UiT The Arctic University of Norway

Department of Computer Science

A Robot System for Personalized Language Education

Implementation and evaluation of a language education system built on a robot

Yigit Can Dundar

Master's thesis in Computer Science…INF-3990-1…June 2020

Abstract

As modern computer hardware becomes more affordable, the application areas of Artificial Intelligence (AI) expand and become more feasible to implement efficiently. This aspect opens up opportunities where AI robots can be used to support and improve many areas of human lives. One such area is education, and this thesis project combines AI and education to evaluate whether the results from that union can provide effective education. As a solution, a personalized language education robot that teaches the basics of the Italian language was implemented. The robot collects the user's movie preferences using verbal communication and uses that information during the study sessions to personalize the learning experience. After implementation, the system has been tested on actual students to evaluate its educational effectiveness. The students filled out questionnaires that asked them whether they thought the system was effective in teaching a language, whether they thought personalization was a motivating factor or not, and they also provided their feedback as to how to improve the system in the future. The results showed that the personalized language education robot was, for the most part, effective in teaching the basics of the Italian language. However, the feedback from the users highlighted some key improvements that may be necessary if such a system were to be utilized to cover a broader subject.

Keywords

Personalized Education, Artificial Intelligence, Facial Recognition, Language, Education Effectiveness, Natural Language Processing, Robotics

Table of Contents

List of Tables

List of Figures

1 Introduction

As modern computer hardware becomes more powerful and accessible, it opens up opportunities for computer systems that were otherwise non-feasible or expensive. This situation has allowed complex applications of Artificial Intelligence (AI) systems to begin taking shape. AI can be described as computer systems that have been designed to interact with the world through capabilities and intelligent behaviors that resemble humans [1]. AI can assess the available information and then take the most sensible action to achieve a stated goal [1]. The common goals for AI can be to provide expert-level suggestions, make predictions on a subject, be a part of a video game, organize day-to-day tasks, match patterns, and classify objects. AI has advantages over humans when it comes to performing many calculations in a quick manner, transforming data, searching through data to look up specific information, and more [28]. One area that AI has been utilized is education. In fact, from the earliest days of computers, researchers have strived to develop computer tutors that are as effective as human tutors [25]. And a review of the field of AI in education showed that the research papers published in the past two decades highlight an impressive process of growth, maturation and evolution within the AI in education field [26].

1.1 Background and Motivation

The application of AI in education has been the subject of academic research for more than 30 years [1]. From applications on smartphones and tablets to robots and desktop computers, AI in education can take many forms. Robots, in general, are materialized views of AI systems that have a specific purpose or a role. Within the scope of education, robots can take the role of a tool, peer, or tutor in a learning activity [3]. In order to teach a subject to a user, robots need to have social capabilities where they can communicate with the user effectively and naturally. They can achieve this social capability by utilizing text-to-speech, multiple language support, physical gestures, and so on.

When considering research in education of any kind, an important topic that comes up is referred to as educational effectiveness. Educational effectiveness is one of the measurable facets of the larger concept of education quality [34]. So, if the education quality is higher, the better the students can learn a subject.

There have been many instances of robots (and AI) being utilized for educational purposes in the past. In one case, a robot taught English vocabulary, story-making, and story understanding to children that improved their language skills [11]. There are some features that the robot is suggested to have to improve its effectiveness. For instance, in another study, the results of an 18-day field trial, where a robot was used to create relationships with children in a mutually informative way, showed that the robot needed to have more things in common with its users [14]. Another study further solidifies the need for this aspect [12]. In that study, an adaptive approach to language tutoring was proposed and was tested, which showed that the adaptive system was successful in the short term with suggestions on how it could be effective in the long term as well [12]. To sum up the findings from the papers [11], [12] and [14], the idea of using a language tutoring robot is more feasible as long as certain criteria are met. These criteria can be outlined as:

- The robot needs to motivate and engage its users
- The robot needs to have some things in common with its users
    - General knowledge of hobbies, likes, preferences, etc.
- The robot needs to adapt to its users' learning capabilities and shape its education methods accordingly

Following these criteria, personalization techniques can potentially be used to improve the effectiveness of the robot. Personalization, in this case, refers to the idea of approaching the education experience with the users' preferences and needs in mind. Thus, providing them with a more unique and personal experience.

Figure 1 - NAO Robot

To meet the outlined criteria and provide a personalized education experience, the humanoid robot, as seen in Figure 1, created by SoftBank Robotics called NAO [2], is used. NAO has features like facial recognition, text-to-speech, motor functionality, voice commands, and more. These features open up opportunities where new systems can be implemented to run on top of already existing functionality. In this case, a language tutoring system will be implemented to run on the robot, which will incorporate personalization aspects based on the user's movie-related preferences to teach the basics of the Italian language. The robot learns its current user's name and movie-related preferences and then proceeds to teach them the basics of the Italian language by utilizing the gathered user information.

The system is then tested on actual students to evaluate its educational effectiveness and gather general feedback on how to improve the system. During the tests, both the robot's and the users' behavior is observed and are noted down. The results from the tests are then presented and discussed.

The motivation for this thesis is to explore if the previously mentioned criteria and technologies can work together and be an effective method of education when used for teaching a language.

1.2 Problem

The current education system, which is widely used around the globe, has been around since the Renaissance, where most of the teaching is done by an instructor lecturing to a room full of students while only some of them are paying attention [23]. Although time-efficient, this system tends to leave some students behind in their current studies. Because, as with every other aspect of human lives, people have different ways they can learn a subject as well. Some people learn best at their own pace of study; some are more comfortable learning in a classroom setting; some of them are more motivated for a specific subject than others, and so on. This variation in education preferences brings up a question: With the help of robotics technology and AI, what can be done to provide a more personal and effective language education?

1.3 Purpose

This thesis describes and evaluates an idea of utilizing AI in education to see whether it can be an effective education method for both current and future generations. Also, the design and implementation details of the language education robot system alongside its effectiveness evaluation are presented to expand the available knowledge in AI in education for the research community.

1.4 Goal

The goal of this degree project is to provide a working language education robot system and explore whether or not using this system can be an effective tool when teaching a language to a person. After the education robot implementation is finished, the system will be tested on actual human subjects. The results from those test sessions will be presented alongside a discussion of the results. The result of the entire project will be an insight as to whether the idea of a personalized language education robot proves to be effective or not for both the research community and possibly future generations.

1.4.1 Benefits, Ethics and Sustainability

Growing number of students per classroom and the demand for greater personalization of curricula for children with diverse needs are fueling research into technology-based support that augments the efforts of parents and teachers [27]. This thesis provides a potential solution to these demands using modern robotics and AI technologies to create a personalized education experience for users with different educational needs and preferences. The system can be used to supplement the current education system or be used as a standalone personalized education method. Thus, potentially cutting costs in education and being beneficial for students, teachers, and schools.

From an ethical standpoint, since the robot requires private and personal information from the users, that information is secured and not given to or shown to third parties. Also, the robot is not in any way, shape, or form trying to replace teachers from classrooms but rather assist and supplement the education by providing a more personal connection with the students.

In order for this system and similar systems to be sustainable, further research in the area of AI in education needs to be conducted. Also, further technological advancements need to be made to make the underlying hardware more accessible, powerful and affordable.

1.5 Methodology / Methods

In terms of research methodologies that can be used for this thesis project, there are two main research methods: quantitative and qualitative. Quantitative research methodologies utilize experiments and large data sets to reach a conclusion, whereas Qualitative research methodologies use investigations in an interpretative manner to create theories or artifacts by, commonly, using small data sets [6]. Qualitative research methodologies have been utilized during this thesis project to prove whether the idea of using a personalized language education robot was effective or not. User insight and interpretation were required to evaluate the effectiveness of the system. Also, gathering detailed and subjective information on the system was deemed more useful than gathering large sets of objective data.

Alongside the two methodologies, two respective approaches are available. These approaches are called inductive and deductive. While the inductive approach establishes a general proposition from facts, the deductive approach derives conclusions from known premises [6]. The inductive approach is more suited for qualitative research, and the deductive approach fits the needs of quantitative research. Since, in this thesis project, qualitative methodologies are utilized, and the ideas on this project are more in line with inductive reasoning, the inductive approach was the obvious choice.

1.6 Contributions

The contribution of the thesis is to provide a working personalized language education robot system to teach a language using modern computer hardware and AI technologies. In this setting, the robot takes the role of an instructor and teaches the basics of the Italian language to its students. This system can be used to supplement current educational methods or as a standalone method to learn a language. In terms of Computer Science, the system combines Natural Language Processing, Facial Recognition, and Speech Recognition with robotics technology to create a unified example of such systems and components working together. In terms of education, the system provides an alternate teaching method by using a robot that handles the responsibilities of an instructor, which directly interacts with a human being with little to no supervision. To sum it all up, the system combines already existing robotics technology with Natural Language Processing, Facial Recognition, Speech Recognition to provide an alternate, personalized education method. The thesis itself provides the necessary knowledge, challenges, and details into how all of the previously mentioned areas and aspects of the system can be used together for future research and software development.

1.7 Delimitations

The initial degree project proposal highlighted the inclusion of broader personalization techniques during the education part of the system. However, this feature was deemed too broad and too demanding by the supervisors and was therefore toned down to be just about a specific personalization aspect: movies. This delimitation allowed more focus on the writing parts of the thesis while also taking the much-needed workload off of the implementation part.

Before the project proposal, the system was supposed to teach multiple languages from which the users could select whichever language they wanted to learn. This feature would require too much additional work and would not impact the primary goal of the project in a significant way. Therefore, multiple language feature was also cut down from the rest of the project, and instead, the system only teaches Italian to its users.

1.8 Outline

The chapters in this thesis are as follows:

- **Chapter 2** contains information on robotics technologies and specific related work
- **Chapter 3** provides a description of Natural Language Processing and how it factors into the implementation parts of the project
- **Chapter 4** gives the background information related to language education and personalization alongside related work
- **Chapter 5** goes over the methods and methodologies that were utilized during the project
- **Chapter 6** covers the requirements and the design aspects of the system

- **Chapter 7** provides implementation details of the system
- **Chapter 8** contains the test results and evaluation of the system
- **Chapter 9** concludes this thesis by providing the summary, discussion and future work.

2 Robotics Technology

This chapter will present information on social robots, robot programming, Human-Robot Interaction (HRI), facial recognition, and related work.

A robot is often viewed as a device that performs tasks on command [7]. They can be used in various environments and can vary in terms of the complexity of which tasks they can complete. Some robots are more primitive and are used for more specific tasks (like moving an object from point A to point B). Meanwhile, some robots can handle more complex tasks, and they can observe and interact with their surroundings, communicate with humans, offer insight on a subject, and so much more. Robots are still very limited as to what they can achieve on their own. They often require additional input from humans or from their environment to complete their tasks. This lack of true independence can impede their usefulness on certain topics, but that does not mean that they cannot be improved.

2.1 Social Robots

A certain type of robots, called social robots, have social capabilities that they can utilize to communicate with humans or other robots to complete their tasks. These types of robots can be used in education, therapy, entertainment, and businesses. Based on past research, observations and findings indicate that, in order for social robots to be accepted as social entities, they need to possess certain capabilities. They need to be able to express and perceive emotions, communicate with high-level dialogue, learn and recognize models of other agents [8]. On top of those capabilities, they also need to be capable of establishing and maintaining social relationships while exhibiting distinctive personality and character [8]. These capabilities are not easy to achieve. Both hardware and software limitations can pose challenges when it comes to creating a robot with these capabilities. However, with the help of modern hardware and software technologies, SoftBank Robotics created a robot called NAO robot [2] that provides a working example of a socially capable robot. NAO robot can observe its surroundings, recognize faces, express certain emotions with gestures, move around on its own using its legs and sensors, speak in multiple languages, and much more. On top of all of this, the NAO robot provides an open platform in which custom applications and behavior for the robot can be implemented using pre-existing features.

2.2 Programming the Robot

The framework used in the NAO robot is called NAOqi Framework. This framework is the programming framework used to write custom software for NAO [24]. It allows homogeneous communication between different modules, homogeneous programming, and homogeneous information sharing [24]. The framework contains many modules that provide the featured functionality of the robot. Some example modules from NAOqi framework are:

- **ALProxy Module:** Is a client to the served methods of a module [40]
- **ALMemory Module:** Provides a centralized memory that can be used to store and retrieve named values [41]
- **ALTextToSpeech Module:** Allows the robot to speak [42]
- **ALSpeechRecognition Module:** Allows the robot to recognize predefined words or phrases in several languages [43]

The modules can be used in custom code to utilize the built-in functionalities and hardware components of the robot. To write custom software and to access these modules, the NAOqi Framework supports C++ and Python programming languages. C++ is a portable, compiled, open ISO-standardized programming language that can be used to write programs for computers to understand and execute [38]. And Python is a clear and powerful object-oriented programming language [39]. A development environment, created by Softbank Robotics, called Choregraphe [22] can be used to write custom software to run on top of the framework. Choregraphe allows developers to create custom behavior, animations, and scripts to provide additional functionality to the robot. Also, using Choregraphe, it is possible to utilize pre-made modules and functions to create even more complex structures.

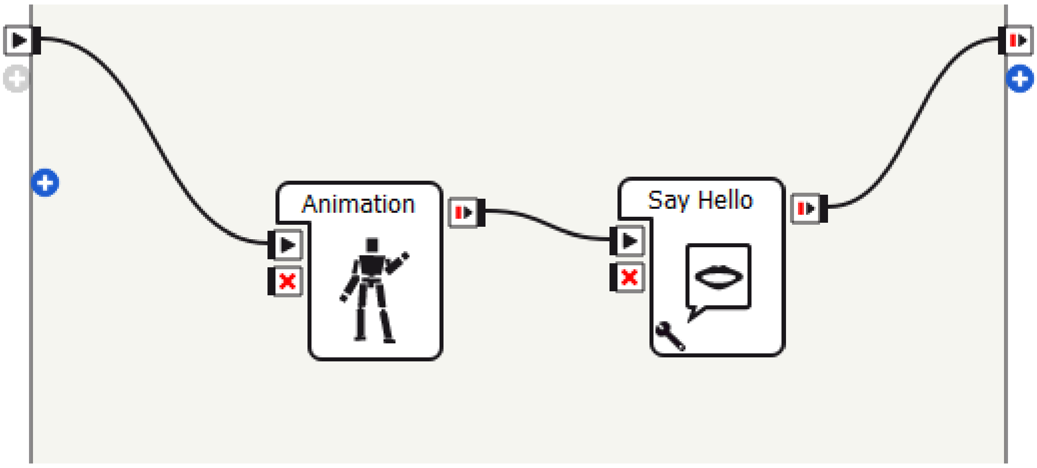

Figure 2 - Choregraphe Simple Behavior Example

In Figure 2, a simple behavior created in Choregraphe is presented. If the above behavior example is executed on the robot, it will do a simple animation with its arms (Animation box) then say “Hello” (Say Hello box) and finish the execution. Each box, Animation and Say Hello, contains a script that has code written in Python, and the boxes were created by Choregraphe’s developers. Also, it is possible to create a custom, empty box to write Python code. This aspect allows developers to both create completely new features on the robot itself and utilize pre-built functionality provided by Choregraphe’s developers as well.

2.3 Human-Robot Interaction

One key aspect that needs to be considered when working with robots that interact in some way with humans is called Human-Robot Interaction (HRI). HRI can be described as the study of the interaction between robots and humans. And, in terms of research fields, HRI is at the intersection of psychology, cognitive science, social sciences, artificial intelligence, computer science, robotics, engineering, and human-computer interaction [9]. Understanding HRI is crucial for implementing effective and healthy robot-based systems that interact with humans. For instance, one study suggests that robots need to be relatable in predictable ways to humans while also encouraging a natural bond with them [10]. This aspect becomes more important when the HRI system relies heavily on human-robot socialization. In fact, another study proposes several metrics to evaluate the social effectiveness of the interaction [29]:

1. **Interaction Characteristics:** Observing or analyzing the interaction characteristics like interaction style or social context.
2. **Persuasiveness:** To what extent the robot alters the behavior, feelings, or attitudes of humans.
3. **Trust:** The amount of trust that the robot is able to achieve with the human.
4. **Engagement:** The efficacy of social characteristics that capture attention and hold interest (personality, emotion, dialogue).
5. **Compliance:** Characteristics that influence the cooperation between the robot and the human (appearance, adherence to norms).

Alongside the metrics outlined above, communication is another factor in HRI system effectiveness. The communication delay, jitter, and bandwidth can have profound effects on human performance [29]. Therefore, ensuring smooth and responsive communication can improve the quality of the system and improve the performance of the human interacting with the robot.

The human's role may also affect the fluidity and effectiveness of HRI. Based on a study, humans can have five roles in an HRI system [30]:

1. **Supervisor:** Humans can monitor and control the overall situation.
2. **Operator:** Humans can modify the internal software or models when the robot behavior is not acceptable
3. **Mechanic:** Humans can physically intervene the robot, so the interaction has the desired effect on behavior
4. **Peer:** Humans can be a peer who can give the robot commands to work together with them.
5. **Bystander:** Humans can act as a bystander who can have some control over the robot's capabilities.

Each role requires different information and awareness [29]. Therefore, the role a human can take in an HRI system can affect the system's performance and effectiveness.

2.4 Facial Recognition

Biometrics are any automatically measurable, robust, and distinctive physical characteristics or personal traits that can be used to identify an individual or verify the claimed identity of an individual [35]. Humans have many unique defining characteristics that are inherent to them, such as their eyes, their fingerprints, their voice, their face, and so on. Facial recognition is a biometrics technology that records the spatial geometry of distinguishing features of the face [35]. For instance, once the facial information is recorded, it can then be used to identify the user of a system and give them the necessary access to certain functionality.

When it comes to detecting faces, different methods and techniques can be utilized. The method to be used may vary depending on the limitations of the hardware or the capabilities of the software. There are two traditional ways of recognizing faces. These are geometric (feature-based) matching and template matching [55]. The main idea of geometric or feature-based matching is to extract relative position and other parameters of distinctive features such as eyes, mouth, nose, and chin [55]. Geometric matching can recognize faces even if the details of the individual features are not resolved [55]. This aspect can allow this method to work in non-ideal environments and situations. In the other traditional method, template matching, the image that is represented as a bidimensional array of intensity values, is compared using a suitable metric with a single template representing the whole face [55]. There are, however, more modern approaches to recognizing faces as well. Holistic matching, for instance, works by taking the complete face region into account as input data into the face catching system [56]. In addition, hybrid methods use a combination of both holistic and feature extraction methods, which usually use 3D images of the person's face [56].

