Indirect Speech Acts with AI Assistants

Špiranec, Marin

University of Zagreb

Faculty of Humanities and Social Sciences

Department of English

Marin Špiranec

Indirect Speech Acts with AI Assistants

Master's Thesis

Advisor: Dr. Marina Grubišić

Zagreb, 2022

Table of contents

Introduction

"Google assistant is ready to help, anytime, anywhere" (*Google Assistant on Your Phone*, n.d.), says the mission statement of Google's globally known AI-based virtual assistant that has become a part of the daily lives of many a user. Helping people with simple and tedious tasks like setting reminders, checking the weather or simply doing a random internet search, Google assistant boasts its capability with over 1 million possible actions (*Google Assistant on Your Phone*, n.d.). As virtual assistants make use of technologies like machine learning, speech recognition, dialog management, text-to-speech synthesis, data mining and analytics just to name a few (Sarikaya, 2017, p. 68), the sheer number of possibilities a virtual assistant can do is rapidly increasing. The fact that a virtual assistant is not limited to a phone or a computer anymore, but can be integrated into cars, wearables and homes, only expands its possibilities.

Other virtual assistants like Apple's Siri, Microsoft's Cortana or Amazon's Alexa have also taken the world by storm, estimating that by the year 2023, 8 billion virtual assistants will be in use worldwide (Brill, 2019, p. 1401). As artificial intelligence technology is getting more advanced, virtual assistants are becoming more able of mimicking natural human conversations and reading between the lines, as well as catching subtle nuances in speech. Metaphorical notions, abstract concepts and ideas are slowly settling into the virtual minds of digital assistants, so the need for research in that direction is increasing rapidly.

The goal of this paper is to test the degree of indirectness a virtual assistant can handle without losing the core intention behind the message. In order to test that, a theoretical framework from two different fields will be required, each covering one conversational partner, a human and a machine, respectively. The first part will be dealing with pragmatics, focusing mainly on Austin's speech act theory, i. e. indirect speech acts, and the second part with computer science with an emphasis on natural language processing. After that, the structure of the research is going to be presented, followed by the results and a discussion of the possible implications. The scope of this paper is going to be fairly limited in regards to the analysis, as the goal is not to present an extensive conversational analysis with the virtual assistant, but just to use some of its tools to highlight the scope and/or limitations of the assistant's comprehension[1] of intentions and indirectness.

[1] It is important to note that comprehension is not meant in its literal sense here, but rather the virtual assistant's ability to process speech and infer correct interpretations.

Part I - Pragmatics

Speech acts

In order to analyse a conversation with a virtual assistant, a first step would be to establish a theoretical framework that will help shed some light on what is exactly happening (at least from the perspective of the human speaker). Moreover, it is important to note before explaining the framework, in the case of this thesis, a conventional analysis presupposes that and is essentially applicable when the two conversation participants are human beings, instead one of them being a machine. This does not automatically throw most of the concepts and rules out the window, but it does alter them in a certain fashion. For this reason, concepts like Grice's maxims (1968) and felicity conditions (Yule, 1998, p. 50) will not play a decisive role in the framework for two reasons. Firstly, the maxims or felicity conditions would not relate to the virtual assistant itself, but the programmers of the assistant that determine the conversational principles the assistant depends on and to what degree. Secondly, the assistants take inputs at face value and with the help of their own principles, i. e. programming, they determine what the intention is, which means that they always assume the same circumstances, as it is their goal to always find what the user is looking for, even if it is nothing.

The most significant concept the discussion is going to revolve around is the speech act, so that will be the starting point of the framework. There are many definitions and iterations what exactly constitutes speech acts, but a commonly accepted and simple one would be that they are "actions performed via utterances," (Yule, 1998, p. 47) but this may be too vague for the purposes of this paper. To enrich the definition just a bit, Austin (1975) shows, that speech acts are made up of three distinct aspects, namely the locutionary, the illocutionary and the perlocutionary aspect (p. 103). The locutionary act constitutes the sole activity of making an utterance, i. e. producing a meaningful message in a certain language. The illocutionary act is realised by performing the locutionary act; it is the declarative aspect of the message, i. e. what is actually meant. Lastly, the perlocutionary act represent the effect that the utterance should have on the addressee. To better illustrate the aspects with an example, consider the following sentence 'Close the door,' where the locutionary aspect is realized by simply uttering the sentence, the illocutionary aspect is a request, as this is the intent behind the sentence, and the perlocutionary aspect is to persuade the addressee to (hopefully) close the door, although this could prove difficult considering the directness of the speaker. Before continuing onto a different concept, a few words should be said about the aspect with the biggest importance for this discussion, and generally most discussions in pragmatics, the illocutionary aspect. Searle (1965) states that performing illocutionary acts means to "engage in a rule-governed form of

