1. **What is Part of Speech Tagging?**

This is a method for analyzing sentences where you categorize each word in the sentence, for instance into nouns and verbs. Part of Speech Tagging is not that useful in itself, but it can be the first step towards solving a number of issues, like parsing and interpreting a sentence, or examining if it has the correct syntax.

For example (boat,N), (walk,V) where N = noun and V = verb.

Languages that do not have high inflectional morphology, such as English, are inclined to ambiguity, and hence it is difficult to do this process in those languages. Ambiguity can also happen in languages that are tonal, for example during verbalization, like Chinese.

1. **Explain Morphology.**

Morphology studies the internal structure of words and divides them into smaller pieces, called stems or morphemes. Some words only have one morpheme. They are categorized as simple words. Words with several morphemes are called complex words. An example of a simple word is ”walk”, and one of the complex words derived out of it is”walked”.

1. **Explain the difference between semantics and pragmatics.**

Semantics studies the literal meanings of words and phrases. You derive the meaning according to grammar and vocabulary.

Pragmatics is when different words have a special meaning in an environment. A word can have other meaning in some environments and can be dependent on information of previous conversations.

1. **What is the difference between semantic and syntax?**

Syntax describes the grammatical structure of a sentence and if this is correct. Semantic on the other hand describes the meaning of sentence. This difference is important in context of if a sentence is “correct”. For example the sentence “A fish walks along the street and watches a T-Rex”. This sentence is grammatically completely correct. This means it is incontext of the syntax correct. But in context of the semantic it does not make any sense.

1. **What is the goal of a syntactic analysis, and how can you perform it?**

It's generally called parsing. It can be used to check that the word order or syntax in a sentence is correct. You analyze the words in a sentence and divide them into categories, and then depending on the language, you make sure that it follow a certain order. A common issue is that if you use the correct word type order, sometimes the sentences may still not be valid, because some words may not be valid in that specific context.

1. **What is the general goal of Natural Language Processing in AI?**

It can be used to both try to make and interpret words or sentences. This can involve both speech or written text.

1. **What is a context free grammar?**

Based on the context free grammar rules a sentence is broken into structured parts that the system can easily understand. A context free grammar consists of a set of non-terminal symbols (noun phrase, verb phrase), a set of terminal symbols (words of the text), a set of rules (left hand side is a non-terminal and the right hand side consists of one or more non-terminal or terminal symbols) and a start symbol.

1. **What is the Backus-Naur form?**

It is a notation technique for context-free grammars and is often used to describe the syntax of computer programming languages, document formats and communication protocols. It defines terminal symbols, nonterminal symbols, start symbol and rewrite rules and is used wherever exact description of languages are needed.

1. **What is the difference between a chatbot and NLG? (NLG förklaras inte alls i svaret)**

A chatbot is an artificial intelligence-powered piece of software in a device (Siri, Alexa, Google Assistant etc), application, website or other networks that try to identify consumer’s needs and then assist them to perform a particular task like a transaction, hotel booking etc . Chatbots can be rule-based, where answers are based on predefined rules, which leads to a predetermined scripted conversation. Secondly, chatbots can also be self learning bots, based on machine learning approaches and are seen as more generative. For this task, NLP is utilized to derive meanings from a human language in a smart and useful way.

1. **What is the difference between natural language understanding and natural language generation?**

Natural language understanding (NLU) is about analyzing the input data and mapping it into a useful representation. An example for NLU is the app “Photomath”, where you make a picture of an equation a get a step by step solution as outcome. Natural language generation (NLG) transform data into understandable natural language text. An example for NLG would be an automatic generated form letter.

1. **What are problems in NLP?**

First of all language is highly ambiguous and must be disambiguated. Different accents, slangs, pronunciations, typos,...

Polysemy: Words in natural language usually have a fair number of different possible meanings (plant, ...). Hence the proper sense of each ambiguous word in a sentence must be determined.

1. **Where are NLP system used?**

* text processing - word processing, e-mail, spelling and grammar checkers
* Interfaces to databases - query languages, information retrieval, data mining, text summarization
* expert systems - explanations, disease diagnosis
* linguistics - machine translation, content analysis, writers' assistants, language generation
* Autocorrect, personal assistants, voice recognition, translators, and search engines.

1. **Why is ambiguity an issue in NLP?**

Ambiguity is an issue because ambiguous words or sentences can’t be handled with simple computer rules and requires context to extract the meaning from, and since most text parsing is context free it can be hard to do.

1. **Why is ambiguity an issue for NLG?**

Lexical ambiguity is common in natural language and short of trying to circumvent every single ambiguous word there is a risk of the NLG generating a sentence that can be misinterpreted.

1. **Explain why a rule-based NLP is unsuited for generalized use.**

A rule-based NLP is made by manually constructing rules to parse language,since all natural languages contain large amount of ambiguity and irregularities it would be a near infinite job to make rules to handle it all.