The NAO robot, mentioned earlier, provides the needed modules for facial recognition functionality using its camera and the underlying computer hardware.

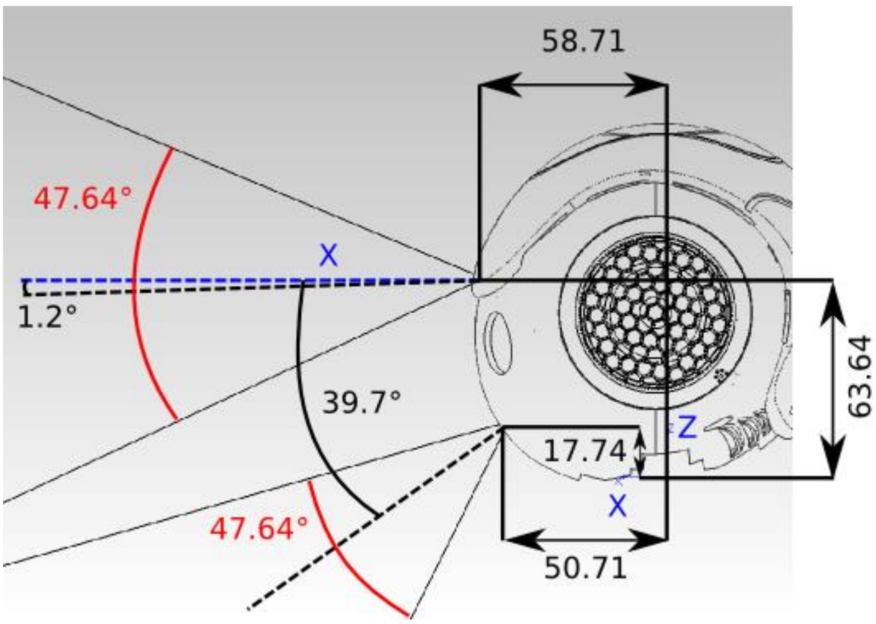

Figure 3 - NAO Robot Video Cameras [57]

As shown in Figure 3 [57], the NAO robot has two cameras located on its face. These cameras are capable of capturing images and video while also being used for identifying objects in the visual field [57]. The robot can recognize human faces using its cameras and built-in facial recognition module. The NAO robot uses a vision module called ALFaceDetection to detect, and optionally recognize faces in front of him [58]. In order for the robot to recognize people, the face shown to the robot needs to have a neutral face (no emotions shown) for detection, and then the robot needs to learn the detected face to be able to recognize the person [58].

2.5 Related Work

The studies mentioned in this section all provide a working example of robots being utilized for education, alongside some findings from tests that have been conducted on actual human subjects.

Comparative Study of Effects of Language Instruction Program Using Intelligence Robot and Multimedia on Linguistic Ability of Young Children [11]

This paper evaluates the feasibility of using intellectual robots as language instruction tools for children. Using picture books as content, the authors have developed a reading program, which had three main stages. In the first stage, called the basic stage, the robot read the entire text from a picture book to enable the children to understand the story and the structure of the book. In the next stage, the practice stage, the children were tasked with practicing their reading and reinforcing their understanding of the study. In the third stage, called the active stage, the children were allowed to have varying linguistic experiences such as story making and understanding. The robot used its LCD display, moved its arms and legs, and expressed emotions using lamps located around its face. They conducted several tests on children at the age of four to test their story making, story understanding, vocabulary, and word recognition abilities. The test results showed that a robot using bi-directional interaction improved the children's linguistic abilities.

The thesis project, much like the paper described above, provides a system which utilizes a robot to teach a language. However, the main difference between both systems is that the system provided by the thesis project utilizes personalization, actual speech, and conversations instead of utilizing a display and gestures to conduct the teaching, which in turn could lead to different results.

Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making [12]

This paper presents an approach to adaptive language tutoring in child-robot interaction. The authors used a modified and extended version of the Bayesian Knowledge Tracing model as an approach to adaptive language tutoring. More specifically, they use the model to decide which tutoring action to take by the robot based on the learner's knowledge and skills. The authors implemented the model as an application in a child-robot second language tutoring game. The robot was also accompanied by a tablet PC in order to conduct initial tests. The first test was a vocabulary-oriented game that was explained to the children, followed by a test run to make sure that the child was able to understand how to play the game properly. And then, the actual interactive language game was played between the robot and the child subject. During testing, the robot gives positive feedback by praising the child, using implicit corrections, and using confirmation gestures like smiling and nodding to ensure that the child stays motivated and learns the words properly without referring to explicit corrections. After the initial tests, an evaluation study was conducted to assess the effects of the adaptive model on a larger scale. A total of 40 participants took the tests. And, the analysis of the results indicated that the adaptive model was successful in teaching the words during human-robot interaction. However, post-test results showed that there was no significant improvement over the control group. They explain this inconsistency as either a result of the way as to how they conducted the experiments during the post-test phase or as a result of strong inter-individual differences among learners. They conclude the findings by saying that the results, in parts, were very promising and that the future work could focus on refining the model to be effective post-test as well.

The system developed as part of the thesis project provides a language education system that makes use of personalization. The paper above provides a working sample of a language education system that can adapt to its users' needs. This is similar to the thesis project, where the robot used during the teaching sessions tries to learn the users' preferences to improve education. The difference between the two projects is that the thesis project gathers preferences on a specific topic (movies), where the above paper focuses more on adapting to its users' skills.

Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning [13]

Contrary to conventional teaching agents, where the robot takes the role of a teacher or a caregiver, the authors of this paper propose the opposite scenario in which the robot receives instructions or care from children. The authors hypothesize that by using this care-receiving robot, they might construct a new educational framework where learning by teaching is promoted. The authors outline the main use-case scenario of the care-receiving robot as a tool used for learning support or learning reinforcement for children. They claim that using a care-receiving robot can strongly motivate children to take care of the robot. Which, in turn, can motivate the children to complete the topics for learning. The authors chose a venue of an

English language school for Japanese children where it was possible to conduct experiments with the children in a classroom setting, who were aged between 3 to 6 to test their claim. First, pilot trials were conducted. During the trials, the robot was introduced to the subject children with the purpose of observing whether it could promote the care-taking behaviors of children by trying to learn basic vocabulary from children. One observation indicates that the children were more inclined to take care of the robot when the robot gave incorrect answers to the teaching questions. After the idea was deemed feasible, the authors moved over to the main test phase. In this phase, the robot was used to investigate whether it could promote learning by teaching and also categorize the forms or types of the children's teachings. The results showed that the children were successfully motivated to use the robots, and they also learned more effectively compared to not using the robots. However, it is mentioned that the children spent more time with the robots than without. Therefore, the results might not be fully reflective of a real-life, human teacher to robot teacher, comparison.

The robot system developed on the thesis project teaches its users a language using personalization to motivate them. The above paper explores the idea of using robots to motivate children to learn a language. The method they utilize (children teach the robot instead of the other way around) differs from the method used in the degree project where the robot handles the teaching part. The similarity between the two systems is that they both focus on improving language education by also trying to motivate the users to interact with the robot.

Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial [14]

In this paper, the proposition of whether robots could form relationships with children in a way that the children might learn from the robots as the robots learn from them is explored. In order to study this idea, the authors conducted an 18-day field trial held at a Japanese elementary school. They used two robots that interacted with the first and sixth-grade pupils near the classrooms. The robots were equipped with several sensors, which helped them identify and interact with the children. During their interactions, robots spoke English with the children with the help of gestures. The first week of the trials resulted in children interacting with the robots frequently, whereas the children's interest in the robot declined drastically during the second week of testing. Overall, the robot was able to encourage some children to improve their English skills, more so if they already had some prior English language knowledge. The authors think that, based on the results, interactive robots should have more things in common with their users to be more effective, which is a challenge in terms of both social and technical aspects.

The thesis project's goal is to determine whether using an educational robot is an effective and motivational tool for education. The paper outlined above explores the motivational effect that two, language education-oriented, robots have on children. The paper is similar to the thesis project in a way that both studies explore the feasibility of using robots to teach a language by motivating its users. The difference between the two is that the paper mentioned above uses robots to interact with a lot of different users, most commonly at the same time, whereas the robot in the thesis project interacts with one user at a time. Limiting the concurrent users to just one can allow the robot to provide a more personalized approach. Which the authors of the paper mention as a potential solution to the “lack of motivation” problem.

3 Natural Language Processing

In this chapter, Natural Language Processing (NLP) will be described alongside key aspects of NLP. The chapter will also touch upon speech recognition and how it ties to NLP, and some related work will be presented.

The way humans interact with computers usually involves using external tools for input like a keyboard and a mouse. Using those tools could be considered as being natural to computers, but what about something that is natural for humans, like speech? The answer to that question is Natural Language Processing (NLP). NLP is an area of research and application that explores how computers can be used to understand and manipulate natural language text or speech to do useful things [17]. Using NLP, it is possible for a computer to both analyze given sentences and to create new sentences from scratch. Therefore, NLP allows computers to understand human speech (or text written in natural language) and to give back responses to them in a natural manner as well.

3.1 Speech Recognition

In speech recognition, sounds uttered by a speaker are converted to a sequence of words recognized by a listener [36]. Speech recognition allows a computer-based system to acquire spoken, natural, input from a user and then turn it into something it can understand and work with. This method of acquiring input is useful when other forms of input methods (keyboard, mouse, etc.) are not usable in a use case scenario. The way speech recognition works depends on the underlying model used during processing. A study outlines the commonly used models as [54]:

- **State machines:** Consists of states, transition among states, and an input representation.
- **Rule systems:** One of the main tools used when dealing with knowledge of phonology, morphology, and syntax.
- **Logic:** Used for modeling semantics and pragmatics.
- **Probabilistic models:** Used for capturing every kind of linguistic knowledge.
- **Vector-space models:** Used in information retrieval and many treatments of word meanings.

Most tasks in speech and language processing can be viewed as resolving ambiguity [54]. An input is said to be ambiguous if multiple, alternate linguistic structures can be built for it [54]. This ambiguity stems from the way humans speak and can be hard to resolve without context. For example, a sentence like “The cat is gone.” can be interpreted as the cat is outside the house or as the cat is dead. The models mentioned before can help resolve these ambiguities. Thus, allowing the computer to understand the meaning of a given human speech in a better way.

Within the scope of NLP, speech recognition works as a gateway to analyzing recognized spoken language. With speech recognition, it is possible to gather utterances from users to be parsed and analyzed for semantics or named entity extractions, and so on. This interrelation does introduce some dependencies between NLP and speech recognition. For instance, if the speech recognition module is not able to accurately recognize speech, the output text from the module will be inaccurate as well, which, in turn, reduces the accuracy of NLP parsing. Therefore, it is important to find the balance, in terms of accuracy, between speech recognition and NLP to understand human speech properly.

3.2 Parsing and Generation

The two main activities in NLP can be categorized as parsing and generation. Parsing is the process of analyzing a sentence to determine its syntactic structure according to a formal grammar [4]. Parsing is not directly concerned with giving meaning to a given sentence, but it enables it. After parsing is done, it is possible to analyze the determined structure to find out what the sentence actually meant. This process of finding the meaning of the sentence is referred to as semantic analysis [4]. This process is rather complicated, and it depends on a number of external aspects like context, common sense reasoning, results of the parsing, and more [4]. With the co-operation of parsing and semantic analysis, it is possible for a computer to analyze and give meaning to an input sentence.

Generation is the process of creating a structurally sound and meaningful sentence within some context. Usually, the way humans decide on what to say depends on the current space, time, and situational constraints [5]. Generation, within the scope of NLP, works in a somewhat similar way. The overall task of generating a sentence covers a broad spectrum ranging from planning some action to then executing it [5]. This range of actions contains many small decisions in between that alter the way the final output is determined. So, the generation task can be characterized in terms of mapping information from some non-linguistic source (like a knowledge base) into some corresponding linguistic form [5].

3.3 Part-of-Speech Tagging

Part-of-speech (POS) tagging aims at labeling each word with a unique tag that indicates its syntactic role [16]. Words in a sentence can take syntactic roles such as nouns, verbs, adverbs, and so on. Using POS, it is possible to find out the structure of a sentence or check if the given sentence conforms to the syntactic rules.

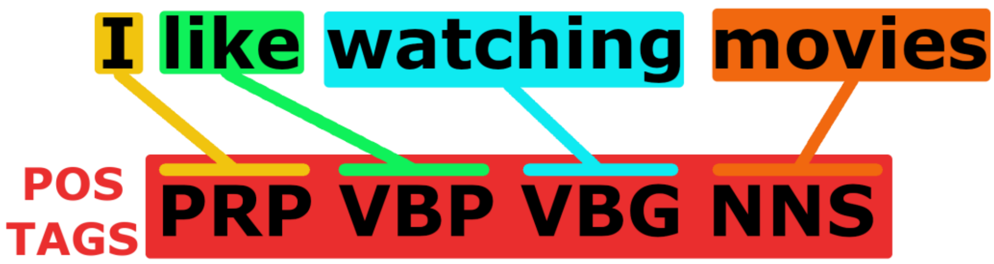

Figure 4 – POS Tagging

The results of an example tagging scenario can be seen in Figure 4. In this figure, the sentence "I like watching movies" has been POS tagged, and each of its elements have been given a tag to represent their role in the sentence. The element "I" has been tagged with "PRP" which stands for personal pronoun, the word "like" has been tagged as "VBP" which stands for verb present, the element "watching" has been tagged as "VBG" which stands for verb gerund and the element "movies" has been tagged as "NNS" which stands for noun plural. With the help of these tags, it becomes possible to analyze the sentence further. POS tagging does not give any information as to what exactly the sentence meant. But, by understanding the structure of a sentence, it is easier to find out the meaning of it.

3.4 Named Entity Recognition and Extraction

A Named Entity (NE) can be seen as a first-level generic semantic information that can be found on documents like text, audio, video [32]. Named entities in documents can be the name of a person, a title for a movie, a unique name for an object, and more. Recognizing and extracting such entities can help understand the meaning of the sentence, and the extracted entity can then be used to serve another purpose. Named Entity Recognition (NER) labels automatic elements in the sentence into categories such as "PERSON" or "LOCATION" [16]. The categorized elements can then be extracted from the sentence for other use cases. One use case could be to gather information about a person's preferences on a subject to provide personalized services like advertisements, recommendations, or education.

The process of recognizing and extracting named entities is not quite straightforward. Different sentence structures, complex named entities, mixed-language entities, and other such factors, make recognizing and extracting named entities more difficult. Even if such difficulties are accounted for, recognition and extraction tend to be inaccurate at times. In order to make the NER system more accurate, techniques like POS tagging, machine learning, or chunking can be utilized. In the past, such techniques have been utilized to provide accurate NER results [33].

3.5 Related Work

Presented in this section are some related work that combines NLP and robotics to serve specific purposes.

Using Natural Language Processing (NLP) for Designing Socially Intelligent Robots [15]

In this paper, the possibility of using NLP as a tool to enable social robots to create dialogues that build trust and comfort between the users is explored. The robots learn from interpersonal interactions with users, learn the user's personal information and preferences to build a user profile, which is then used in the dialogue. The author has designed a dialogue system, using NLP, that can collect basic personal data (age, name, likes, dislikes, and so on) and either create a new user profile or update an already existing one. In this scenario, the robot also uses facial recognition, aided by artificial neural networks, to help aid profile creation and detection. The robot uses personal information to customize future dialogues in order to foster long-term interaction by building trust and comfort. The author tested the system by implementing the designed system on the NAO robot and then testing the robot's ability to extract personal information and use it to customize the dialogue on ten college students. The test is conducted for two weeks. In the first week, the robot is introduced to the users, and it starts gathering information about them. In the second week, the robot's abilities to retrieve and utilize the personal information are tested where the users are asked to rank the relevancy of the information and how much do they feel the robot knew them. And as a result, 63.5% of students were satisfied with the system performance. As future work, the author claims that gathering deeper knowledge about the users (like family or friendship tree) can help provide a better understanding of the users and potentially improve the user satisfaction rating.

NLP, in the thesis project, is used in certain parts of the system to analyze and extract useful information from user inputs for future use. The paper above combines NLP with intelligent robots to create a system that can gather personal information of users, utilizing natural speech. In that regard, the thesis project is similar to the system mentioned in the paper. The difference is that the paper above is only focusing on gathering accurate information and not utilizing that information to serve a different purpose, the thesis project, however, utilizes the gathered personal information to aid in teaching a language.

Spoken Language Processing in a Conversational System for Child-Robot Interaction [31]

In this paper, a conversational system for child-robot interaction is presented. The purpose of the paper is to highlight some practical issues concerning spoken language processing that can occur from the usage of a robot. The authors of the paper have developed an event-based integration approach using the NAO robot using a middleware. The system they developed has three game-like activities that the children can interact with. The quiz activity allows the robot and the child to ask each other multiple choice quiz questions from different domains. The imitation activity allows either the child or the robot to present a sequence of simple arm poses that the other tries to memorize and imitate, and the dance activity allows the robot to teach the child a series of dance moves. The robot utilizes automatic speech recognition combined with natural language understanding to recognize and analyze child speech. The authors mention that child speech is different from adult speech and that more robustness against recognition errors needs to be in place to understand child speech properly. The robot can also generate sentences, responses, and general feedback using natural language generation. The authors have applied a yearly cycle of specification-development experiments. In these experiments, every subject was invited to play with the robot three times with very little supervision. The initial evaluation has yielded encouraging results, and the authors think that more detailed experiments need to be conducted to analyze the system further.

The NLP implementation in the thesis project focuses on parsing to extract user preferences for use in the study sessions. Whereas, the paper summarized above combines both parsing and generation to provide three simple game-like activities for children. The main difference between the two projects is that the paper outlined above gives more complex feedback to its users using NLP generation techniques, which the thesis project does not utilize. The similarities, on the other hand, lie with the parsing part of the communication where both systems utilize speech recognition and NLP techniques to understand and analyze human speech.

4 Education and Personalization

This chapter will highlight the importance of education in general, provide insight into the importance and key aspects of both language education and personalized education alongside some relevant work within those areas.