behaviour" (p. 255). He distinguishes between two sets of rules that govern illocutionary acts, constitutive and regulative rules (p. 255). As their names suggest, constitutive rules establish an activity which is dependent of those rules and regulative rules manage activities which are independent of those regulative rules. Those rules make people understand how to, for example, make promises and what promises encompass, i. e. they constitute the performance of the speech act, as well as the way in which it should be performed. There is a small discrepancy, as for Searle (1965), the illocutionary act is the minimal unit of linguistic communication (p. 254), as opposed to the whole speech act. The rules, therefore, govern only this aspect of the speech act, but this will not play a significant role for the analysis. Lastly, as Strawson (1974) suggests, it is important to keep in mind the distinction between "act" and "force" when referring to the aspects of speech acts (p. 291). An act will be a certain act depending on the force it has, because the force constitutes an act. This should suffice as a rudimentary definition of speech acts, as widening the scope of mediums for speech acts to material activities (Wunderlich, 1984, p. 498) such as grimaces, shrugs and hand motions offer no purpose in this particular research of an interaction with a virtual assistant. The only applicable factor of the material world in the case of this paper's research is the human voice, which will also be considerably limited in the effect it can have on the assistants' interpretations, seeing that virtual assistant cannot distinguish between differences in prosody.

Now that a working definition for speech acts has been established, a few words can be said about their typology. Mey (2000) states, that there are two main ways in which speech acts can be classified. The first one, the "splitter" approach, encapsulates the idea that there could be as many speech acts as there are verbs that can denote different speech acts, which would be an enormously large number. The second one is known is the "lumper" approach, which is a name reserved for those who "lump" their speech acts into a few large categories. For the purposes of this paper, the latter approach shall be taken for the analysis, since it is both more convenient and digestible for the discussion. There are two notable classification types in the lumper approach, which should be mentioned here; that of Searle (1976) and Wunderlich (1984). Searle (1976) divides speech acts (or in his case illocutionary acts) into a five-part classification, according to the direction of fit, where "fit" shows the relation between the speaker's words and the world. The first class are representatives, which represent a state of affairs. Speech acts that fit into this class are statements, assertions, conclusions and descriptions (Yule, 1998, p. 53). As for the direction of fit, the words fit the world, as no change is the world is actually being made. The speaker just tries to say what they see in the world. The

second class are directives, which instruct the addressee to do something. Speech acts like commands, orders, request and suggestions (Yule, 1998, p. 54) belong to this class, which has a word-to-world direction of fit, as the speaker changes the affairs in the world by commanding the hearer. The third class are commissives, where the speaker commits themselves to do something. Promises, threats, refusals and pledges (Yule, 1998, p. 54) fit into this class, which has a world-to-word direction of fit. The fourth class is called expressives, which is used to express certain psychological states and emotions. Speech acts like congratulations, expressions of pleasure or sorrow or any other emotion (Yule, 1998, p. 53) fit into this class. As they are purely subjective expressions, they have no direction of fit. The last class is declarations, which change the world as they are uttered. Those speech acts are rooted in institutional power and can only be performed by individuals or entities that hold certain power in the world (Yule, 1998, p. 53), e. g. bosses when firing their employees by exclaiming "You're fired!" The direction of fit is both word-to-world and world-to-word, as the change/verbalisation happens simultaneously. The classification by Wunderlich (1984) is even more economical. He categorises speech acts according to the three main sentence types (p. 502). Assertives are speech acts that have a declarative structure. This category would encompass four of Searle's classes, except, funny enough, declarations. Directives are speech acts that correspond to the imperative structure and questions, lastly, correspond to the interrogative sentence structure. This classification can also be seen as the interplay between the three basic structural forms and the three basic communicative functions (Yule, 1998, p. 54).