1. **Machine translation is said to be a AI-complete problem, what is meant by this?**

It means that solving the problem is essentially equivalent to solving the central problem of making an AI as intelligent as a person, this is because formachine translation to be perfected it requires all the kinds of information humans possess such as grammar, semantics and context**.**

1. **Explain why is language processing is hard?**

Language processing is hard because of the complexity in the human language and the fact that we have so many different languages. One of the hardest things is that we can imply things in a sentence without actually saying it. For example, “I'm hungry” can imply that you want that someone to make some food for you and not just stating that you are hungry.

1. **Explain why the Sentence -> sentence conjunction sentence, representation effective for parsing?**

If a sentence can be represented by a sentence, conjunction and a sentence then we can parse an infinite long sentence by just splitting it in the conjunctions. A sentence can also be a NP, VP (noun phrase then verb phrase). This representation does also have some drawbacks, if the computer can follow this rule it can in theory generate an infinite long sentence, which will be hard for humans to read and understand.

1. **What is natural language processing?**

Natural language processing (NLP) is an Artificial Intelligence method where natural language is given as input to a computer system. The system then process tasks on the given input and delivers output. Input and output can be a speech or written text. There exists two forms of NLP, Natural language understanding (NLU) and Natural language generation (NLG).

1. **How does parsing a natural language work?**

The sentence “Tom ate an apple” will be determined based on the context free grammar rules. Starting with the start symbol “sentence”. The sentence consists of a noun\_phrase and averb\_phrase. Based on our rules a noun\_phrase can be a proper\_noun or a determiner and a noun.

The verb\_phrase is a combination out of a verb and a noun\_phrase. The terminals are then mapped to the non-terminals.

1. **What problems might occur during natural language processing?**

Lexical ambiguity - where the system don’t know if a word is a noun or a verb. For example the word ‘match’: “These two samples match” or “I will light this candle with a match”.

Syntax level ambiguity – A sentence can be understand in different ways. Example: “The chicken is ready to eat”.

Referential ambiguity – When more than one person is involved in a situation and one person is doing something and it’s only referring via pronouns. You can’t be sure which person is meant.

1. **Take the text “Barcelona was on fire last night. They destroyed the enemy team”. What would a computer understand?**

A computer is not able to “read between the lines”. It would take the sentences literally and think that the city of Barcelona burned last night and “they” killed the enemy team. It is notable to understand the context that “Barcelona” means a football team.

1. **Why is NLP part of AI? Which part of NLP is AI?**

Especially Stemming and Lemmatization but also the tokenizing of sentences is are based on machine learning. NLP is part of AI because of different reasons. The first is that the previous described parts using AI techniques.

According to the description “everything that seems to be intelligent is an AI” it is an AI because it seems intelligent.

1. **What is the difference between Lemmatization and Stemming?**

Stemming only cuts words in pieces to get the stem of a word.

Lemmatization is considered to be a task where only inflectional endings are removed in order to obtain the origin(base dictionary) form of the word, and that word is also known as lemma.

For example the word “studies”. Stemming would form it to “studi” and “es”. Lemmatization would form it to “study”.

1. **What is stemming, regarding information retrieval in NLP?**

Stemming is where the end of words are removed to get a better search term when looking for information. An example is that a user may search for the phrase “where are elephants?”, the system may search for articles containing just “elephant” and find better results than just the files containing the plural form.

1. **Explain some of the steps that happen during language parsing.**

1. Tokenizing - The text is tokenized in words and sentences.(Tokenization is a low level process of NLP where one identifies individual words or punctuations within a sentence. Texts often contain tokens/characters that separate different content, an example is forward slash in the following: “10 mg/day”)

2. stemming/lemmatization - The word is transformed back to its basic form or its stem.

3. part-of-speech tagging - In this step the words are marked with their type, for example noun or verb. Also connections between words are made.

1. **Explain: What is “Stop words”.**

Stop words are filtered out from the text processing. They are commonly used but has little impact to the meaning of the text. eg. “the” & “is”

1. **What are 2 ways of parsing natural language?**

Using a parsing tree – Words divided into verb phrases and noun phrases and then articles, verbs etc.

Transition networks – Here several transition networks exist which represent rules, they are like a directed graph where you follow the nodes which represent categories such as verb, adjective etc. And like in logical programming it takes a sentence into one of these networks and pattern match each word in order with each of the nodes, if it does not match a transition network it moves on to the next one.

1. **What are some issues you can face with parsing speech?**

One issue is that many words have different meaning, duck, for example can be an animal or an action.(N or V)

Another issue is that some sentences can be interpreted differently depending on what you emphasise, “I carried the girl with the spade” could mean that I either carried a girl which had the spade or that I carried the girl using the spade.   
A third issue is that some sentences need context, for example “John gave Bob the sandwich. He smiled.”. If only the second of those sentences was considered it is impossible to say who smiled.