Education is an important part of human lives. In past studies, education has been observed as an important factor in economic growth in African states [44], the level of education has been suggested as being an important factor in the employment of head nurses by highlighting the value of research methods and supportive leadership [45] and, education levels have been determined as a key factor in improving the rates of organ donation [46]. Furthermore, getting an education can open up job opportunities by making the person more qualified for a position. And, there are many education fields to choose from, for instance, a person can choose to study science, literature, language and more. A person can choose to focus on a specific field or choose to study several at a time. It all depends on what that person is trying to achieve once they are educated.

4.1 Language Education

Language education is a field of education in which a student can learn to speak, understand, and write in a specific language. An incentive to learn a language can arise from the reason that, as the world communities develop business and political relationships, there is a greater need for individuals to develop multilingual competence [47]. In general, learning another language can allow humans to communicate with others who share that same language. When it comes to learning, a collective term called learning styles is used to describe the ways in which an individual characteristically acquires, retains, and retrieves information [48]. To match the learning styles in education, there exist teaching styles that correspond to them. In terms of language education, a study suggests several teaching styles to utilize during language education lessons which are [48]:

- Motivating learning
- Balancing concrete information and conceptual information
- Balance formal training with more open-ended, unstructured activities (cultural contexts for the target language)
- Making liberal use of visuals
- Assigning repetitive drill exercises
- Balancing inductive and deductive presentation of course material

Based on these suggestions, if the instructors pick some suggestions that look feasible and utilize them during lessons, the most suitable teaching style will naturally evolve with a potentially dramatic effect on the quality of learning [48]. Covering the needs of the students, in terms of learning styles, can potentially lead to a more memorable and effective education.

4.2 Personalized Education

Students learn in different ways. For instance, students can learn by seeing and hearing, reflecting and acting, reasoning logically and intuitively, memorizing, and visualizing [48]. In response to these individual needs of students, personalization in education not only facilitates students to learn better by using different strategies to create various learning experiences but also caters to teacher's teaching needs in preparing/designing varied teaching/instructional packages [49]. In a classroom setting where a diverse group of students exists, the needs and wants of every student in terms of learning styles can be covered with a more personal and individualistic approach to teaching. In the past, there have been several personalized approaches being applied in education. A study highlights several attempts to deliver personalized education where universities and their faculty engage in a variety of activities that include developing relationships with students both in and out of the classroom, reducing class sizes, implementing various types of educational plans, and facilitating collaborative learning experiences in the classroom [50]. While there are possibly many answers to how personalization in education could be applied, the previously stated approaches seem to be the more common responses to utilizing personalization in education. The main takeaway with the idea of personalization is to adapt to the learners' needs. One such way of adapting could be to utilize personal preferences on a specific topic of interest for that learner. Perhaps, in this manner, the learner could feel more motivated to study, and the subject to be learned could be more effectively understood.

4.3 Related Work

In this section, related work in personalized language education will be provided and then compared to the thesis project by highlighting differences and similarities between them.

Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle [51]

The authors of the paper start by mentioning the importance of fluent international communication and how learning English is very popular in non-English speaking countries. They also highlight the importance of learning English vocabulary and how difficult it may be when learning English in general. They mention how forgetting learned vocabulary is a serious problem while learning English vocabulary. And, as a solution, they present a novel personalized mobile English vocabulary learning system that tailors English vocabulary learning materials to individual learners based on their vocabulary abilities and learning memory cycles. The system evaluates the learner's vocabulary and then recommends proper learning materials for that learner. The system, in general, allows the learners to use the least time to achieve the most efficient learning of vocabulary. The users can use the system whenever and wherever they want as long as they have a network connection on their mobile device. By using the system, the learners can learn new vocabulary suggested to them and also review their vocabulary learning progress to alter the way the system recommends new vocabulary to be studied. Once the system had been implemented, the authors conducted experiments in which a group of university students was invited to participate in using the system to learn English vocabulary. The participants were informed on how to use the system and were given five weeks to study English vocabulary using mobile devices in which the system was installed. After five weeks of using the system, the participants were invited to a post-test session to evaluate their English vocabulary abilities and to complete a questionnaire to assess their degree of satisfaction after learning. The results showed that the review strategy of the system was very helpful. And, the authors think that the results revealed that the personalized mobile English vocabulary learning system can significantly enhance the learner's English vocabulary abilities and promotes learning interest.

The personalization aspect within the thesis project focuses on movie-related preferences of the users to tailor their learning experience and motivate them to study. The difference is that the system presented in the related work focuses more on the learners' vocabulary learning abilities to adapt the course of education to their needs. The similarities between the two systems are that they both try to adapt to user needs by providing flexible and personal study sessions that do not impose strict limitations in terms of time and space of the education session.

Personalized Intelligent M-learning System for Supporting Effective English Learning [52]

The paper highlights a growing community of web users as wireless mobile devices and how the development of educational technologies also tend to be more mobilized, portable, and personalized. The authors of the paper state that mobile learning is an effective form of flexible learning, and learning English is very popular in non-English speaking countries. Based on their statements, they promote a system that takes the advantages of mobile learning by breaking the limitations when it comes to time and space of learning. In this personalized intelligent m-learning system, electronic English news articles are automatically retrieved from English news sites by an intelligent crawler agent, which are used as the course materials. The system recommends news articles to the learners based on their reading abilities which are assessed by the system. The system also takes the difficulty of the article into account when recommending it to the user. The authors have not provided actual experimental details or test results; however, they believe that the implemented system provides benefits in terms of providing a mobile and ubiquitous learning environment for English learning.

The system implemented in the thesis project utilizes a robot to carry out personalized education. Also, the personalization aspect focuses more on users' interests. The system implemented in the related work focuses on the learners' language capabilities when it comes to personalization and utilizes mobile devices to carry out education. Asides from those differences, the two systems are similar when it comes to their goals, where both systems utilize modern tools and devices to provide an effective personalized language education.

Personalized and Contextualized Language Learning: Choose when, where and what [53]

Initially, the author of the paper describes how the growth of mobile and ubiquitous learning technologies has opened up new avenues and learning areas for learners. The author highlights how language learners can continue their learning process outside the classroom whenever and wherever they want. Alongside this information, some challenges related to such mobile learning systems are mentioned. Challenges being: personalization and contextualization of the learning resources. As a response to the challenges, the author describes a prototype system called PALLAS, which is a personalized and context-sensitive foreign language system for real-life simulations using mobile devices. The system provides support to the learner by providing personalized and contextualized access to learning resources. The author also describes how learning resources need to reflect real-life situations since the learning process is done outside the classroom. To envisage the usage of the system, the author has created a scenario to identify the initial functional requirements of the system. In this scenario, the author describes a situation where a student sees a notification from the system, telling her that the art gallery has a French art exhibition. She visits the exhibition, which also allows her to get in contact with some French people that she could practice her French with. Whenever she has a problem understanding a French word, she queries the system's built-in dictionary. In order to provide the personalization aspect of the system, the system creates a user profile that contains information on their age, skill level, native language, interests, and taken courses. The system also incorporates some environmental factors as well, like location, time and day, and mobile devices used by the learner. The author explains how, due to resource restrictions, the system has not been able to be populated with language content. However, three language teachers were asked to evaluate the system from their perspectives, and the results from this evaluation show that the teachers thought the system was consistent with their teaching and learning philosophy and approaches. Two of the teachers also stated how the system took the language learning process outside the boundaries of a classroom and that it added content and reality to the course by making language learning a part of every-day life. The overall conclusion from the evaluation was that the system increased the flexibility of learning for students and that it was a suitable means of providing personalized learning.

The system implemented as part of the thesis project creates a user profile, fills it with movie-related preference information, which it then utilizes during teaching sessions to go over both every-day aspects of the Italian language and its grammar rules. In this manner, the system implemented in the thesis project and the aforementioned related work show similarities where both systems rely on the usage of user profiles and personal information to provide a personalized language education experience. The difference between the two systems is that the system described in the related work has not been tested on actual end-users where the thesis project has been tested with actual users.

5 Methods and Methodologies

In this chapter, research and engineering related methods and methodologies used during the thesis are described in detail alongside validation, verification, reliability, and replicability information.

5.1 Research Methods and Methodologies

Within the research community, two research methodologies show the most common coverage of many research types. These two research methodologies are called quantitative and qualitative. Quantitative methodologies use experiments and large data sets to reach a conclusion, whereas, qualitative methodologies rely on investigations in an interpretative manner using smaller data sets to create theories or artifacts [6]. As explained earlier in the Introduction chapter, the main research method that was used during this degree project was qualitative. With the reason being that, in order to be able to evaluate the effectiveness of the system, more detailed and subjective data from the users were required, which alleviated the need for a large data set in this case.

Since the evaluation of the system depends on the user's interpretations and opinions of it, the philosophical assumption, interpretivism, was deemed suitable for this degree project. The aim of interpretivism is to discover the meanings that people assign to a phenomenon by exploring richness, depth, and complexity, often, in an inductive manner [6]. By doing so, a general understanding of the phenomenon can be reached [6].

The method used during research was non-experimental for this thesis. The non-experimental method describes or predicts behavior or opinions and can also describe relationships between variables [6]. This method is often used to study the users' behavior or opinions of functionalities and interfaces [6]. Since the thesis work involves user opinions on the effectiveness of the system, the non-experimental method was chosen.

The inductive approach (or reasoning), as explained earlier in the Introduction chapter, is about establishing a general proposition based on particular facts [6]. The data for this approach is collected and analyzed to gain an understanding of the phenomenon and establishing different views of the said phenomenon [6]. Since the evaluation of this system is very opinion heavy, inductive approaches helped verify if the initial proposition was effective or not. Therefore, it made sense to use the inductive approach for this project.

Data collection for this project was done by using questionnaires. Each test subject, anonymously, filled out questionnaires that asked them about their opinions of the system, their prior knowledge of Italian language, their past experiences with robots, their opinion on whether the personalization topic (movies) was a motivating factor or not and an open-ended general feedback question on what could improve the system in the future. The anonymity of questionnaires allowed the users to give out their unbiased opinions. The questionnaires also enabled more detailed questions that elicited detailed responses. When combined with the anonymity, the detailed answers provided much-needed insight into the effectiveness of the system alongside important feedback as to how to improve the system, and possibly systems like it, in the future.

After the data was collected, the gathered information was analyzed using analytic induction. Analytic induction is an iterative method that alternates between collections and analyses [6]. The iterations continue until no cases dismiss the hypothesis or theory [6]. Therefore, by utilizing this data analysis method, the claims of the implemented system were compared to actual results in an iterative manner, which allowed effective and efficient analysis of the test results. The analyses made from the data also showed whether the system was effective in teaching a language or not. Using this data collection method showed that the system was mostly effective except in a number of areas that needed improvements.

5.2 Validation and Verification with Reliability and Replicability

An important process in research is validation. The purpose of validation is to ensure that each phase of the chosen research methodology rigorously adheres to the highest standards of quality [59]. The validation method changes from study to study, and for this research, the construct validity method was utilized. Construct validity ensures that a research effort measures what it is supposed to measure [59]. To validate the thesis project, a user test has been conducted to gather user input and feedback through questionnaires to measure the educational effectiveness of the implemented system.

In qualitative research, verification refers to the mechanisms used during the process of research to incrementally contribute to ensuring reliability and validity and, thus, the rigor of a study [60]. There are several strategies when it comes to verification. Based on a study, five verification strategies are highlighted [60]:

- **Methodological coherence:** Ensuring congruence between the research question and the components of the method [60]
- **Sampling sufficiency:** Appropriate sample that consists of participants who best represent or have knowledge of the research topic
- **Concurrent data collection and analysis:** Forms a mutual interaction between what is known and what one needs to know [60]
- **Thinking theoretically:** Ideas emerging from data are reconfirmed in new data [60]
- **Theory development:** Theory is developed through the outcome of the research process and, as a template for comparison and further development of the theory [60]

With these strategies in mind, for this research, to cover methodological coherence, the system was designed with the research question in mind at all times. To properly fit the research question, the requirements for the system were deliberately chosen and designed for. In terms of sampling sufficiency, students from varying backgrounds and levels of education were chosen as participants most fitting to represent the potential end-users of such a system. Both during and after testing, the collected data was iteratively checked for response correctness and was analyzed to get an immediate sense of what the system was capable of and whether the data provided the answers for the research question or not. As the data was being collected and analyzed, feedback from the users started to show potential cases for improvement and hints at future research. This aspect started to develop an iterative theory which also started being more solidified as more data was collected and analyzed. As a result, a broader, more complete theory was formed, and future research potentials were highlighted.

Reliability is related to the concepts of consistency and repeatability in the data collection [59]. For this research project, parallel forms reliability was utilized to ensure research reliability. Parallel forms reliability refers to the consistency of different, but related measurement tools when applied to the same sample [59]. In the case of this research, the measurement tool refers to the implemented education system, and the sample refers to actual users who are students. Since this research is similar in some ways, and different in other ways, to related studies, the collected data and results were fairly similar. Thus, ensuring research reliability.

The research done in this thesis project alongside system design and implementation details were provided as they were in detail so that this research could be replicated to achieve potentially similar results. It is worth mentioning that the test results could differ based on the test participants. Because the results of the tests are opinion heavy and, different people from different backgrounds with varying education levels could have different opinions about the system, which could alter the results in a positive or negative way.

5.3 Engineering Methods

As a basis, software engineering tasks were utilized as engineering methods for this degree project. These tasks are:

- Requirements Gathering
- Design & Architecture
- Implementation
- Testing

Normally there are five tasks total in software engineering with the fifth one being maintenance. However, as can be seen on the bullet list above, maintenance is not included as part of the engineering tasks that were utilized for this thesis project. The reason why the maintenance task is not considered is because of the fact that the implemented system was not going to be maintained for prolonged periods of time after the testing task has been concluded.

5.3.1 Requirements

In order for a software system to provide the service that is required by its users, the functionality of the system must be determined. In order to do this, requirements must be gathered from either the end-users, shareholders, or the staff that is responsible for developing the software product. Requirements are split into two distinct categories: functional and non-functional. Functional requirements are direct specifications as to how the system should function technically. These types of requirements are very precise and descriptive. Non-functional requirements, on the other hand, represent specifications of the system’s non-functional aspects like security, usability, reliability, and so on. These types of requirements are generally vague, but they have an important impact on user interaction and experience. Both of these requirements must be determined clearly in order for the developed system to serve its purpose in the most proper way possible.

In the case of the degree project, the requirements are self-determined, which means that there is no third party specifying the requirements. This aspect opened up possibilities where the system functionality was more flexible, which made it easier to adapt to necessary setbacks and adjustments to the scope of the degree project during the timeline of the entire project. The requirements for the developed system in this degree project aim to cover the goals of the project while also providing additional features to provide better human-robot interaction.

5.3.2 Software Development Method

While developing the system, the Agile software development method was used. The Agile method is an iterative approach to software development, where the system goes through multiple iterations before it is ready to be released. These iterations include the first four tasks of software engineering (requirements gathering, design, implementation, and testing) and are carried out at each iteration. Utilizing this method enables the system to adapt to changes in requirements more quickly and effectively.

The reason why this method was utilized in this degree project was that: at any given moment, there could be one or more, small or drastic changes to the requirements of the system. These changes require quick action and adaptation in order to not halt the development progress for prolonged periods of time. And that is the reason why the Agile method was utilized in this project.

5.3.3 Testing

The developed system was tested both during and after the implementation. The subsections below will describe how both tests were carried out in detail.

During Development

At each development iteration, the system was tested by the developer. The testing involved, interacting with the robot from scratch, each time, to go over every dialog option that was implemented up to that point. To make sure that most use-case scenarios were tested, different inputs were used in various combinations to check whether the robot was able to handle each input correctly. Also, to test worst-case scenarios, inputs that were not able to be adequately processed by the robot were included in some test cases to see how the robot reacted to those inputs.

Post-Development

After the system was developed and tested thoroughly by the developer, it was tested for real-life use case scenarios. These tests were carried out with humans that were not related to the development in order to make sure that there was no positive bias towards the system. Also, this aspect of testing provided additional insight as to how the robot interacted with new users. Before the tests, the test users were given information as to how the robot works, what the limitations of the robot are, and that they should go over every dialogue option. All test users were asked to go through all study options to cover every aspect of the language education provided by the robot. The test users were also asked to talk to the robot about their movie preferences to test the impact of personalization during study sessions. The rest of the explaining and interaction was done by the robot itself. After the test is concluded, the test users were asked to, anonymously, fill out a questionnaire going over their experience with the robot and whether they thought the education they got was effective or not. After the users filled out the questionnaires, the results were evaluated and discussed.

6 Software Development of the Robot System for Personalized Education

As a solution to the research question, a robot that teaches the Italian language to its users using personalization has been designed. In this system, the robot uses text-to-speech, speech recognition (English and Italian), Natural Language Processing, and facial recognition to teach the users the basics of Italian language using personal, movie-related information. The robot first learns who the current user is by recognizing their face and asking their name. Then, it can ask the user what their favorite movie, actor, and movie genre are and extract that information from recognized speech using speech recognition and natural processing. The robot can use the gathered personalized information during study sessions, which cover: basic Italian vocabulary, basic Italian present tense, and common Italian social conversation phrases and sentences. Thus, providing personalized language education to its user. To develop this system, first, the requirements of the system needed to be specified. And then, a system design needed to be created, which would pave the way for the actual implementation.

The requirements are divided into two categories: functional and non-functional. And, the design aspects of the system are explained in a top-to-bottom manner (high level first, low level after).

6.1 Requirements

The requirements for the system were self-determined, meaning that they were not determined by a third party. While determining the requirements, the goals of this thesis were used as a basis. From which, additional requirements were added on top so that the system provided a better user experience. The requirements for the system are outlined in two categories: functional requirements and non-functional requirements.