Indirect speech acts

After having discussed speech acts in general, the next step is to specify a special kind of speech acts, which constitute the backbone of this paper, indirect speech acts. Yule (1998) offers a good starting point, saying that the distinction between a direct and an indirect speech act is constituted on the basis of structure (p. 55). The communicative function of a direct speech acts corresponds to the structure used to express it, i. e. posing a question with an interrogative sentence. Following this logic, the function of an indirect speech act does not correspond to its structure, so using a declarative sentence to ask a question would constitute an indirect speech act. According to Searle (1975), the structure and function of an indirect speech act do correlate to each other, and do not at the same time, i. e. the speaker means what they say, but they also mean something else (p. 266). There are two illocutionary forces in one speech act. The famous example "Can you pass the salt?" then asks the addressee about their ability to pass the salt and

by extension, if they are able, to actually pass it, both having the force of a question and a request at the same time. To further this claim, Levinson (1995) states that the illocutionary force is built into the sentence form (p. 263) and that, similar to Searle, this is its literal force, while the indirect force is inferred in addition to it (p. 264). Levinson (1995) calls this view the literal force hypothesis. This multiplicity of forces is also argued by Clark (1979), who highlights that as one of the properties of indirect speech acts (p. 200).

Knowing that the function and structure differ in indirect speech acts is not nearly enough to actually find and interpret one correctly. Consider the sentence "Do you have my papers?" This can be interpreted as a normal question, where the speaker just wants to know if his papers are in the hearer's possession. Perhaps they were supposed to be delivered to them (the hearer) and the speaker wants to know it that has been done. Consider this same sentence, but with the added information that the speaker when uttering the sentence holds out their open hand as if they are waiting for the hearer to give them something. The interpretation shifts now from a direct speech act representing a question into an indirect speech act representing an indirect request. According to Austin (1975), there are speech devices which can indicate the illocutionary force of speech acts (pp. 73 – 77), making an interpretation easier for the hearer. Those include:

A) Mood
B) Adverbs and adverbial phrases
C) Connecting particles
D) Tone of voice, cadence and emphasis
E) Accompaniments of the utterance
F) Circumstances of the utterance

Indicating devices A-C will be applicable in a conversation with the virtual assistant because they can be observed in written language, whereas devices D-F will have no purpose in helping the virtual assistant ascertain if there is an indirect speech act, as those devices depend on the physical world, especially device E and F. Yule (1998) also talks about those devices, which he calls IFIDs (illocutionary force indicating device) (p. 49), and Searle (1965) as well, calling them function-indicating devices (p. 257). Other speech act indicating circumstances like felicity conditions (Yule, 1998, p. 50) will not play a big role in the analysis, as indirectness will be the key factor to be tested and the only one which will vary.

Convention and Implicature

Austin (1975) states that the illocutionary act is "a conventional act; an act done as conforming to a convention" (p. 105). Searle conforms to a similar idea that the connection between an utterance and its illocutionary force is a matter of linguistic convention (qtd. in Asher & Lascarides, 2001, p. 186) and Wunderlich (1984) states that a speech act holds a certain degree of conventionality to it (p. 502), while Clark (1979) sees conventionality as being one of the six properties of indirect speech acts (p. 201). They all seem to agree that convention plays an important part with speech acts, but that poses the question, what kind of convention is meant by that?

Morgan (1978) distinguishes between two types of convention in that regard, that of language and that of usage (p. 242). Conventions of language give rise to the literal meanings of sentences, while conventions of usage dictate how those sentences are used, dependent on their purpose. Furthermore, he gives two approaches to explaining the conventionality of speech acts, the natural and the conventional (p. 245). He points out how the question "Can you pass the salt?" is interpreted as a request "naturally" because both participants adhere to Grice's maxims to get to the correct interpretation, and because this interpretation "feels" correct. It is not based on linguistic convention, but rational behaviour (p. 245). "Conventionally" speaking, this sentence is taken idiomatically, that is its former implicature became the literal meaning of the sentence. It cannot be interpreted as anything else than a request, as no native speaker of English would interpret the sentence as a question, wondering why someone would ask about their ability of passing objects around (except as a joke). According to Morgan (1978), this is "arbitrary, a matter of knowledge of language" (p. 246). In that same way, it becomes common sense and logical for speakers of English, that the sentence "Can you pass the salt" implies a request and not a question. However, Mey (2000) argues, that a distinction should be made here, because the implicit force in actual language use is not the same as logical implication (p. 99), so this implication is not rooted in a common, logical sense, but rather in a conversational sense. An entirely different phrase like "The salt has taken a walk" could have meant the same thing, had the phrase embedded itself into the conversational knowledge of a language under the right circumstances. A good example of implicature which relates to conversational knowledge of a language community is Morgan's (1978) short-circuited implicature (p. 250), which is an implicature that resides in the common knowledge of a given language community and is instantly recognizable for what it is intended, as it is with the sentence "Can you pass the salt?" Clark (1979) emphasizes, that this form of the sentence is more conventional and

