nature relative to the zeitgeist. Consider the sentence ‘*I saw her duck*’. One interpretation could be that I saw a duck (noun) that belonged to her. Another interpretation of this same sentence could be that I saw her perform the act of ducking (verb) to avoid an obstacle. Depending on the context, sentences with similar structures and subjects can still convey very different meanings. Take the following sentences: ‘*I ate rice with a spoon*’, ‘*I ate rice with curd*’and ‘*I ate rice with Rahul*’. All of these sentences convey that I am eating rice. In the first instance, I use a spoon as a utensil, while in the second sentence, I use curd as an accompaniment. The third sentence, however, suggests that I am eating rice in the presence of a person. Teaching computers to understand how to interpret these sentences differently and resolve ambiguities has been a significant motivation behind the development of NLP and understanding techniques.

**2.1 Computational Linguistics and Natural Language Processin**

Like any other medium of information, human languages have undergone various stages of development—they originated at some point in time, propagated far and wide, and have borne witness to the evolution of human society. While we do not know exactly when and how the earliest humans spoke, numerous theories have been proposed about the origin of language. Alister Hardy and Elaine Morgan proposed the *aquatic ape theory* in 1997, highlighting that there are certain traits we do not share with our primate relatives but that we do share communication traits with aquatic animals. For example, whales and dolphins are known to communicate with members of their species using sounds.

Linguistics is the discipline that engages in the scientific study of languages. It is an interdisciplinary system (see Figure 2.1) where linguistics and its branches, such as sociolinguistics, psycholinguistics, and neurolinguistics, seek to answer significant philosophical questions, such as: What rules do languages follow? How do languages evolve? How do we learn and process meanings in our minds? How are different modalities of languages related to each other? Linguists like Noam Chomsky and Steven Pinker hypothesise that language is something innate. From a very early age, we begin to mimic natural sounds and lip movements. We grow up associating these sounds with certain objects, qualities, and attributes of the environment around us. By studying these theories, we can infer what the earliest languages might have been like, but our understanding only extends so far back into the history of human languages (Jurafsky and Martin 2009; Pinker 2010; Vaneechoutte 2014).

**[Insert Figure 2.1]**

Figure 2.1: Language-related Disciplines (Tsujii 2021) – Linguistics, Cognitive Science, Psychology, Natural Language Processing (NLP), Artificial Intelligence (AI), and Computational Linguistics. All of these disciplines study language from different perspectives.

Many other fields, such as neuroscience and psychology, also show great interest in languages. Computational linguistics is a sub-field of both linguistics and computer science that focuses on the interactions between human language and computers, serving as a bridge between the broader field of linguistics and engineering processes. While computational linguistics is more concerned with understanding language structure and developing computational models, NLP emphasises the design and analysis of algorithms and systems for *tasks* that rely on processing human language input.

**2.2 Overview of the Natural Language Processing Pipeline**

The standard *pipeline* used in NLP involves several steps. The natural language input for the pipeline discussed in this book is typically modelled as a collection of machine-readable text documents, known as a *corpus;* a larger collection of these documents is referred to as *corpora.* When employing NLP systems, the standard pipeline consists of a sequence of steps, as illustrated in Figure 2.2, to address one or more NLP-based tasks.

**[Insert Figure]**

Figure 2.2: Stages of the language processing pipeline for textual data input.

**Tasks in Natural Language Processing (NLP).** Depending upon the task, the output from the NLP pipeline could be in the form of a sentence-level or word-level class label, a sequence of words, a piece of text, and even paths of a graph node-to-edge sequence (see Figure 2.3). To better illustrate the tasks, we will consider an example: ‘*I do not support WHO. They underfund Indian diseases’.* Below is a non-exhaustive list of popular tasks in NLP.

**Sentiment Analysis:** Detecting the type and intensity of emotional tone or opinion expressed in some text. Here, the output fr the entire sentence is a label such as positive, negative or neutral. Based on the usage of