Functional Requirements

The functional requirements of the system cover the general features and functions of the system with detailed descriptions. And, the functional requirements of the system are:

- The way of interaction between the robot and the user must be made via human speech
- The language that will be taught must be Italian
- The language used during non-study interactions must be carried out in English
- The robot must be able to learn its users’ faces and names
- The robot must be able to create a user profile for new users
- The robot must be able to recognize and remember previous users
- The system must be able to recognize human speech properly
- The system must be able to analyze recognized speech and extract word roles and named entities (human names, nouns, movie names)
- The robot must be able to give the users options at critical steps to alleviate mistakes

    - Provide a safety net when learning the users' name with a user name confirmation sequence in order to learn a user's name properly
    - Repeat back certain responses that include unique names in order to give feedback to the user that what they said were properly understood or not (extracted movie title, actor name and movie genre)
- The robot must greet the user with their name
- The robot must provide a dialog to learn the users' movie related preferences
    - Favorite movie
    - Favorite movie genre
    - Favorite actor or actress
- The robot must clearly present the study options that are available to the user
- The robot must incorporate movie-related preferences and users' personal information during study sessions

Non-Functional Requirements

The non-functional requirements cover the system's more user experience related specifications with more general and vague descriptions. The non-functional requirements of the system are:

- The system must be secure
    - The personal information of the users must not be revealed to other users or third parties
- The system must be easy-to-use
    - The robot should provide means to navigate through provided functionality easily and convey the instructions on how to navigate the functionality clearly to its users

6.2 Design

When designing the system, first, a dialog diagram was created to give a high-level description of the dialog system for the robot. In this diagram, as can be seen in Figure 5, the robot AI has to follow through forced paths, which can lead to other paths based on input. This input is either gathered from the environment or as a command from the user. Following the diagram, the robot has to introduce itself first. After the introduction, the robot needs to detect a face. If it detects a face, it will check if that face is already known. If it is a known face, the robot greets the user and loads the stored user information file or, if the face is not known, the robot has to ask if the user would like to create a new profile or try detecting the user's face again. The retry face detection functionality here is included to eliminate recreating profiles in the case of faulty face detection scenarios. Where the user may already be registered, but the facial recognition system was not able to classify the detected face properly, which could cause an unnecessary new user profile creation. After the face is recognized properly, and the user is not yet registered, a new user profile needs to be created. The robot has to learn its current user's name first. After the name is understood by the robot, the face and name information are stored under a user profile file stored within the local storage of the NAO robot. Once the user profile creation path is completed, the robot comes to the core decision phase of the application. In this decision phase, the user has two options. The first option is to socialize, and the second option is to start studying with the robot. If the user picks the socialize option, the robot will chat with the user and learn what their favorite movie, actor, and movie genre are which are to be used in the study sessions later on. After the movie preferences are gathered, they are stored in the user information file for future access. If the user picks the study option, the flow of the program moves over to the study phase of the application, where the user can choose which part of the study program they would like to study. The study program offers three key study options which are: vocabulary, grammar, and social conversations. The users can pick whichever study option they want at any given moment.

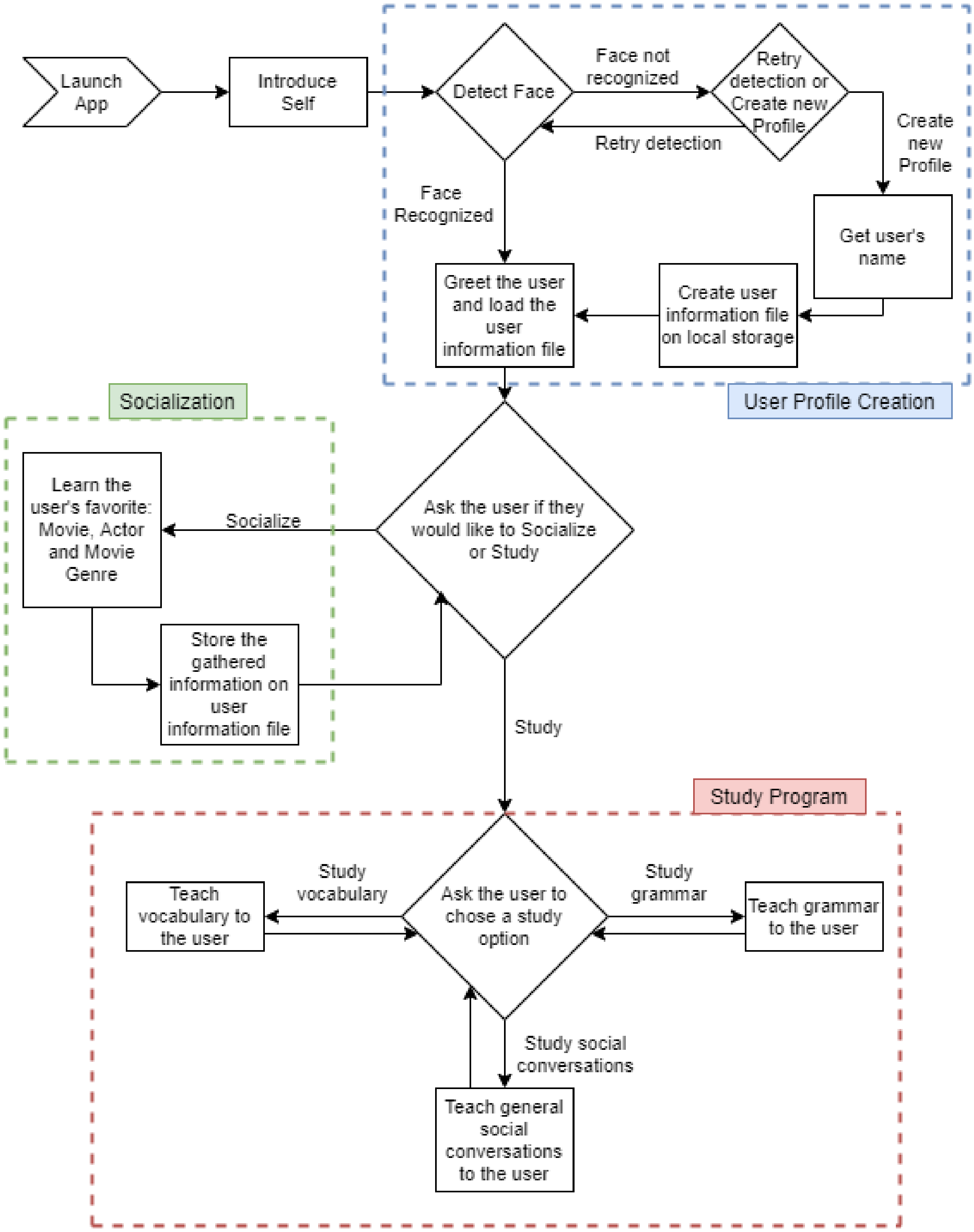

Figure 5 – Robot Dialog Diagram

6.2.1 Facial Recognition and User Profile Creation

In order to detect and recognize faces, the default facilities of the NAO robot were utilized. The NAO robot uses the camera located on its face to stream continuous sets of captured images and analyzes them. These images are analyzed using the built-in facial detection implementation, and then the result is either a “recognized” or a “not-recognized” classification. Although the NAO robot offers these functionalities, they were not enough on their own. A custom behavior needed to be designed around those features to provide the functionality that was required. This functionality is to register users based on their recognized face image alongside additional information of the user. So, in this case, the user’s face is their way of logging in to the system and is used as a means of authentication.

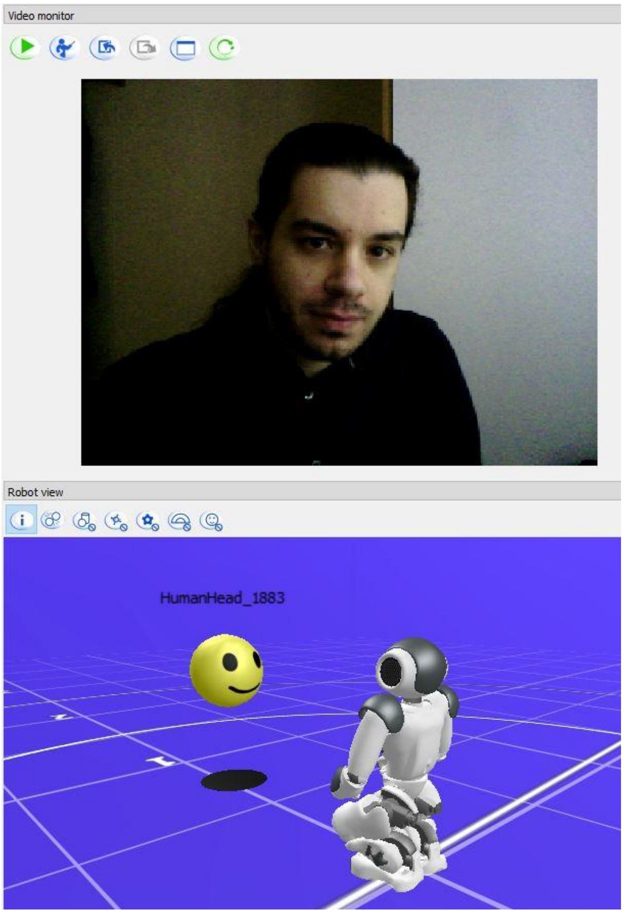

Figure 6 – Choregraphe Facial Recognition

Shown in Figure 6 is how the robot sees and identifies a person. The image was taken from the Choregraphe interface when the robot sees a human face through its camera. As can be seen from the image, the human face seen through the camera has been given identification “HumanHead_1883” which can then be learned and associated with a custom name to be used for identification and authentication.

To create a new user profile. The robot first needs to detect the user's face using the aforementioned facial recognition feature. After a face is detected, the robot then checks whether a user profile has already been created for that face. If it cannot match the detected face with any registered faces, the robot then prompts the user with a message asking them whether they would like to retry the detection process (in the case of possible faulty detections) or set-up a new user profile. If the user wants to set-up a new profile, the robot then has to ask them their name to complete a basic user profile. The user's name will be used while greeting, socializing, and studying with the user. The user's name is also used when creating a local user information file, which contains the user's name, their favorite movie, actor, and movie genre information as well. This information needs to be loaded up each time a registered user logs in to the system using facial recognition.

6.2.2 Socialization

The socialization aspect of the robot is required to provide the personalization feature during study sessions. In order to achieve this goal, the robot utilizes text-to-speech when asking questions and natural language processing to understand keywords from the user's speech. The main topic of socialization is movies. The robot asks three questions in a sequence to learn the user's:

- Favorite movie
- Favorite actor or actress
- Favorite movie genre

At each step in the line of questions, the robot needs to extract the necessary movie preference keyword and repeat it back to the user for confirmation. Once the user confirms that the extracted preference information is correct, that information is stored on the user information file for future use. The questions asked, by the robot, during socialization are:

- ➢ What is your favorite movie of all time?
- ➢ Who is your favorite actor or actress of all time?
- ➢ What is your favorite movie genre of all time?

A speech-recognition implementation and natural language processing were required to recognize user speech and then to analyze the recognized sentence. Once socialization is complete, and the gathered information is stored within the user information file, the personal information can then be used during study sessions to create the personalized education experience. If the user skipped the socialization aspect, for privacy or other personal reasons, the robot can, instead, use a pre-determined user name, movie, actor, and movie genre information during study sessions.

6.2.3 Study Program

When the users want to study the Italian language, they are presented with three study options that they can choose in any order they want. These study options are:

- **Vocabulary:** In this option, some common Italian words are taught to the user alongside their English translations
- **Social Conversations:** In this option, common Italian social conversation sentences, questions and phrases are taught to the users by utilizing personal information
- **Grammar:** In this option, the basics of the Italian present tense is taught to the user by utilizing personal information

Each study option has its own unique content that the robot goes through, and the vocabulary and social conversations options include a more practical approach to the teaching activity. This practicality comes from the required interaction from the user at certain times. The interaction, in this case, is to repeat the last heard Italian word, sentence, or phrase back to the robot to be evaluated by the robot. This evaluation can then alter the rest of the study session by giving priority to the incorrect responses from the user to try and get them to learn that specific language content properly. The grammar option, on the other hand, does not require user input and evaluation to instruct the user. The reason for that aspect is that requiring user input during relatively non-input necessary (one-way) instructions can lead to excessive interactions that can extend the duration of the study session. That can then lead to potential loss of focus or motivation on the user's side.

For each study option, the education content needs to be determined. And this content needs to include opportunities for personalization and the topic that is covered by this project: movies. Each study option required to have its appropriate content created with that specification in mind. The language content for each study option follows the specifications:

- For vocabulary content:
    - General greetings
    - Movie related words
    - Household objects
    - Family members
    - Food items
    - Nature and animals
    - School-related words
- For social conversations:
    - Greetings and goodbyes
    - Common gratitude and excuse phrases
    - Common questions and answers
    - Movie related preference expressions with personalized information
- For grammar:
    - Masculine and feminine words
    - Regular verbs

- Irregular verbs
- Preference expressions with and without personalization

The content created from these specifications is used in a repetitive manner during study sessions to ensure that the information is passed on to the user correctly and is appropriately understood by the user as well. While the study sessions for vocabulary and social conversations take place, the user input gets evaluated for correctness. And, after all of the content has been at least presented to the user once, the incorrect responses have priority during the repetition phase. The reason for this decision was to give higher priority to potentially misunderstood content first instead of going over what the user already understood.

The study program that is provided by the robot covers high-level aspects of the Italian language. It is, by no means, full coverage of the entire language but a smaller, more surface-level coverage of it. The goal of the vocabulary option is to enrich the users' vocabulary knowledge and teach them how to pronounce words in that language correctly. The purpose of the grammar option is to provide a basic understanding of the Italian present tense so that the users can potentially create cohesive and grammatically correct sentences during their usage of that part of the language. The goal of the social conversations option is to prepare the user for common social scenarios that naturally occur when meeting new people, expressing personal preferences, or when making plans with other people.

7 Implementation

The robot system was implemented using Choregraphe [22]. Choregraphe is a development environment created by Softbank Robotics, which allows the development of custom functionality and behavior on the robots that they have created. The programming language that was used during development was Python 2.7. During implementation, the system went through four iterations. At each iteration, a core function of the system was implemented and then the current state of the system was tested. Based on the tests, the necessary fixes and adjustments were made, and the next iteration was initiated.

Iterations	**Tasks**
Iteration 1	Create the dialog outline
Iteration 2	Implement Natural Language Processing
Iteration 3	Implement user profile creation
Iteration 4	Implement the study program

Table 1 - Implementation Iterations and Tasks

Table 1 shows the number of iterations and their respective implementation tasks associated with them; in the order they were initiated.

First Iteration: Creating the dialog outline

The initial task in the first iteration was to create the dialog diagram (as seen in Figure 5) using the built-in functions provided by NAO and Choregraphe. The dialog diagram was created, and it contained the basic structure for dialog flow and the necessary commands to navigate those dialog options. The implemented dialog diagram was also tested, which revealed several issues with the built-in functionalities of the robot. Two default functions on the diagram, although properly configured, did not work as expected. The first function that caused issues was with the default speech recognition function. The default speech recognition only works if a set of words are pre-defined in a vocabulary.

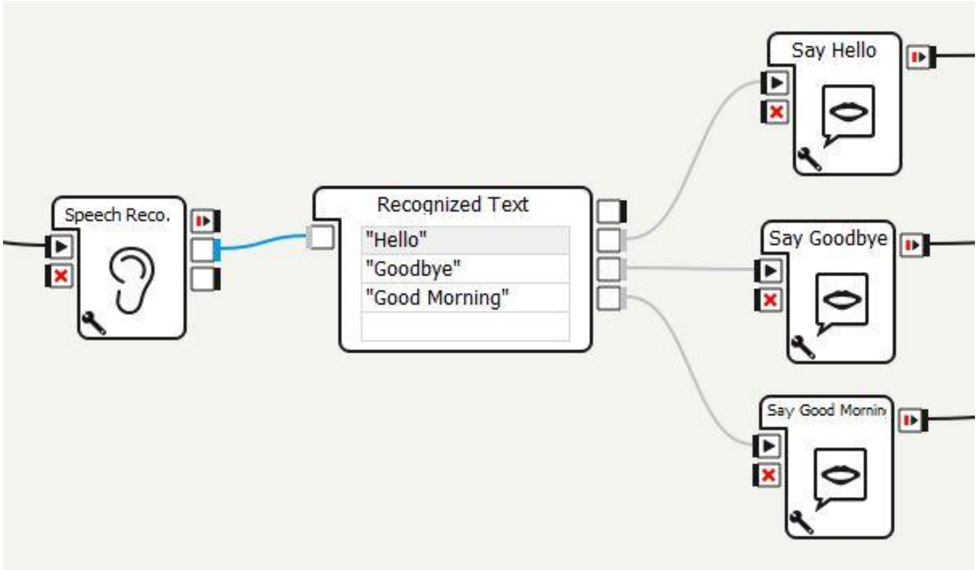

Figure 7 – Default (built-in) Speech Recognition Example

The default speech-recognition module tries to detect the words that are in the vocabulary and outputs those words only. As seen in Figure 7, only "Hello", "Goodbye" and "Good Morning" can be recognized from human speech using the default speech recognition module. Due to the design of the system, this functionality was not sufficient. The ideal speech recognition functionality needed to provide wildcard detection. The wildcard in this scenario refers to words or sentences that are not pre-defined in a speech recognition vocabulary and can be any word or sentence at all. The result of that ideal speech recognition would be an output of the entire sentence, or words, that were recognized as they were heard. The other issue with the default functions that were provided was with the text-to-speech functions. The provided functions would often skip their determined inputs for text-to-speech, which caused issues with providing instructions on how to navigate the dialog diagram. Custom workarounds needed to be created to fix these issues. To solve the first issue, related to speech recognition, a speech recognition Python library was looked at called PocketSphinx, which is a lightweight speech recognition engine [18]. However, due to the read-only nature of the NAO's operating system, the library could not be properly installed on to the robot itself. Choregraphe does provide the functionality to handle this issue. However, using Choregraphe's solution to this issue only resulted in more issues. More specifically, issues related to the dependencies of the library. Finding and including every dependency of the library would only result in further complexity and time consumption. Therefore, the PocketSphinx library was not suitable for this case and was not used in the final system. The actual solution to the first issue was using Google Cloud Speech services to analyze the audio recorded by NAO and outputting the most accurate guess as a whole. The second issue, the text-to-speech issue, was solved by creating a custom Python script. This script still relied on the built-in library methods but used a more direct approach since the default functions were too complicated for the use case needed and would occasionally lose their position in the dialog flow. After the necessary workarounds were implemented, the system was at a functional state. The flow of the diagram worked properly, the robot would use text-to-speech at the right times, and the custom speech recognition implementation was able to analyze heard sentences and words properly and output them as they were heard.

