

# Natural Language Processing Engineer interview questions

This **Natural Language Processing Engineer** interview profile brings together a snapshot of what to look for in candidates with a balanced sample of suitable NLP interview questions.

## Natural Language Processing Engineer Interview Questions

A **Natural Language Processing (NLP) Engineer** develops products that rely on the intelligent processing of human language by a computer. Example applications include building an intelligent tutor, a system that automatically summarizes news articles or a one that recognizes and understands human speech. Beyond a very strong basis in natural language processing, an ideal candidate will be strong in related fields such as machine learning, text mining, information theory and information retrieval. These natural language processing interview questions will also test familiarity with specialized tools and experience with projects working with natural language data such as nltk (Python), Apache OpenNLP or GATE. Knowledge of linguistics is often a very big plus. Fluency in one or more foreign languages is equally valuable. This tends to be a very technical role and so research skills can also be very important. Computer science is typically the background of choice for this type of role but some very successfully approach this with a linguistics background that emphasizes computational linguistics.

### Role-specific questions

*(Natural language processing)*

- What is part of speech (POS) tagging? What is the simplest approach to building a POS tagger that you can imagine?
- How would you build a POS tagger from scratch given a corpus of annotated sentences? How would you deal with unknown words?
- How would you train a model that identifies whether the word “Apple” in a sentence belongs to the fruit or the company?
- How would you find all the occurrences of quoted text in a news article?
- How would you build a system that auto corrects text that has been generated by a speech recognition system?
- What is latent semantic indexing and where can it be applied?
- How would you build a system to translate English text to Greek and vice-versa?
- How would you build a system that automatically groups news articles by subject?
- What are stop words? Describe an application in which stop words should be removed.
- How would you design a model to predict whether a movie review was positive or negative?

*(Related fields such as information theory, linguistics and information retrieval)*

- What is entropy? How would you estimate the entropy of the English language?
- What is a regular grammar? Does this differ in power to a regular expression and if so, in what way?
- What is the TF-IDF score of a word and in what context is this useful?
- How does the PageRank algorithm work?
- What is dependency parsing?
- What are the difficulties in building and using an annotated corpus of text such as the Brown Corpus and what can be done to mitigate them?

*(Tools and languages)*

- What tools for training NLP models (nltk, Apache OpenNLP, GATE, MALLET etc...) have you used?
- Do you have any experience in building ontologies?
- Are you familiar with WordNet or other related linguistic resources?
- Do you speak any foreign languages?