings
9. Sentence-level semantics
 - Including: logical forms, coreference
10. Ethics & bias, exam review
11. Revision week (no new content)

The content of the course builds up from the smaller pieces of language (words and morphemes), through the ways that words can combine structurally (syntax) to how those structures create meaning within sentences (semantics) and across sentences (discourse).

At the start of each week, the content and activities for that week will become available on Learn and OpenCourse, and we'll provide instructions to guide you through them. Each week, we typically start by presenting linguistic information about that aspect of language, and then we discuss computational approaches (models and algorithms). This approach aligns with the Learning Outcomes (see below), which involve knowledge and skills related to both linguistic and computational aspects of NLP.

But as you can see from the Learning Outcomes, this course is about more than just facts and procedures. To do good work in NLP, you will need to understand the pros and cons of different approaches, and which methods are appropriate when. You will also need to consider the broader context in which NLP systems are used, and potential implications for society. So, at various points in the course, we will also discuss these broader methodological and ethical topics.

Learning Outcomes

On successful completion of this course, you should be able to:

1. Identify, construct, and analyse examples of different kinds of ambiguity in natural language (e.g., ambiguity in part-of-speech, word sense, syntactic attachment). Explain how ambiguity presents a problem for computational analysis, and some of the ways it can be addressed.
2. Describe and apply standard sequence and classification models; describe parsing and search algorithms for different levels of analysis (e.g. morphology, syntax, and semantics) and simulate each algorithm step-by-step with pen and paper.
3. For a range of NLP tasks, outline a processing pipeline for that task, including standard data sets, models, algorithms, and evaluation methods. Given a particular pipeline or part of the pipeline, identify potential strengths and weaknesses of the suggested dataset/method (including both technical and ethical issues, where appropriate), and provide examples to illustrate.
4. Implement parts of the NLP pipeline with the help of appropriate support code and/or tools. Evaluate and interpret the results of implemented methods on natural language data sets.

Our partnership with you

We get students from many different backgrounds in this course, and each of you brings something different, whether it is examples from your native language, the ability to design or critique an experiment, or programming skills. One of the best ways to learn about the parts you are less familiar with is by interacting with each other, and we will provide opportunities to help you do that. Most exercises we give you are formative: your course mark does not depend on whether you got the right answer, and we encourage you to work with other students on these types of exercises. The fact that your course mark does not depend on the outcome *does not mean* that the exercises are optional, however: if you want to do well on the exam, you must practice, and the exercises are designed to give you (some of) that practice.

However, we will also assign some work to assess your knowledge and skills, either individually or with a pre-specified partner. In these cases, you may not get help from other people unless we explicitly say so. This applies whether those people are friends, other students, or outside services.

Think of this class as a partnership between us and you. We agree to:

- provide materials and support that will help you to learn;
- set assessments that will fairly assess the learning outcomes

In turn, you must agree to:

- devote the time needed to work through challenging materials;
- maintain standards of academic integrity in your work

Academic integrity covers many areas, but as a few examples, **you may not**:

- claim other people's work as your own. note that this includes the output of systems like ChatGPT, since these systems are simply synthesising other people's work;
- make your solutions to assessed work available (to other students or more broadly), *either before or after the submission deadline*, unless permission is explicitly given by the instructors;
- post or otherwise redistribute course materials outside the course, unless permission is explicitly given on the materials or by the instructors;
- engage in other forms of academic misconduct.

For more details, see the [School page on academic misconduct](#).

Course materials, activities, and assessment

This course is delivered using a mixture of online materials and timetabled in-person activities. The main content is delivered through lectures and assigned readings, with further timetabled activities used for interactive learning in the form of tutorials and labs.

Lectures and Readings

Each week there will be three lectures focusing on the main content of the course, and some associated readings. The readings and lectures complement each other, and you are responsible for understanding the material in both. The readings typically contain more detail, while lectures focus more on motivation and examples, or views not covered in the reading.

Reading

The main textbook for this course is *Speech and Language Processing* by Jurafsky and Martin, which you can find in Learn under Library Resources-> Resource List. In the reading assignments, **JM3** refers to the draft third edition and **JM2** is the second edition. You can also find individual chapters of JM3 on Jurafsky's website, but these sometimes update unexpectedly in the middle of the semester. If you download the whole PDF from our Resource List, you'll be sure to have the same version we are using (from Jan 2023).

We have permission to distribute the JM2 and JM3 PDFs to our class, but you may not upload them to other sites or otherwise re-distribute them.

When we refer to sections of the textbook, section 0 refers to whatever introductory material comes before section 1.

Past students have requested that we prioritize readings, so high priority readings are marked with a (*). If you are really short on time you should focus on these and return to others later. But you are ultimately responsible for all material in the assigned reading, and keeping up is the best strategy! Some students find that it is useful to do the reading before watching the lecture, while others prefer to do it afterward. Please do what works best for you, and if you aren't sure, experiment.

We also provide some optional readings for the course. These may help broaden or deepen your understanding, but are not required in order to do well in the course.