Second Iteration: Implementing Natural Language Processing

The second iteration task was to implement Natural Language Processing on NAO to analyze the recognized sentences and words that were outputted by the custom speech recognition implementation. Since NAO does not have any built-in solution for NLP, a custom Python library seemed to be the ideal choice. Natural Language Toolkit (NLTK), a Python library that provides NLP functionality, was used to handle this task. NLTK is a leading platform built for building Python programs to work with human language data [19]. Using the default way of installing Python libraries to the robot, the NLTK library was successfully installed, and no issues were had with dependencies. The official documentation of NLTK was used to get started with NLTK functions. The documentation provided proper and useful information on how to use NLTK functions, and implementing this part of the system was relatively straightforward. Shortly after, the dialog options regarding the socialization feature of the system were added to the system, and the current state of the system was put to the test with real use case scenarios. These tests included user name, favorite movie title, favorite actor, and favorite movie genre extractions from recognized sentences. The results showed that the implemented system was able to extract named entities from input sentences properly.

Third Iteration: Implementing user profile creation

The third iteration task was to implement the user profile creation part of the system. In order to do this, all of the previous implementation tasks alongside some built-in functionality of NAO needed to be utilized. In order to recognize and learn faces, the built-in NAO facial recognition functionality was used. This, however, was not enough on its own as the robot also needed to gather some additional information about its current user. In order to gather more, verbal, information from the user, the custom speech recognition and NLP implementation was utilized to learn the users' name and their movie-related preferences. All of the gathered user information is then stored locally on the NAO robot to be used for the study sessions and user authentication later on. This implementation task was also tested with several user names and faces to ensure that it was functioning correctly.

Fourth Iteration: Implementing the study program

The fourth, also the final, implementation task was to implement the study program. All three study options (vocabulary, grammar, and social conversations) needed to be implemented with both built-in and custom-made functionality. The first study option that was implemented was the vocabulary option. In order to implement it, a Python dictionary (unordered list of key-value pairs) was used to create the content for the vocabulary option. Content, in this case, refers to storing Italian words (keys) alongside their English translation (values). NAO robot's text-to-speech functionality was used to read out loud the instructions, Italian words, and their meanings during the study session. To do this, robot needed to switch between Italian and

English text-to-speech often, and this aspect was achieved by using the robot's language switch option in code. The user also needed to interact with the robot in this study option as a means of practice. So, to do that, the custom speech recognition was utilized for recognizing both English input and Italian input from the user. Luckily, this was trivially achieved using a separate configuration input for Google Cloud Speech services. The second study option that needed to be implemented was the social conversation option. This option implementation was very much the same as the vocabulary option except for a single additional functionality. This study option includes sentences that require the usage of personalization information gathered from the user. This aspect of the study option required further processing to be done on the language content to include the personalization information (user name, favorite movie, etc.). To achieve this functionality, the language switch feature needed to be utilized even further alongside simple string alteration operations (IE: replacing a character with another string in a sentence). The third and final study option that needed to be implemented was the grammar option. This option differs a lot from the other two options as the robot focuses more on instructions related to the Italian present tense. Since the topic was quite broad, it needed to be separated into smaller chunks to reduce the amount of redundant information being passed down to the user on repetitive use cases. The decision was to divide the instructions into three distinct segments: overview, verbs, and preference expressions. By doing this, the users can study whichever part they want, however many times they want without having to go through parts that they have already understood properly. The study options have then been tested by going over each language content several times with correct and incorrect responses when necessary to determine whether the content was adequately teachable to the users.

7.1 System Overview

The system, in its entirety, can be seen in Figure 8. The image is from the Choregraphe interface and represents the entire system structure that gets executed by the robot. The system incorporates built-in functionality alongside custom code to create the full language education experience.

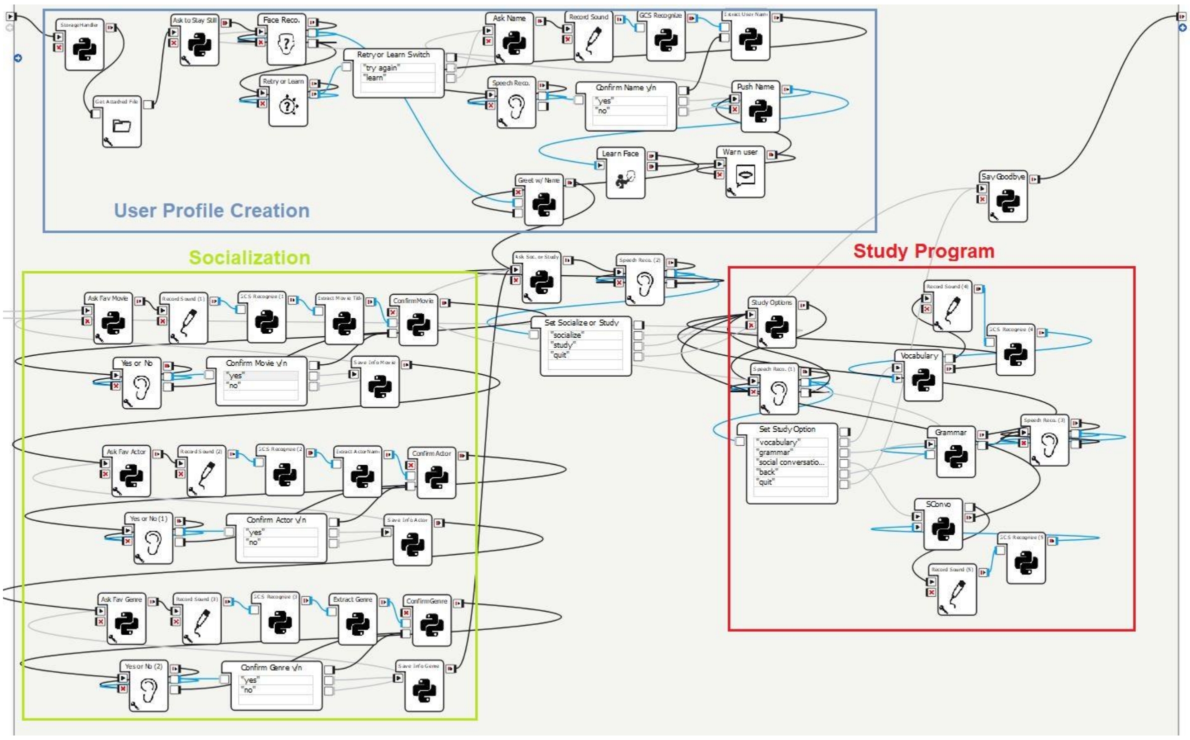

Figure 8 – The Implemented System on Choregraphe

Figure 8 highlights the critical sections of the system, which are: User Profile Creation (Figure 9), Socialization (Figure 10), and Study Program (Figure 11). Due to the complicated nature of the system overview, the essential sections will be explained within their own scope.

User Profile Creation section of the system utilizes facial recognition, speech recognition, and text-to-speech.

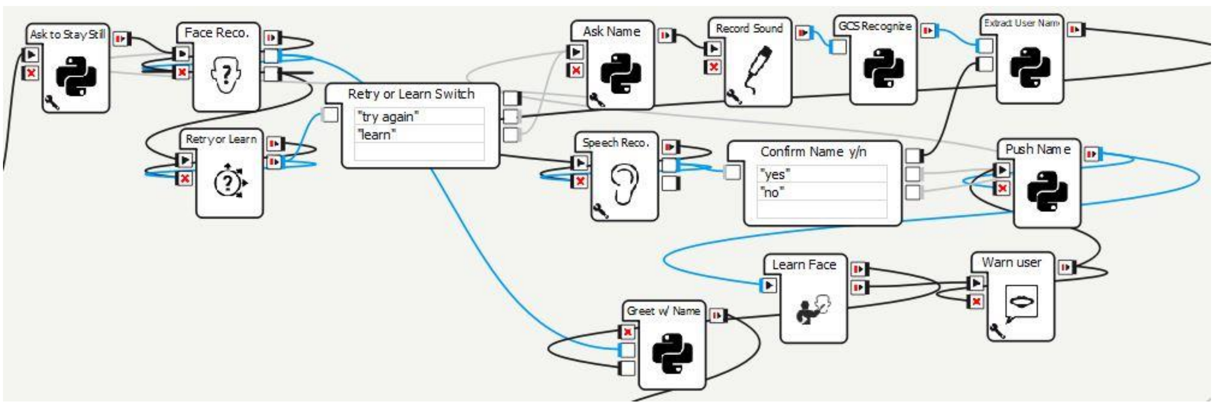

Figure 9 – User Profile Creation Implementation

Figure 9 brings a closer look at the overall User Profile Creation part of the system. The robot, starting from left to right, begins by asking the user to stand still and in front of its camera to recognize their face (Ask to Stay Still box). The robot then proceeds to recognize the face (Face Reco. box). If the face is not recognized, the robot asks the user whether they would like the robot to try detecting their face again (in cases of false negatives) or learn their face to create a new user profile. The “learn” option will prompt the robot to ask the user their name (Ask Name box), record the user's speech (Record Sound box) then send the recording to Google Cloud Speech services for recognition (GCS Recognize box). When the speech is recognized, the recognized text is analyzed for named entity extraction (Extract User Name box). The user's name is extracted from the recognized text using NLP then the robot reads the recognized name out loud to the user and asks them whether it is correct or not (Confirm Name y/n box). If the name is not correct, the flow of operation moves back to asking the user their name again. Or, if the recognized name is correct, the name is pushed to memory to be used for the rest of the session and then stored in local storage for later use cases as well (Push Name box). Once the name is pushed, the face is learned and is associated with the recognized user name. At this point, the user profile creation is complete, and the users can start using the system.

From that point on, the users have two options: to socialize and to study. The robot asks them which option they would like to choose. And, depending on their choice, moves over to either the socialization part of the system or to the study part.

If the users choose the socialize option, they will be presented three sequential questions to extract their favorite movie, actor, and movie genre information, respectively.

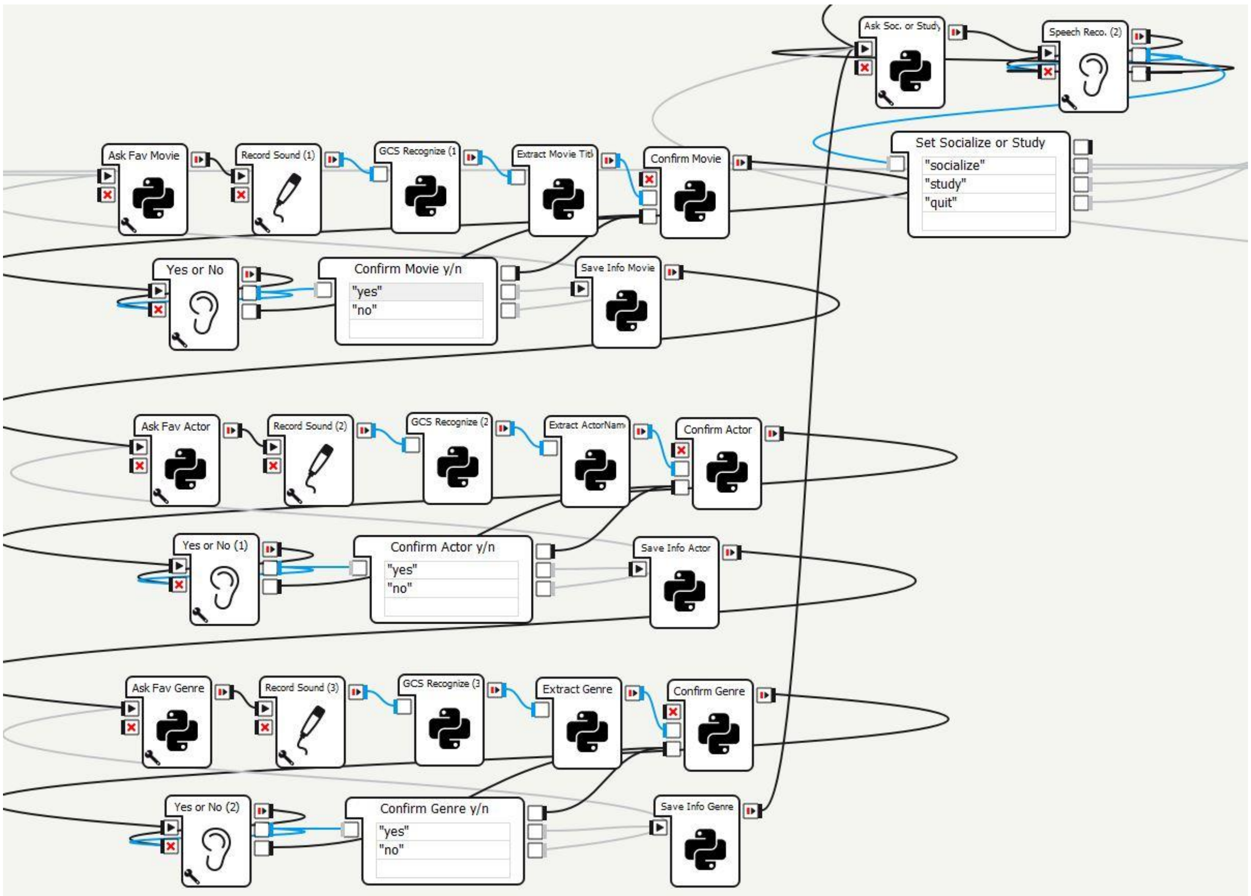

Figure 10 – Socialization Implementation

Figure 10 represents the socialization part of the implementation. The "Ask Fav Movie" box is where the socialization dialog flow starts from (Top-Left). The robot asks the user their favorite movie title, records their speech (Record Sound (1) box) then sends the recording to Google

Cloud Speech services for speech recognition (GCS Recognize (1) box). The recognized speech is returned as text and is then used for named entity extraction using NLP (Extract Movie Title box). The extracted movie name is then read out loud to the user for confirmation (Confirm Actor y/n box). If the movie name is incorrect, the flow of operation moves back to asking the user their favorite movie box (Ask Fav Movie box). Or, if the movie name is correct, the movie name is pushed to memory and stored in local storage under the file created for that specific user (Save Info Movie box). After the movie question and answer, the sequence is finished, the flow of operation moves over to the "Ask Fav Actor" box. The following processes are the exact same as the previous movie name sequence. After the favorite actor information is acquired, the flow moves over to the "Ask Fav Genre" box. The processes here are, again, the same as the two question and answer sequences before. Once the favorite genre information is acquired, the flow of operation moves back to the "socialize or study" selection boxes. From there, the users can move over to the study section.

If the users choose the study option, they will be presented three study options (vocabulary, grammar, and social conversations) that they can choose from.

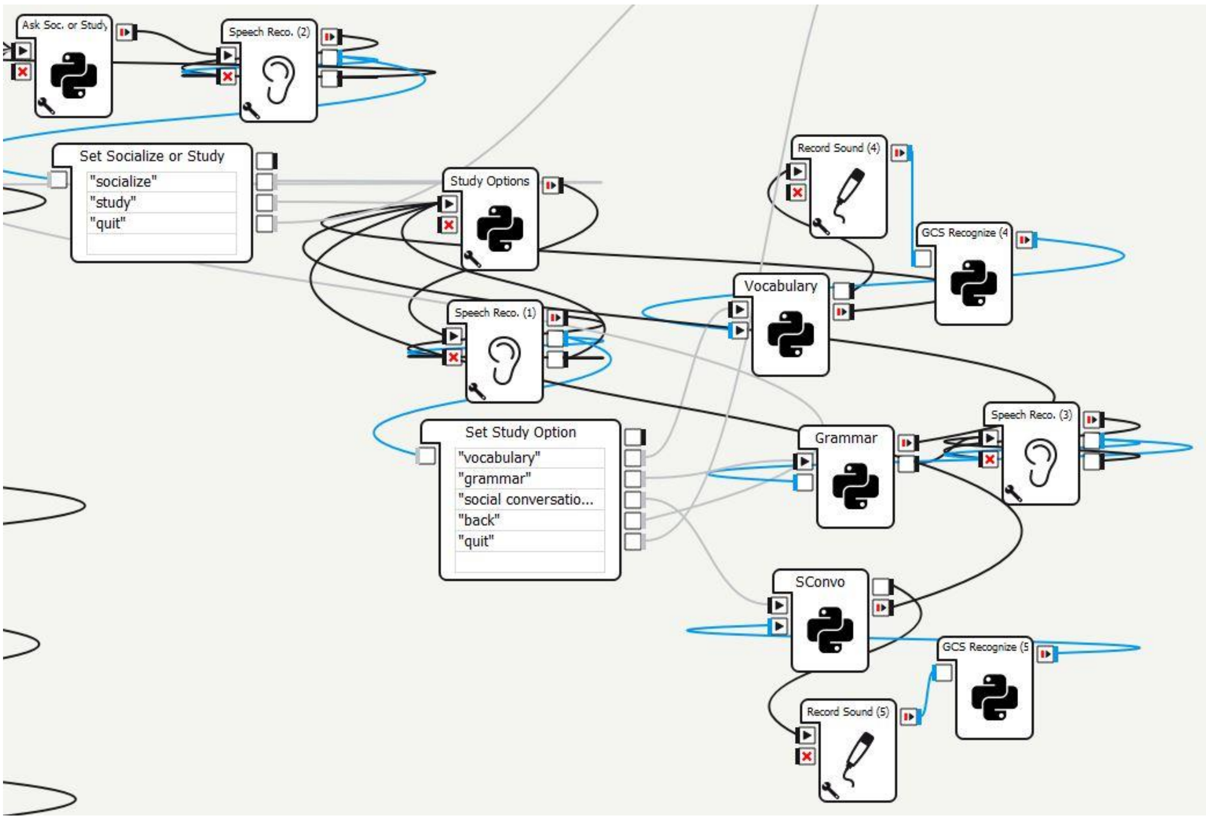

Figure 11 – Study Program Implementation

Figure 11 displays the study program part of the implementation and dialog flow. The flow of operation starts from the "Study Options" box, where the robot asks the users which study option that they would like to study. The three study options are presented to choose from to the user: vocabulary, grammar, and social conversations. If the user picks the vocabulary option, the flow of operation moves over to the "Vocabulary" box. The robot gives instructions on how the vocabulary study session will proceed. The instructions being that the robot will utter a word in Italian that the user can repeat back to the robot after it finished speaking. The user is also given instructions on how to get back to the study option selection. The way they can get back to the study option selection is by saying "Stop" or "Go back" after the robot says a word in Italian. After presenting the instructions, the robot picks a random word from the vocabulary to read out loud to the user. After a word is read out loud, the user speech is recorded and then passed over to Google Cloud Speech services for both Italian and English speech recognition. The reason as to why both Italian and English speech recognition is required is due to the robot both analyzing user's Italian word pronunciation and listening for English "Stop" or "Go back" commands. Once the user speech is recognized, it gets pushed back to the "Vocabulary" box for further analysis. If the recognized text is a "Stop" or "Go back" command, the flow of operations moves over to the study option selection (Study Options box). Or, if the recognized text is the word from the vocabulary in Italian, the robot describes the meaning of the word then picks another random word from the vocabulary. The robot also compares the recognized Italian speech with the word that was read out loud. By doing this, the robot can mark the response as either "Correct" or "Incorrect". The incorrect responses are stored, to be repeated later, once all of the vocabulary content has been presented to the user. The social conversations part of the study option (SConvo box) works in a very similar way to the vocabulary option. The only difference being the utilization of movie-related preferences. Within the social conversations option, there exist several sentences that use the movie-related preferences of the user to provide the personalization feature. From the "Study Options" box, if the users pick the grammar option, the flow of operation moves over to the "Grammar" box. This option works differently from the rest of the options where the user interaction is only required to navigate within the grammar content and to go back to the study option selection. The grammar option offers three sub-sections that the users can study separately. These sections are an overview of the Italian present tense, information on Italian verbs, and expressing preferences. None of the sub-sections require the user to repeat back what they heard to the robot; rather, the robot simply presents descriptions of the topics covered by the sub-sections. In the overview sub-section, the robot goes over the present tense sentence structure and gender identities of Italian nouns with examples. In the verbs sub-section, the robot gives descriptions of Italian regular and irregular verbs with examples. And, finally, in the preference expressions sub-section, the robot describes how to express personal preferences using movies as an example. The preference expression examples also utilize the user's movie-related preferences to provide the personalization experience.