1. **What types of NLP are there?**

Natural language processing can be classified into rule-based and statistics-based. The Former one was primarily used until 1990s, and it relied on the set of hand-written rules.The latter paradigm, dependent on machine learning techniques makes use of statistical inference in order to learn new rules by itself, by checking multiple sources, i.e.documents and real world examples.

1. **Explain the difference between Constituency Parsing (or Deep Parsing) and Dependency parsing.**

Constituency Parsing(Deep Parsing) is a type of parsing that provides even more information to the part-of-speech tagging. Constituency parsing tree divides the text into sub-phrases, where the edges are not labeled, and the terminals are the words in a sentence. On the other hand, dependency parse is based on relating words with respect to their relation. Each vertex represents a word here.

1. **What is Sentiment analysis?**

The key of sentiment analysis is to analyse a certain text and try to grasp/understand the opinion expressed by it. Needless to say, it mostly works on subjective texts. The final sentiment of the text is determined with respect to the text’s polarity score, and it is positive, neutral or negative. There are 2 approaches: Supervised machine learning or deep learning approaches and Unsupervised lexicon-based approaches.

1. **Explain the differences between rule-based NLP vs statistical NLP**

Since the introduction of machine learning in natural language processing has leaned more and more towards using statistical models instead of hand crafted rules. When using a statistical method it is easy during the learning procedure for the method to focus on the most common cases. When you craft your own rules for your NLP it is hard to devise were the most of

the programs effort should be put. With the automatic learning procedure you can have when using a statistical method, you will also get a more robust system. A system that will be able to handle words it has never seen before or misspelled words. Making a soft system using hand written rules is very difficult. Systems based on automatically learning the rules are also much easier to improve. You only need to provide more and better input data which is relatively easy to do. On the other hand, to improve a

system were you have hand crafted the rules, you can only improve the system by making your rules better. Making a rule more complex and more suited for all your cases is much harder than simply supplying more cases.

1. **What is Lexical ambiguity and what is Syntax level ambiguity?**

Lexical ambiguity is when a word has multiple meanings. For example, “board”, it can be a noun as well as a verb. Syntax level ambiguity is when a sentence can be parsed in several ways.

1. **Briefly explain the five steps of NLP: Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Integration and Pragmatic Analysis.**

Lexical Analysis involves identifying and analysing the

structure of words in a text. Lexical analysis divides a chunk of text into paragraphs, sentences and words.

Syntactic Analysis involves analysis of words within a sentence to find the relationships between the words. It can reject grammatically wrong sentences.

Semantic Analysis involves drawing the dictionary meaning from the text. The text is checked for meaningfulness by mapping syntactic structures to objects in the task domain.

Discourse Integration involves connecting the

meaning of a sentence with the context of previous sentence in the text.(Never mentioned in FL)

Pragmatic Analysis involves re-interpreting the text to find out what is actually meant. This requires real worlds knowledge.

1. **Discuss the different grammars involved in NLP?**

The following grammars are involved in NLP: Noun phrases (NP), verb phrases (VP), sentences (S), determiner (DET), noun (N), transitive verb (TV), intransitive verb (IV), preposition (PREP), prepositional phrase (PP), adjective (ADJ)

1. **How are Convolutional Networks connected to NLP? Describe CNN and in what subject it is used in NLP?**

CNN is used in vectorization, where you contribute to “better learning of shared parameters”. In a matrix each vector could represent a word and use different filters to contribute the output. CNN:s are usually used in image recognition, but also used in NLP because it is fast.

1. **Describe the terms sentence generation and text realization. What Subject do they belong to?**

They belong to the subject Generation - “the process of producing meaningful phrases and sentences” . Sentence generation is the ability to choose required words to form meaningful phrases/sentences. Text realization is how to provide a sentence structure from sentence generation (put together sentences).

1. **Why do you have to be careful to use too much data with a lexicon in an NLP?**

With too much data it is hard to get generalisation in the NLP system.

1. **Explain the difference between natural language and formal language?**

Natural languages are ambiguous where a sentence can have more than one meaning and the correct one can be difficult to work out. While formal languages such as programming languages are usually designed to avoid ambiguities in order to ensure that the computer consistent decisions.

1. **How have natural language processing been affected by the availability of the World Wide Web?**

Before, a natural language process would normally be to answer a query from a repository which the system then had to attain deep understanding of that query on its own which is difficult. With the World Wide Web as its repository, the system only have to use a template of the answer because it is highly probable that it would match with something in it.

1. **Explain the acquaintance algorithm.**

The acquaintance algorithm is used to identify languages by using sets of n-letter words, also called n-grams, and with statistics that shows how likely the sets appear in certain languages determine which language a document is written in. For example given the tri-grams “and”, “ing” and “the”, it is highly probable that the language is English.