Quizzes

Most lectures have an associated quiz in gradescope, which you should complete during your independent study time following the lecture. Quiz questions are intended to be fairly straightforward, and should not take much time or computation to answer if you understand the material. The quizzes are to help you check your own understanding, so you will be able to immediately see if your answer is right or wrong, and you can change your answer as many times as you want. These quizzes are for formative feedback, and your course mark does not depend on whether you got the right answer or not. However, if it looks like you are falling behind on the quizzes, we may check in with you to see if you're having difficulties.

Self-assessment exercises

In most weeks, we'll also distribute a worksheet of self-assessment exercises. These exercises will typically require you to solve problems to arrive at a particular solution, but a few of them may be more open-ended. You will normally be expected to work through the tutorial exercises by the Monday following the week when they were distributed. You do not need to submit your solutions, . We will also post solutions with a bit more explanation on Monday mornings.

While it may be tempting to check solutions before working through the problems yourself, that is not a good way to learn the material you need to know for the exam; it is too easy to fool yourself into believing that you would have gotten the right answer. Please work through the exercises and ask questions if you get stuck, or if you are still confused even after seeing the solutions. You can ask your fellow students or on the Piazza discussion forum, and we'll normally have plenty of time for questions at the weekly live session as well. Only once you have attempted the problem yourself should you review the solution.

Labs

In some weeks, you will need to do a lab (computer-based exercises). Labs are designed to be done in pairs, so feel free to sit next to someone and introduce yourself when you arrive in the lab! The labs will help you practice working with language data and using tools that are needed for the assignments, but your answers to the lab questions need not be submitted and they do not count toward your final mark.

Tutorial group meetings (discussion groups)

In other weeks, you'll meet in a tutorial group of about 16-20 students with a tutor. In these weeks, we'll distribute some more open-ended questions for you to consider and discuss in your groups, with guidance from the tutor. These questions may ask you to consider your own examples or opinions, perhaps different from other students'. The questions will be distributed in advance, and in some cases you will need to do a bit of preparation before the meeting to make participation worthwhile.

Assessment

Assessment for the course will consist of:

- Two assignments, requiring a mixture of programming and report writing. Worth 15% each.
- A final exam, worth 70%.

Deadlines for the assignments are listed under Assessments on Learn. Exam dates are arranged centrally by the University and are normally published late in October.

How to get the most out of this course

You already have lots of experience at university, but perhaps in a different context or country. Or this may be your first time doing it! Here are some tips to help.

Block out time and create a schedule. This course has only around 3-4 hours of required timetabled activities each week, but a lot of self-study materials. This provides flexibility but can also be a big challenge, because it's easy to fall behind. To help you, we have provided suggested deadlines for what you should have completed by each day of the week (see the "Typical Week's Schedule" above). We strongly recommend that you set aside specific times each day to work on this class, get into a habit

and stick to it. You should expect to spend about **12-13 hours per week** on this class, so start by blocking out that time. If you find the course is much easier or harder than expected, you can always adjust later.

Be an active learner. Unfortunately, just reading a book or watching videos is not enough to learn the material. Research shows that you will learn best by actively engaging with the material. One way to do that in a traditional lecture is to take notes, which forces you to process the material and decide what's important as you write. It's tempting to not take notes when viewing recorded lectures or reviewing slides, because you can "always go back to them later". But we suggest that you do still take notes, either during the lecture or immediately afterward. Write down what you think are the most important parts and any questions you have. You may be able to resolve your questions by reading the textbook; if not, post to the class forum, or bring them to your discussion group or TA drop-in hours.

We've also provided a lot of other materials to help you be an active learner: the quizzes, labs, and tutorials are all ways for you to actively engage, and some of the videos also have questions for you to think about before you move to the next video. It may feel easier to skip some of these opportunities, but learning effectively does take effort!

Ask questions...and answer them too. Whether you are stuck or just wondering about something, please don't be shy about asking. There are different ways to ask questions (see the Communication section below), so find the one(s) that suit you. We encourage you to read the class forum (Piazza) regularly, and especially to answer other students' questions if you can. Explaining something to another person is one of the best ways to make sure you understand it yourself, and you'll also help them out!

Engage with other students. We will provide ways for you to meet other students---for example, your lab and assignment partners and your discussion group members. You can choose to talk to these people only as much as minimally required, but it's well worth putting in a little extra effort to connect more often, or to seek out other study partners, or even just to answer some of your classmates' questions on the class forum. It will make you feel more engaged with the class and give you more opportunities for learning.

Communication and where to get help

How we will communicate with you

Course announcements will be posted to Learn. Once you are registered for the course, you'll also get these Learn announcements in your email. Until then, make sure to check Learn at least once a day.

Course staff will also monitor Piazza to respond to (private or public) questions posted there. In general, you should not expect immediate replies from course staff on piazza; there are far more students than course staff. Instead, we encourage you to answer questions from other students, which is a great way to test your own learning. Course staff will be on the Forum at least once a day on weekdays, and correct any misunderstandings that arise in student answers.

Where to get help

Please see the Course Help and Contacts section of the ANLP Learn site, which will tell you who to contact depending on the type of question or concern you have about the course.

A final word

Please remember to be kind: to yourself, to other students, and to us. Many of us are dealing with the uncertainties of the larger world, and there will undoubtedly be unexpected hiccups. We are also working with several new university systems this year. We are doing our best to make this course work for everyone, and we hope you will too!