The following sections will provide more in-depth information as to how speech recognition, natural language processing, user profile creation, and study program works.

7.2 Speech Recognition

The built-in speech recognition module on the NAO robot was insufficient for this system. The built-in speech recognition only works with a pre-determined set of words or vocabulary. Since the robot has to learn the users' names, favorite movies, actors, and more, it was necessary to have a speech recognition module that could output any human speech in text form as it heard them. In order to get around that issue, Google Cloud Speech services were utilized in this implementation. Google Cloud Speech is an API used to access Google's Speech-to-Text services. Google Speech-to-Text enables developers to convert audio to text by applying powerful neural network models in an easy to use API [20]. How this service ties into the system is like this:

- The robot records user speech for a certain period of time
- The robot stores that audio file on local storage
- The robot reads the stored audio file as bytes and encodes it with Base64 encryption
- Robot prepares the encoded audio file to be uploaded to Google Cloud Speech services by creating a configuration file
- Robot posts the payload using REST API to Google Cloud servers
- Google Cloud servers respond back with the recognized text
- Robot acquires that recognized text and then works with it

The usage of Google Cloud Speech services introduces a network connection requirement and the latency that comes with using such a service. However, due to the nature of the project, this method was the most feasible and efficient way to recognize speech based on the requirements of this project. Also, moving the speech recognition to the cloud allows the workload on the robot to be reduced. Which makes room for other processes to run their course.

There are, however, certain parts of the system that can work very well with the default speech-recognition functionality provided by the NAO robot. These parts require pre-determined inputs, so there is no need for a more complicated speech-recognition method to process it. In fact, using the default speech-recognition functionality for these types of inputs is fast, efficient, and easy to set-up. Also, using the Google Cloud Speech solution here would only introduce additional latency to the response due to the network connectivity requirement. So, in cases where the speech-recognition inputs are pre-determined and straightforward, the default speech-recognition functionality was sufficient enough.

7.3 Natural Language Processing

In order to enable NAO to handle NLP tasks, Natural Language Toolkit (NLTK) Python library was used. NLTK is an extensive library with lots of functions to use when handling NLP tasks. The parts of NLTK that this system utilizes are word tokenization, part-of-speech (POS) tagging, and chunking. Word tokenization splits the input sentence into smaller subsets of that sentence. POS tagging tags the elements of the sentence based on their role in that sentence. And chunking enables grouping up several pre-determined tag groups and patterns into a new tag to allow the sentence to be analyzed further. Before any of these processes are executed, grammar needs to be defined for chunking. This grammar contains definitions for which tags belong to which new tag group and in which order. By doing this, common patterns can be classified into their own category, which leaves uncommon patterns (mostly noun phrases) exposed for extraction. The way this is implemented in the system is that the common phrases are ignored when it comes to finding the actual name of an entity in that sentence (user names, movie titles, etc.). The grammar defined in this system is as follows:

Grammar Pattern #1: IGNORE: {<PRP.*><.*>*<VBZ|VBP>}

Grammar Pattern #2: IGNORE: {<VBZ|VBP><.*>*?<PRP.*><.*>*}

Grammar Pattern #3: IGNORE: {<NN.*><VBZ|VBP>}

The patterns of statements that are inside the curly brackets are grouped into the "IGNORE" tag, which the extractor function ignores when extracting named entities. The grammar above tries to look for three patterns and group those patterns. The first pattern is a statement that starts with a proper noun and ends with a verb that can have any word in between. The second pattern begins with a verb, which is followed by a proper noun that can have any word in between and after the proper noun. The third pattern begins with a noun, followed by a verb. This last pattern is due to an issue that was observed during implementation with specific sentence structures. These sentences contained lowercase proper nouns, which were tagged as nouns by the POS tagger. Therefore, checking for that pattern reduces potential incorrect outputs.

Based on the grammar, a sentence with a structure like: "My favorite movie is X" can be analyzed and chunked to ignore the "My favorite movie is" part, which is a common phrase in this case. After the common phrase is ignored, the entity name extraction function can then extract the rest of the words in that sentence as they are and, in the order that they were heard. In this instance, only the movie title "X" is detected by the extractor, and the output of the extractor function is the movie title. However, this aspect of the analysis introduces some limitations. The input sentence structure must follow certain patterns in order for the extractor to output the correct words. If the sentence is complex, the output might not be correct. For example, a sentence like: "Y is the movie that I like the most" will most likely lead to an incorrect output. Whereas, if the sentence were like: "Z is my favorite movie of all time" it would most likely lead to a correct output. The reason for this outcome is that the custom grammar defined for extraction tries to find patterns where a proper noun and verb are located in specific ways in a sentence. Thus, if there is more than one proper noun or more than one verb in a sentence, the extractor may or may not be able to chunk the sentence properly, which can lead to incorrect outputs. The grammar could potentially be improved in the future, but it is not a simple task to fit all sorts of patterns into a group and provide correct outputs all the time. The desired outputs from sentences can also vary in their structure and patterns. For instance, movie titles can have varied patterns associated with their structure. Some movie titles include proper nouns, verbs, adverbs, and some of them are just adjectives or verbs in different tenses. These variances make detecting the named entities harder and can also lead to incorrect outputs.

7.4 User Profile Creation

Creating a user login system on a robot is not as straightforward as creating said functionality on a web browser or a smartphone application. Since there is no graphical user interface for the user to interact with, that interaction needs to be handled using speech and facial recognition. Using facial recognition, in this case, allows a user to log in to the system. But first, that user needs to be registered. In order to register a user, the built-in NAO facial recognition and face learning modules were used to detect a face and then store that face information. When a face is not recognized, it means that the robot is talking to a new user who needs to be registered, or it could be the result of bad lighting or other difficult to process environment conditions. To handle both situations, the robot needs to rely on the user for guidance in that matter. The robot asks the user whether they would like the robot to learn their face or to try recognizing their face again. In order for the robot to be able to learn a user's face, it requires a name to associate with that user. Gathering this name with the default speech-recognition module was not possible based on the requirements, so the custom speech-recognition combined with NLP implementation was used to recognize user speech and extract the user's name from the recognized text. Since names are unique and have many variations, the implementation here also provides a way of confirming that the robot understood the name correctly. Once the user is satisfied with what the name that the robot understood, the robot then creates a simple user information file with the fields:

name, fav_movie, fav_actor, fav_genre

This information file is stored as a JavaScript Object Notation (JSON) file to read and write to the file in an efficient manner. JSON is a lightweight data-interchange format that is easy for humans to read and write and easy for machines to parse and generate [21]. After the JSON file is created and stored locally, the users can choose to socialize or study with the robot. The user needs to tell the robot which choice they prefer by using speech. And the robot utilizes the default speech-recognition functionality here due to the nature of the input (pre-determined input). By choosing to socialize with the robot, the user's movie-related preferences are gathered much in the same way their names are gathered. The robot, in the socialization part, asks the user's favorite movie, favorite actor, and their favorite movie genre. At each question, the user's speech is recognized using the Google Cloud Speech services, and the unique name is extracted from the recognized text using the NLP implementation. The user then confirms the extracted unique name and the next movie preference is gathered the same way until all of the required information is gathered (favorite movie, favorite actor and favorite movie genre). Also, after each confirmation, that specific information is stored inside the user-specific JSON file for future use.

7.5 Study Program

After the user profile creation, the user can study with the robot. If they choose to study, they are presented with the three study options: vocabulary, grammar, and social conversations. In order for the user to pick which option they want to study; the built-in NAO speech-recognition module is used to recognize user speech. Since the input is pre-determined, using the Google Cloud Speech solution here would only complicate the process and introduce latency. After the user picks a study option, the flow of the application moves to that specific study option, and a study session begins. During two of the study sessions, vocabulary, and social conversations, the custom speech-recognition implementation is utilized. This time, however, the speech-recognition implementation needs to detect the Italian language as well. To handle that requirement, the language information within the Google Cloud Speech configuration file has been changed to Italian for those use cases. And with that small adjustment, the robot is capable of recognizing both English and Italian spoken language. There are three study options the users can pick, vocabulary, social conversations, and grammar. And these study options differ in the way they are executed and taught to the user. The following sub-sections (7.5.1, 7.5.2, 7.5.3) will go through each study option in greater detail.

7.5.1 Vocabulary

If the user picks the vocabulary option, the robot will teach them common Italian words (includes movie-related vocabulary as well) with a back and forth interaction with the user. This back and forth interaction refers to the robot saying an Italian word and waiting for the user to repeat that word in Italian back to the robot. But, before this interaction takes place, the vocabulary that will be taught to the user needed to be determined. In order to do this, a 50-word dictionary was created from scratch. This dictionary contains words that are related to general greetings, nature, houses, rooms, school, food, family members, and, most importantly, movie-related vocabulary. This dictionary also contains the translations of the words to English as well. And to tie it all together, a Python dictionary was created with the aforementioned dictionary values to be used in code with relative ease. Alongside this dictionary, there are two other Python dictionaries utilized in this part of the program as well. These two dictionaries being: a correct response dictionary and an incorrect response dictionary. These two dictionaries are initially empty, and they get filled out as the study session progresses further. The high-level algorithm as to how vocabulary education is carried out and how the dictionaries work together is as follows:

1. The robot gives instructions as to how the vocabulary education will be carried out
2. While the user does not issue the stop command to the robot
3. Pick a random word from the main vocabulary dictionary
4. Read out loud the selected word in Italian
5. Wait for user input
6. Read out loud the word in Italian again and then read out loud its English translation
7. If the user input is a stop command, go to step 14
8. Compare user input to the last uttered word

9. If the input is correct, remove that entry from the main dictionary and add that entry to the correct input dictionary
10. Else if, the input is incorrect, remove that entry from the main dictionary and add that entry to the incorrect input dictionary
11. Check if the main vocabulary dictionary is empty
12. If the main vocabulary dictionary is empty and if the incorrect dictionary is not empty, pick a random word from the incorrect dictionary, put that word back into the main dictionary, remove it from the incorrect dictionary and go to Step 4 or if the incorrect dictionary is empty, pick a random word from the correct dictionary, put that word back into the main dictionary, remove it from the correct dictionary and go to Step 4
13. If the main vocabulary dictionary is not empty, go to Step 3
14. Stop the vocabulary study session and go back to study option selection

The instructions the robot gives at the start of the study session is: "I will say a word in Italian, and you can repeat it back to me. Also, you can say stop or go back if you want to go back to the study option selection". Basically, this instruction explains the only interaction expected by the user. The reason why the user input is evaluated for correctness is to determine which words from the vocabulary to repeat back to the user first. But, in order for a correct or incorrect word to be repeated, the main vocabulary dictionary needs to be emptied first (the user needs to go through every word in the main dictionary first). And, when it comes to repeating the vocabulary elements, the incorrect responses have the first priority.

7.5.2 Social Conversations

If the user picks the social conversations option, the robot will teach them common Italian phrases, sentences, and movie related preference expressions. These phrases and sentences include greetings, goodbyes, name introductions, basic questions, and, most importantly, expressing movie-related preferences. This study option is the first instance where the personal user information, gathered during socialization, comes into play. Much like the vocabulary dictionary, another dictionary dedicated to social conversations was created. However, the social conversations dictionary contains sentences where there is an additional character that needs to be replaced with personalized information, which is represented with the character 'X'. This aspect requires additional processing to be done on the string that will be read out loud to the user. The character 'X' needs to be replaced with the appropriate personalization information. Also, like the vocabulary option, social conversations session makes use of two additional dictionaries, correct dictionary, and incorrect dictionary, in the same way, they were used in the vocabulary option. The algorithm of the social conversation study session is as follows:

1. The robot gives instructions as to how the social conversation education will be carried out
2. While the user does not issue the stop command to the robot
3. Pick a random sentence or phrase from the social conversations dictionary

4. If the selected sentence or phrase contains the character ‘X’, replace that character with the appropriate personalized information
5. Read out loud the selected sentence or phrase in Italian
6. Wait for user input
7. Read out loud the sentence or phrase in Italian again and then read out loud its English translation
8. If the user input is a stop command, go to step 15
9. Compare user input to the last uttered sentence or phrase
10. If the input is correct, remove that entry from the social conversations dictionary and add that entry to the correct input dictionary
11. Else if, the input is incorrect, remove that entry from the social conversations dictionary and add that entry to the incorrect input dictionary
12. Check if the social conversations dictionary is empty
13. If the social conversations dictionary is empty and if the incorrect dictionary is not empty, pick a random word from the incorrect dictionary, put that word back into the social conversations dictionary, remove it from the incorrect dictionary and go to Step 4 or if the incorrect dictionary is empty, pick a random word from the correct dictionary, put that word back into the social conversations dictionary, remove it from the correct dictionary and go to Step 4
14. If the social conversations dictionary is not empty, go to Step 3
15. Stop the social conversations study session and go back to study option selection

The instruction given to the user at the start of the study session is: “I will say a sentence or phrase in Italian, and you can repeat it back to me. Also, you can say stop or go back if you want to go back to the study option selection”. As can be seen in the algorithm above, the general structure of the study session is very similar to the vocabulary option. Except for a slight addition to handle the personalization requirement. The dictionary for social conversations contains sentences that utilize all of the gathered personal information (user name, favorite movie, favorite actor, and favorite movie genre). If the user did not socialize with the robot (if no personalization information is available), the robot picks personalization information from pre-determined lists for each personalization information category (user name, movie title, actor name, genre name). The robot does that to make sure that even if there is no personalization information available, the study session covers those kinds of sentences or phrases as well. Also, while comparing the user input to the last uttered sentence, if there was personalization involved, the algorithm does not expect the personalization information to be correct within the input. Meaning that, as long as the actual Italian part of the input is correct, the additional personalization information that is detected is not necessary to be evaluated. The reason why this specific evaluation is done this way is due to the reason that the personalized information is in English, and while recognizing Italian speech, the English words that get uttered by the user might not be recognized properly. This outcome can lead to user input being considered incorrect, whereas the actual important part, the Italian part, of the input may have been correct after all.

7.5.3 Grammar

If the user picks the grammar study option, the robot will move the users over to the grammar study session. In this session, the robot reads out loud very basic information about Italian present tense, basic verb information, and preference expression structure. In this section, user interaction is not required to further the study session progress, but rather, it is used to pick out subcategories of the grammar study. These subcategories are:

- **Overview:** Provide basic information about Italian present tense alongside gender associations of Italian nouns
- **Verbs:** Provide basic information about Italian regular and irregular verbs
- **Preference Expressions:** Provide basic information about Italian preference expressions using a unique verb case with personalization involved

The users can pick any of the above categories, whenever they want. During each category session, the robot goes through a static set of information for the most part. However, both verbs and preference expressions are not as static and can vary from time to time. During the verb study sessions, the robot gets to pick from six different verbs twice per session. Which can produce a slightly different session each time the user picks the verb category in succession. The preference expression study sessions rely on gathered personalization information. So, those sessions can also differ slightly based on the personalization information availability. The high-level algorithm for this study option is as follows:

1. The robot gives an introduction and lists out the available grammar study categories: Overview, Verbs and Preference Expressions
2. The robot waits for user input
3. If the user input is Overview, the robot goes through the overview content and when finished, goes to Step 1
4. If the user input is Verbs, the robot goes through the verbs content and when finished, goes to Step 1
5. If the user input is Preference Expressions, the robot goes through the preference expressions content, and if the content includes sentences where the character 'X' is present, it replaces that character with the appropriate personalization information, then, when finished, goes to Step 1
6. If the user input is "stop" or "go back", the grammar study session stops and the flow moves back to the study option selection

This study option follows a very different implementation compared to the other study options. The reason for this is that there is a lot of information to cover, and if at each step during instruction, user input is required, the study sessions can take more extended periods of time than what was allocated by the user. Much like the social conversations option, the grammar option also includes the usage of personalized information during instruction. And, again, also like social conversations, if there is no personalization information available at that moment, the robot picks, randomly, from a list of pre-determined personalization information and goes through the study content regardless.

8 Results and Evaluation

This chapter contains the results from tests performed on the system, followed by an evaluation of the results.

Before testing the system on people, it needed to be tested to ensure that the system was able to function properly without major issues. This test started right after the implementation phase of the system had ended. And, in order to thoroughly test the system, every major part of the system was tested with varying inputs. The major parts were: user profile creation, socialization, and study program. After the functionality tests, the actual user testing was conducted.

8.1 Functionality Testing

Two main inputs, the user's name and their face are used by the system to create a user profile. This aspect was tested by using different images of human faces, which were shown to the robot's camera to be recognized by the facial recognition module, and by using various human names that have been given to the robot as the user's name. The gathered information has been used to register test users to be later logged in to the system with their respective user profile. The results from this test showed that, as long as the lighting conditions are adequate (The person's face and facial features are clearly visible to the robot's camera), the facial recognition system can successfully detect faces of people of different nationalities, ethnicities, and genders. The user name input tests showed that names that were not of English origin were recognized as their closest noun phrase estimates. For instance, the name "Mayuresh" is recognized as "Maya rush", the name "Can" is recognized as "John". This aspect of the recognition system is due to the fact that the Google Cloud Speech only works with one specific language at a time for given audio input. And, if a sentence contains a mixture of words from multiple languages, the specified language for speech recognition is the only language that can be accurately recognized. The user name recognition aspect of the system works properly for names of English origin.

There are three questions asked during the socialization phase of the system. The questions ask the users about their favorite movie, favorite actor/actress, and favorite movie genre. All questions have been given answers in four sentence structures with different movie titles, actor names, and movie genres to test how the system evaluates the responses to these questions. The four sentence structures used during testing are as follows:

- My favorite X is Y
- Y is my favorite X
- I like Y
- Y

The 'X' in the above sentences are filled in with either "movie", "actor" or "actress" and "genre". Whereas the 'Y' character is filled in with the actual movie title, the actual name of the actor or actress, or it can be the name of the user’s favorite movie genre. The reason why four different sentence structures were used was to test the Natural Language Processing implementation's ability to extract named entities from sentences with varied sentence structures. Also, the above sentence structures were deemed the most common responses for such questions regarding preferences. The results showed that the NLP implementation was able to successfully extract the named entities from such sentences as long as the named entity is not another complex sentence. What “another complex sentence” means is that the named entity is not a full sentence structure that includes a verb of some form. When the named entity for extraction is a complex sentence, the extracted entity can be split into parts, and the resulting title or name can be incorrect. The reason for this outcome stems from the fact that the defined grammar for NLP parsing looks for patterns within the sentence based on the location of the verb. That can then lead to recognizing patterns and marking parts of the named entity as part of the sentence, which is not considered part of the named entity any longer. When it comes to actor names, the Google Cloud Speech implementation is better able to recognize varying actor and actress names compared to regular names. This could be due to how Google Cloud Speech services have been trained. Also, the actor and actress names are possibly more frequently encountered by the services, which may have made the service more aware and accurate with popular human name combinations.

During the study sessions, the robot utters a phrase or a sentence and waits for the user to repeat what they heard back to the robot for evaluation. To test this aspect of the study sessions, correct and incorrect responses were given to the robot. This was done to determine whether the robot was able to distinguish correct and incorrect responses during evaluation accurately. Also, the accuracy at which the speech recognition system recognized human speech in Italian was tested as well. After the robot says a word, sentence, or phrase in Italian, the user has the option to repeat what they heard back to the robot to see whether they were able to understand and pronounce what they heard. The robot evaluates that response to provide repetition in the case of incorrect responses after all of the study content has been gone through at least once. So, the test, in this case, was done to check whether the robot correctly understood and evaluated user input. A test was conducted in which all of the study content has been gone through five times in total for each study option (vocabulary, grammar, social conversations). The results showed that the speech recognition implementation was able to recognize words, sentences and phrases in Italian correctly, and the robot was able to properly evaluate the user response as either being correct or incorrect. Also, the robot was able to repeat the content properly, which was given an incorrect response by the user.

8.2 User Testing

User tests were conducted with six volunteer users who were given around 30 to 45 minutes to interact and study with the robot. Before their interaction began, they were informed of what the robot could do, what their options were, and which dialog path that they should follow. The users were also asked to keep their responses to the robot as simple as possible (in terms of response sentence structure). To make sure that the interaction between the user and the robot was functional. The robot's behavior was monitored during the test sessions.

The test starts with the user profile creation part of the system. The robot scans the current user's face, asks them their name, and then creates a user profile with which the user can start to use the other parts of the system. After the profile creation, the user goes through the socialization part of the system. While socializing, the robot asks the user their favorite movie, actor, and movie genre. After the user's movie-related preferences are learned, the user moves over to the study part of the system. The user gets to pick from the three study options (vocabulary, grammar, social conversations), then proceeds to study the option they chose. Then the user is asked to pick one of the other options until all three of the study options are covered. Once the user goes through all study options, the test is concluded, and the data collection part of the evaluation takes place.

To get the feedback from the user, first, a verbal (informal) dialogue with the user is held. During this dialogue, the user is asked, by me, what their overall opinions on the system were and what they would change/add to make the system better. Once the verbal feedback is gathered, the users are given a questionnaire that they need to fill up to conclude the data collection part of the test, which is also the end of the user test session.

8.2.1 User Testing Results

The results from the tests will cover user feedback, questionnaires, and observations made on the system during test sessions.

The verbal feedback, acquired from the users, suggested several technical, functional, and quality-of-life related improvements to the system. Most users thought that the system could have been more responsive to commands and faster in processing. Some users also thought that having a screen or a display to show the Italian words and sentences read out loud by the robot to better understand how that specific word or sentence is written and pronounced. Some users commented on the robot's speech in Italian and how hard it was to understand it when it was speaking fast or was reading a long sentence out loud. The final feedback came from one user where they thought that if the robot had a way to give feedback on the user's progress while also providing the necessary tools to examine the user's current skill level (via tests, quizzes), the education would feel complete. The same user also thought that the robot could be less formal and have more personality that could be customized to fit the users' preferences. Overall, most users thought that a system like this could be useful in education if the necessary improvements were made on it.

After the verbal feedback, every user filled out the questionnaires. The questionnaire was created using Google Forms, which is a free, online tool to create question and answer forms [37]. There are five mandatory questions and one optional follow-up question in the questionnaire. The full questionnaire can be seen in Appendix A. The results from the questionnaire from question number 1 to number 5 are as follows:

1. Did you have any prior experience working or interacting with robots? (Response is presented in Figure 12)

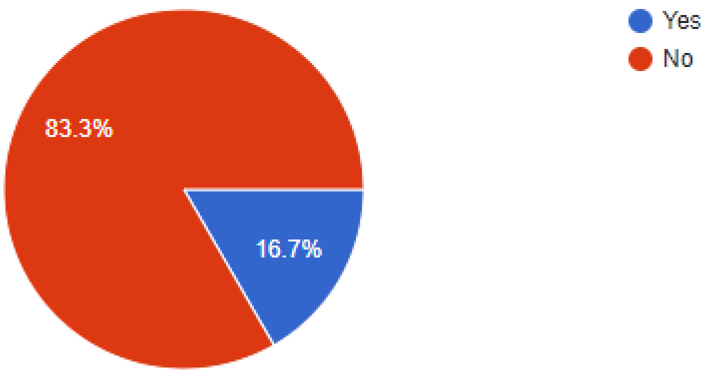

Figure 12 – Answers to Question #1

2. How would you rate your knowledge of the Italian language prior to interacting with the robot? (Response is presented in Figure 13)

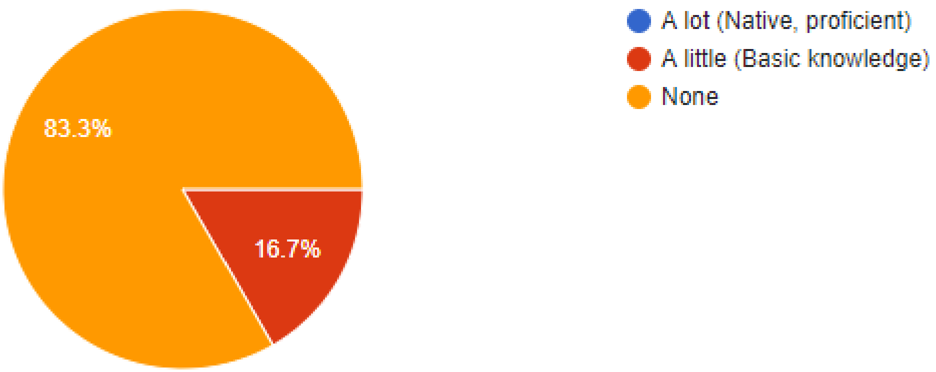

Figure 13 – Answers to Question #2

3. Do you think that the robot was able to effectively teach you the basics of the Italian language? (Response is presented in Figure 14)

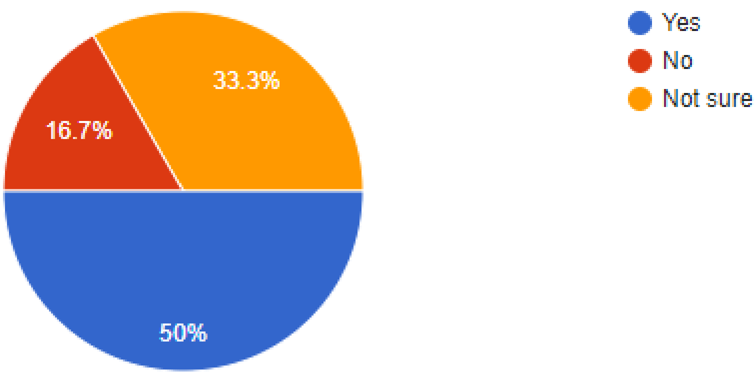

Figure 14 – Answers to Question #3

4. Do you think that the inclusion of movie-related information and subjects during the education was a motivating factor for you? (Response is presented in Figure 15)

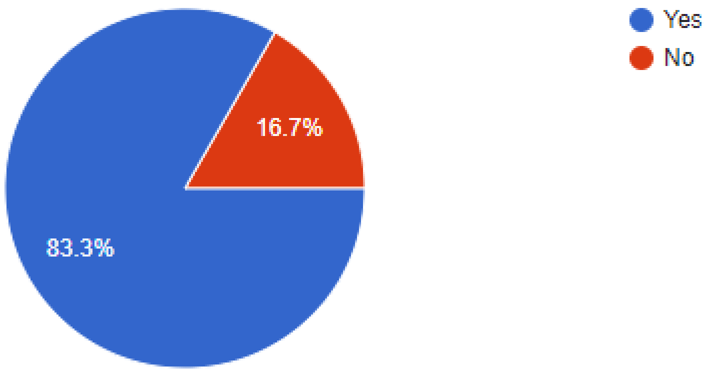

Figure 15 – Answers to Question #4

   a. (Follow up) If you answered "No" to the previous question, what other topic might motivate you?
      i. Sports
      ii. Something related to where the person is from. Questions about home country maybe.

5. In your own words, what more could be improved or added to the system to make it more motivating and effective when teaching a language?
   a. The text which is read out by the robot could also be shown on the screen.
   b. The robot takes a bit too long to think
   c. Feedback and a testing element would make for a more full experience. The personalized topics were the most engaging. The robot is currently very formal, but it might be more approachable with a less rigid or more casual interface. I imagine it would be fun to have personality settings that could be customized to the user (for example, a humor setting that could be turned up and down).In general, anything that makes it feel like the robot is an intelligent system makes it more likely for me to want to continue interacting with it.
   d. Improve the pronunciation and make it understand a wider range of commands
   e. More topics in the section 'Socialize'
   f. It would be so beneficial if it had a screen that shows the spelling at the same time the robot teaches. Also, the robot has a time limit for the responses, and it makes it hard when you want to think about the answer a little bit long, hence, it would be much more efficient if it waits till I respond.

The graphs shown in Figures 12, 13, 14, and 15 were created automatically by Google Forms.

Observations from the test sessions were also noted down for providing additional information as to how the system could be improved. The test users were, for the most part, non-native

English speakers. Some users had prominent accents based on their home country. One user, who was from France, had trouble interacting with the robot due to the robot not being able to recognize their speech properly. However, another user, who was from Spain, had little to no trouble interacting with the robot due to the robot being able to understand their speech properly. Another observation was made with how the NAO robot's built-in microphones pick-up sounds. In this case, a user who spoke in a lower volume was not properly understood by the robot at times, which caused the need for the user to repeat their commands several times to the robot for it to understand it. The final observation that was made during testing was that the facial recognition implementation of the robot was not able to recognize a person's face at all and recognized two different people as being the same user.

8.3 Evaluation

The overall reactions and feedback on the system are mostly positive (Figure 14). However, this does not mean that the system is ready for use in real-life scenarios as of yet. Based on the users' feedback, the system needs a lot more work in order for it to provide a more effective, motivating, and complete education to its users. The robot needs to be more responsive, personal, and adaptive.

For most users, the interaction with the robot system implemented in this project was their first real experience with a robot (Figure 12). The users were very excited to interact with the robot and learn what it could do. The users were also very patient and careful when interacting with the robot in general. This outcome is possibly due to the instructions given to them before the tests begun. The users were very understanding of both the robot's and the system's limitations when given the necessary information about those shortcomings.

Prior knowledge of the Italian language of the users shows that the majority of test subjects had little to no experience with the language (Figure 13). This question was asked to create a distinction between the users based on their skill levels to see how much their opinions would change on the overall effectiveness of the system. However, since there were no answers that picked the option "A lot (Native, proficient)", it was hard to distinguish the results based on the answers to question 2. Gathering opinions on native or proficient Italian language users could provide additional information on the effectiveness of the system.

Most users pointed out that the robot simply took too long to process speech. This aspect could be improved on both software and hardware levels. In terms of software, speech recognition implementation can be more streamlined and smarter. Instead of recording the speech of the user for a constant amount of time, the system could recognize when a user starts speaking and ends their speech and record only that part for recognition. That solution could potentially allow the robot to gather and understand responses quicker, in turn, making it more responsive. The network requirement for speech recognition is another reason as to why there is such a delay in speech recognition. That aspect could be improved with a faster (lower latency) connection or not needing that network element at all. If the speech recognition could be done on local hardware, the responsiveness of the system can improve drastically. However, this does require more powerful hardware, which could be improved by the manufacturers in the future.

Another quality-of-life suggestion made by the users hint at the need for a way to see the robot's speech in text form. This issue could be solved by having a third-party display, which the robot lacks, to display what the robot is saying at all times. The inclusion of such a display could also improve the attention the users might give to the education. For instance, if the robot displayed the image of an apple when it said apple in Italian, it could create a more visual connection to the word, which could improve the effectiveness of the education.

As seen in Figure 15, the personalization topic used in the system, movies, related positively to most people. But some users thought that having a different topic could help motivate them to interact more with the system. This aspect could be improved by expanding the socialization and study sessions to cover more topics. Also, the users could choose which topic motivates them; in turn, this could improve the rate of interaction.

The robot can also provide a means to give feedback (positive or negative) to the user on their current progress. This does require robust speech recognition and analysis. During testing, the robot had trouble understanding human speech when spoken softly or in heavy accents. This issue could create scenarios in which the user actually pronounced the word correctly but the robot did not understand what they said correctly, resulting in a false-negative analysis of the response. If and when spoken language can be understood correctly by the robot, true feedback to the users could be given. At the moment of testing, providing this feedback was not possible for this implementation. This is mostly due to the lack of robustness in speech recognition implementation. Since, at certain points, the speech recognition implementation was not able to recognize heard sentences or words correctly. If the user input was correct, but it was recognized as incorrect, the immediate feedback given to the user might confuse them or de-motivate them. Some users also suggested a way to test their knowledge using the robot. This suggestion, again, ties to speech recognition robustness. But it can be included when the speech recognition robustness can be achieved.

Overall, the results of the questionnaire showed that most people did not have experience working or interacting with robots while also not having a lot of Italian language knowledge. While half of the users thought that the education they received from the robot, was effective, the other half were either not sure or did not think it was effective. This could be due to the education provided by the robot not meeting their expectations. Since based on the feedback, the robot clearly needs more work and features to make it more suitable for more effective education. Another reason as to why they did not think it was effective could be due to the personalization topic not being something that motivates them. Perhaps if the system can incorporate a feedback mechanism and provide more personalization topics, it can provide higher quality education.

The idea of a personalized language education robot was feasible and, to some extent, effective. Utilizing a robot in this scenario has been a good choice since the users were very curious about the robot, excited in interacting with it, and mostly motivated to study alongside the robot. With the necessary improvements, the idea of utilizing such a robot system in education can be beneficial for future generations.

9 Conclusions and Future Work

The thesis presents a personalized language education robot system and an evaluation of its educational effectiveness. The system has been implemented using the NAO robot and modern AI technologies. With the use of facial recognition, speech recognition, NLP, and text-to-speech, the robot is capable of teaching basic Italian language to its users. In order to provide facial recognition and text-to-speech functionality, the built-in modules that exist within the NAOqi framework were used. For speech recognition functionality, Google Cloud Speech services were used, in addition to the NAOqi framework's built-in speech recognition modules to provide a more robust and flexible speech recognition method. NLP functionality was implemented by utilizing a third-party Python library called NLTK. By combining all of these features together, the robot is able to teach basic Italian vocabulary, grammar, and common social conversation sentences, questions, and phrases to its users using verbal communication.

The developed system offers a personal, motivational, and flexible education to its users. These aspects of the system are made possible with the help of personalization and the NAO robot. To answer the research question: With the help of robotics technology and AI, what can be done to provide a more personal and effective language education? The system has utilized personalization, with modern AI and robotics technology, to provide a working and mostly effective solution to the problem. To conduct the research for this thesis, qualitative methods have been used to gather and analyze information about the idea of a personalized language robot. One of the key goals of this thesis was to evaluate the educational effectiveness of the implemented system. To achieve this goal, the system has been tested on actual students to understand which aspects of the system worked and which aspects did not. And, to ultimately uncover whether the system provided an effective language education to its users. The data from the tests were collected using questionnaires, which were then analyzed using the Analytic induction method.

The test results showed that the robot system needed additional work in order for it to provide a more effective and complete language education. While half of the test users thought it was effective, the other half were either unsure or did not think it was effective at all. In order to make the system more effective, it needs to be more responsive, needs to have broader personalization, use a display to show what the robot is saying, have better text-to-speech to make it more understandable, needs to have a robust speech recognition implementation and in general it needs to be more friendly and relatable. Thus, the closer the robot is to an actual human being, the better the education provided by it can be effectively understood by humans.

Compared to related work, the developed system, combining some of the key improvements mentioned within the related work, achieved similar results in the short term. Compared to the study shown in [11], where a robot has been shown to improve the linguistic capabilities of children, the results from the user tests done in this thesis project showed that the education given to the users was effective for the most part. Another study shown in [12] produced effective short-term results using a robot language tutoring system that adapted to the users' skills when tested on children. Compared to the study [12], the results from the thesis project showed that utilizing personalization as a way of adapting to users' preferences showed promising results as well. The thesis project, when compared to another study shown in [13], which successfully motivated children to learn a language using a robot, achieved similar results where the participants of the test mentioned that they were excited and motivated to interact with and learn from the robot. A study, shown in [14], showed that robot systems could be motivational tools for learning a language if they have more things in common with their users. Compared to the study [14], the thesis project solidified the theory of the study [14] by showing that when the robot system has more things in common with its users (movie-related preferences), it can be a mostly effective and motivational tool in terms of language education. On the more technical side of things, a study shown in [15] presents a social robot that utilizes NLP to extract its user's personal information using natural speech. The results of that study [15] showed that gathering more in-depth and personal knowledge of the users could improve the users' satisfaction rating. When compared to that study [15], the thesis project also showed that gathering more personal information about the users could help improve its service based on user feedback. Another study demonstrated in [31] presented a system that could understand and recognize child speech to provide game-like informative activities with them. Although there are no solid results for the study [31] other than it being titled as promising, the thesis project was able to utilize speech recognition and natural language processing to provide promising results as well. In terms of education and personalization, the study presented in [51] showed positive results when incorporating personalization to teach a language to its user. Much like the study [51], the thesis project also presented positive results when it comes to incorporating personalization techniques during education since most of the participants responded very positively to the inclusion of their personal movie-related preferences during the study sessions. In another study where personalization was utilized to adapt to the user's skill level [52], the authors of the study believed that the system could provide a flexible and ubiquitous learning environment for learning English. Compared to that study, the thesis project showed that personalization could also be used to relate to the users to be effective. And, a robot, that is capable of mobility and can be used anytime, could be used to provide a learning environment that is flexible and ubiquitous. Finally, a study described in [53] presents a personalized language education system that is mobile and flexible. Compared to that study [53], the robot system, developed as part of the thesis project, can provide personalized education using the personal information of the users in a mobile and flexible manner. Which is possible thanks to the NAO robot's ability to be used untethered and at any given time.

9.1 Discussion

The idea of utilizing AI in language education in the shape of a robot, then doing research in that area, developing the system and testing it first-hand was an overall insightful and worthwhile effort. Through this effort, more information on how AI in education can be an effective tool and what kind of improvements and features are needed in such a system have been gathered from tests on actual students. Also, designing and implementing such a system, highlighted the use of modern technologies and some challenges that other developers may face as they implement similar systems. Challenges like library incompatibility, lacking built-in features, network latency, and slow hardware needs to be considered when developing a robot education system. These challenges can possibly be overcome with better hardware, better software support, robust networking, and the developer’s intuition.

The thesis, as a whole, has presented relevant background information, followed the research and engineering methods, involved specifying, designing, implementing, and testing of a software system, and evaluated its results. The thesis project has achieved its goal, which was to provide an answer to the research question and then evaluate its effectiveness in real use-case scenarios. The outcome of the project pointed out some critical areas for improvement in order for the system, and systems similar to it, to achieve even greater and more successful results. The aspect of personalization and robotics technology of the project, combined within the NAO robot, has been proven to be a motivating factor, and going for that direction has provided insightful knowledge within the AI in education field. The developed system showed the strengths and shortcomings of such an idea in the form of user opinions and feedback. There is much to improve and add to the project, but, overall, it was a worthwhile effort in terms of research.

9.2 Future Work

More features can be added to the system, and existing features could be made more robust and responsive to improve the educational effectiveness of the system and provide more quality of life for its users in terms of user experience. Based on user feedback and findings from the research, adding features like more languages, being able to select a specific language to learn, more personalization aspects (sports, games, music, etc.), utilizing non-verbal communication during education (moving arms, nodding head, etc.), and a way for users to test their own abilities in the shape of quizzes or tests for a specific language can help improve the system’s educational effectiveness. Also, if the speech-recognition is more responsive and robust, the users can have a better experience interacting with the robot. Some parts of the system can also make use of better hardware to improve the overall responsiveness and capabilities of the system as well. If the robot is able to process all of the information faster, it can also respond faster. Another feedback, taken from the test participants, highlights the potential educational effectiveness increase if there was a display that showed what the robot was saying using both text and images. These additional features and improvements could potentially increase both short-term and long-term effectiveness of the system by providing more flexibility to the users’ needs while also improving motivation to study with the robot. Once the improvements and the additional features are added to the system, it can also provide complete coverage of a language as well. In terms of testing, the system could be tested on more users from a wider range of age groups and social backgrounds to see which demographic the robot system appeals to the most. More research within this area, AI in education, can provide additional knowledge into how these types of educational tools can be effectively utilized in the future.

References

[1] Luckin, Rose, Wayne Holmes, Mark Griffiths, Laurie B Corcier, Pearson (Firm), and London University College. *Intelligence Unleashed: An Argument for AI in Education*, 2016. https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/innovation/Intelligence-Unleashed-Publication.pdf.

[2] NAO the Humanoid Robot, https://www.softbankrobotics.com/emea/en/nao

[3] Mubin, Omar, Catherine J. Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. "A Review of the Applicability of Robots in Education," 2013. https://doi.org/10.2316/Journal.209.2013.1.209-0015.

[4] Dale, Robert, Hermann Moisl, and Harold Somers. *Handbook of Natural Language Processing*. CRC Press, 2000.

[5] Bateman, John, and Michael Zock. "Natural Language Generation." *The Oxford Handbook of Computational Linguistics 2nd Edition*, April 1, 2014. https://doi.org/10.1093/oxfordhb/9780199573691.013.010.

[6] Håkansson, Anne. "Portal of Research Methods and Methodologies for Research Projects and Degree Projects," 67–73. CSREA Press U.S.A, 2013. http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-136960.

[7] Fong, Terrence, Charles Thorpe, and Charles Baur. "Collaboration, Dialogue, Human-Robot Interaction." In *Robotics Research*, edited by Raymond Austin Jarvis and Alexander Zelinsky, 255–66. Springer Tracts in Advanced Robotics. Berlin, Heidelberg: Springer, 2003. https://doi.org/10.1007/3-540-36460-9_17.

[8] Hegel, Frank, Claudia Muhl, Britta Wrede, Martina Hielscher-Fastabend, and Gerhard Sagerer. "Understanding Social Robots." In *2009 Second International Conferences on Advances in Computer-Human Interactions*, 169–74, 2009. https://doi.org/10.1109/ACHI.2009.51.

[9] Dautenhahn, Kerstin. "Methodology & Themes of Human-Robot Interaction: A Growing Research Field." *International Journal of Advanced Robotic Systems* 4, no. 1 (March 1, 2007): 15. https://doi.org/10.5772/5702.

[10] Arkin, Ronald C., Masahiro Fujita, Tsuyoshi Takagi, and Rika Hasegawa. "An Ethological and Emotional Basis for Human–Robot Interaction." *Robotics and Autonomous Systems*, Socially Interactive Robots, 42, no. 3 (March 31, 2003): 191–201. https://doi.org/10.1016/S0921-8890(02)00375-5.

[11] Eun-ja Hyun, So-yeon Kim, Siekyung Jang, and S. Park. "Comparative Study of Effects of Language Instruction Program Using Intelligence Robot and Multimedia on Linguistic Ability of Young Children." In *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, 187–92, 2008. https://doi.org/10.1109/ROMAN.2008.4600664.

[12] Schodde, T., K. Bergmann, and S. Kopp. "Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making." In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 128–36, 2017.

[13] Tanaka, Fumihide, and Shizuko Matsuzoe. "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning."

J. Hum.-Robot Interact. 1, no. 1 (July 2012): 78–95. https://doi.org/10.5898/JHRI.1.1.Tanaka.

[14] Kanda, Takayuki, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial." *Human–Computer Interaction* 19, no. 1–2 (June 1, 2004): 61–84. https://doi.org/10.1080/07370024.2004.9667340.

[15] Hameed, I. A. "Using Natural Language Processing (NLP) for Designing Socially Intelligent Robots." In *2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, 268–69, 2016. https://doi.org/10.1109/DEVLRN.2016.7846830.

[16] Collobert, Ronan, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. "Natural Language Processing (Almost) from Scratch." *Journal of Machine Learning Research* 12, no. Aug (2011): 2493–2537.

[17] Chowdhury, Gobinda G. "Natural Language Processing." *Annual Review of Information Science and Technology* 37, no. 1 (2003): 51–89. https://doi.org/10.1002/aris.1440370103.

[18] Pocketsphinx, https://github.com/cmusphinx/pocketsphinx

[19] Natural Language Toolkit (NLTK), https://www.nltk.org/

[20] Google Speech-to-Text, https://cloud.google.com/speech-to-text

[21] JSON, https://www.json.org/json-en.html

[22] Choregrapge, http://doc.aldebaran.com/2-4/software/choregraphe/choregraphe_overview.html

[23] Koller, Daphne. "Technology as a passport to personalized education" In *New York Times*, 2011. https://flipped.community.uaf.edu/files/2013/12/Daphne-Koller-Technology-as-a-Passport-to-Personalized-Education-NYTimes.com-DaphneKoller_TechnologyasaPassporttoPersonalizedEducation_NYTimes.pdf

[24] NAOqi Framework, http://doc.aldebaran.com/1-14/dev/naoqi/index.html

[25] VanLEHN, KURT. "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems." *Educational Psychologist* 46, no. 4 (October 1, 2011): 197–221. https://doi.org/10.1080/00461520.2011.611369.

[26] Roll, Ido, and Ruth Wylie. "Evolution and Revolution in Artificial Intelligence in Education." *International Journal of Artificial Intelligence in Education* 26, no. 2 (June 2016): 582–99. https://doi.org/10.1007/s40593-016-0110-3.

[27] Belpaeme, Tony, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. "Social Robots for Education: A Review." *Science Robotics* 3, no. 21 (August 15, 2018). https://doi.org/10.1126/scirobotics.aat5954.

[28] Danner, George E. "Where Do We Put the Humans?" In *The Executive's How-To Guide to Automation: Mastering AI and Algorithm-Driven Business*, edited by George E. Danner, 113–20. Cham: Springer International Publishing, 2019. https://doi.org/10.1007/978-3-319-99789-6_11.

[29] Steinfeld, Aaron, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. "Common Metrics for Human-Robot Interaction." In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*,

33–40. HRI ’06. Salt Lake City, Utah, USA: Association for Computing Machinery, 2006. https://doi.org/10.1145/1121241.1121249.
[30] Scholtz, J. “Theory and Evaluation of Human Robot Interactions.” In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of The*, 10 pp. Big Island, HI, USA: IEEE, 2003. https://doi.org/10.1109/HICSS.2003.1174284.
[31] Kruijff-Korbayova, Ivana, Heriberto Cuayahuitl, Bernd Kiefer, Marc Schroder, Piero Cosi, Giulio Paci, Giacomo Sommavilla, et al. “Spoken Language Processing in a Conversational System for Child-Robot Interaction,” 2012, 8.
[32] Raymond, Christian. “Robust Tree-Structured Named Entities Recognition from Speech.” In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 8475–79, 2013. https://doi.org/10.1109/ICASSP.2013.6639319.
[33] Florian, Radu, Abe Ittycheriah, Hongyan Jing, and Tong Zhang. “Named Entity Recognition through Classifier Combination.” In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, 168–171. CONLL ’03. Edmonton, Canada: Association for Computational Linguistics, 2003. https://doi.org/10.3115/1119176.1119201.
[34] Scheerens, Jaap. *Educational Effectiveness and Ineffectiveness*. Dordrecht: Springer Netherlands, 2016. https://doi.org/10.1007/978-94-017-7459-8.
[35] Woodward, John D., Virginia, and Rand Corporation, eds. *Biometrics: A Look at Facial Recognition*. Documented Briefing / Rand Corporation, DB-396-PSJ. Santa Monica, Calif: RAND, 2003.
[36] Norris, Dennis, James M. McQueen, and Anne Cutler. “Merging Information in Speech Recognition: Feedback Is Never Necessary.” *Behavioral and Brain Sciences* 23, no. 3 (June 2000): 299–325. https://doi.org/10.1017/S0140525X00003241.
[37] Google Forms: Free Online Surveys for Personal Use. https://www.google.com/forms/about/
[38] C++ Programming Language, A Brief Description. https://www.cplusplus.com/info/description/
[39] Python Programming Language, Python Beginners Guide / Overview. https://wiki.python.org/moin/BeginnersGuide/Overview
[40] ALProxy Class Reference, http://doc.aldebaran.com/1-14/ref/libalcommon/classAL_1_1ALProxy.html#details
[41] ALMemory – NAO Software 1.14.5 Documentation, http://doc.aldebaran.com/1-14/naoqi/core/almemory.html#event-micro-event
[42] ALTextToSpeech – Aldebaran 2.5.11.14a documentation, http://doc.aldebaran.com/2-5/naoqi/audio/altexttospeech.html
[43] ALSpeechRecognition – Aldebaran 2.1.4.13 documentation, http://doc.aldebaran.com/2-1/naoqi/audio/alspeechrecognition.html
[44] Seetanah, Boopen. “The Economic Importance of Education: Evidence from Africa Using Dynamic Panel Data Analysis.” *Journal of Applied Economics* 12, no. 1 (May 1, 2009): 137–57. https://doi.org/10.1016/S1514-0326(09)60009-X.
[45] Johansson, Birgitta, Marie Fogelberg-Dahm, and Barbro Wadensten. “Evidence-Based Practice: The Importance of Education and Leadership.” *Journal of Nursing Management* 18, no. 1 (2010): 70–77. https://doi.org/10.1111/j.1365-2834.2009.01060.x.

[46] Yılmaz, Tonguç Utku. "Importance of Education in Organ Donation." *Experimental and Clinical Transplantation : Official Journal of the Middle East Society for Organ Transplantation*, 2011.
[47] Lindholm-Leary, Kathryn J. *Dual Language Education*. Multilingual Matters, 2001.
[48] Felder, Richard M., and Eunice R. Henriques. "Learning and Teaching Styles In Foreign and Second Language Education." *Foreign Language Annals* 28, no. 1 (March 1995): 21–31. https://doi.org/10.1111/j.1944-9720.1995.tb00767.x.
[49] Fok, Apple W. P., and Horace H. S. Ip. "Personalized Education: An Exploratory Study of Learning Pedagogies in Relation to Personalization Technologies." In *Advances in Web-Based Learning – ICWL 2004*, edited by Wenyin Liu, Yuanchun Shi, and Qing Li, 3143:407–15. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. https://doi.org/10.1007/978-3-540-27859-7_53.
[50] Waldeck, Jennifer H. "What Does 'Personalized Education' Mean for Faculty, and How Should It Serve Our Students?" *Communication Education* 55, no. 3 (July 2006): 345–52. https://doi.org/10.1080/03634520600748649.
[51] Chen, Chih-Ming, and Ching-Ju Chung. "Personalized Mobile English Vocabulary Learning System Based on Item Response Theory and Learning Memory Cycle." *Computers & Education* 51, no. 2 (September 1, 2008): 624–45. https://doi.org/10.1016/j.compedu.2007.06.011.
[52] Chen, C., S. Hsu, Y. Li, and C. Peng. "Personalized Intelligent M-Learning System for Supporting Effective English Learning." In *2006 IEEE International Conference on Systems, Man and Cybernetics*, 6:4898–4903, 2006. https://doi.org/10.1109/ICSMC.2006.385081.
[53] Petersen, Sobah Abbas, Jan-Kristian Markiewicz, and Sondre Skaug Bjørnebekk. "PERSONALIZED AND CONTEXTUALIZED LANGUAGE LEARNING: CHOOSE WHEN, WHERE AND WHAT." *Research and Practice in Technology Enhanced Learning* 04, no. 01 (March 2009): 33–60. https://doi.org/10.1142/S1793206809000635.
[54] Jurafsky, Dan. Speech & language processing. Pearson Education India, 2000.
[55] Brunelli, R., and T. Poggio. "Face Recognition: Features versus Templates." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, no. 10 (October 1993): 1042–52. https://doi.org/10.1109/34.254061.
[56] Parmar, Divyarajsinh N, and Brijesh B Mehta. "Face Recognition Methods & Applications" 4 (2013): 3.
[57] Video Camera – NAO Software 1.14.5 documentation, http://doc.aldebaran.com/1-14/family/robots/video_robot.html#robot-video
[58] ALFaceDetection – NAO Software 1.14.5 documentation, http://doc.aldebaran.com/1-14/naoqi/vision/alfacedetection.html
[59] Lucko Gunnar, and Rojas Eddy M. "Research Validation: Challenges and Opportunities in the Construction Domain." *Journal of Construction Engineering and Management* 136, no. 1 (January 1, 2010): 127–35. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000025.
[60] Morse, Janice M., Michael Barrett, Maria Mayan, Karin Olson, and Jude Spiers. "Verification Strategies for Establishing Reliability and Validity in Qualitative Research." *International Journal of Qualitative Methods* 1, no. 2 (June 1, 2002): 13–22. https://doi.org/10.1177/160940690200100202.

Appendix A

Personalized Language Education Robot Questionnaire

The results from these questionnaires will be used to evaluate the system's effectiveness. The information you will provide is completely anonymous and is expected to be unbiased.

* Required

1. Did you have any prior experience working or interacting with robots? *

   Mark only one oval.

   - Yes
   - No

2. How would you rate your knowledge of the Italian language prior to interacting with the robot? *

   Mark only one oval.

   - A lot (Native, proficient)
   - A little (Basic knowledge)
   - None

3. Do you think that the robot was able to effectively teach you the basics of the Italian language? *

   Mark only one oval.

   - Yes
   - No
   - Not sure

4. Do you think that the inclusion of movie related information and subjects during the education was a motivating factor for you? *

Mark only one oval.

- Yes
- No

5. If you answered "No" to the previous question, what other topic might motivate you?

6. In your own words, what more could be improved or added to the system to make it more motivating and effective when teaching a language? *