idiomatic than a more indirect expression with the same meaning like “Are you able to pass the salt?” (p. 201).

A last note about implicature before moving on to natural language processing involves the different types of implicature one can have in a conversation. Grice (1968) distinguishes two types of implicature, conventional and conversational. With conventional implicature, “the conventional meaning of the words used will determine what is implicated, besides helping to determine what is said,” (p. 307) i. e. the implicatum is tied to the words themselves, rather than language use. To further elucidate this type of implicature on Grice’s example, “He is an Englishman; he is, therefore, brave,” (p. 307) the sentence implies that he is brave because he is an Englishman, due to the word “therefore.” Had the example gone like this: “He is an Englishman, but he is brave,” the implicature would be that he is brave, although he is an Englishman, which reverses the relationship between the two. Conversational implicature on the other hand does not focus on the conventional meaning of words, but how they are used in the context of a conversation. In order to understand that, Grice (1968) introduces the conversational principle, which characterizes the willingness to work together and adhere to a common purpose and direction of a conversation (p. 307). To adhere to it, he gives four maxims, which govern the success of cooperation in a conversation. The maxim of quantity concerns itself with the amount of information the speaker gives to the hearer, where giving to much or to little information than it is required in a given conversation would mark a violation of the maxim. The maxim of quality is dealing with truthfulness of the utterances the speaker makes. They should not be false or made with insufficient evidence. The maxim of relation instructs that the utterance should be relevant to the conversation taking place. If the speaker asks the hearer to pass the salt and the hearer answers by talking about spaceships, it would be considered a violation of the maxim. The last maxim is that of manner, which concerns itself with the way something is said. This involves not being obscure, ambiguous, wordy or downright chaotic in statements. Grice (1968) states, that maxims can be violated in different ways, unintentionally or even intentionally, in order to imply different things, e. g. intentionally talking about the weather when being asked an awkward question to imply one’s unwillingness to talk about a certain topic. Although Grice discusses conversational implicature much more broadly than conventional, the latter may prove more interesting in a conversation with an AI assistant, because of the more straightforward nature of the implicature.

Part II – Computer Science

Natural language processing

A linguistic framework paints the picture for a conversation with a virtual assistant only halfway. The human side of communication has been made clear, but the virtual assistant still remains a mystery. Concepts like indirect speech acts or conversational implicature do not mean much for its algorithm while it decodes a message, so a brief overview of natural language processing, the main process that helps the virtual assistant comprehend a message, will be given. Considering natural language processing is a highly specialised and broad field, only the most basic concepts and stages of NLP will be explained here, which will offer the necessary context of how a virtual assistant generally interprets a sentence.

A first step would be a definition for natural language processing. NLP is the "study of mathematical and computational modelling of various aspects of language and the development of a wide range of systems" (Joshi, 1993, p. 393). As Arnold and Tilton (2015) add, NLP "mimics the complex process by which humans parse and interpret language" (p. 131). This is by no means an easy feat, especially because of the properties of natural language like its "inherent ambiguity and […] strict connection to semantics" (Ferilli, 2011, p. 199), which can pose a nearly uncrossable hurdle for a machine that operates on a strictly formal and logical level. For that reason, special resources, that are made of "large databases that represent and encode morphologic, lexical, syntactic and semantic information" (Ferilli, 2011, p. 200), are used in order to make that jump and decode and encode language similar to a human. Those databases are dictionaries, corpora, thesauri and ontologies (Ferilli, 2011, p. 200), just to name a few. Moreover, there are a plethora of natural language processing techniques that exist to make natural language understanding more accurate and efficient. Zhao (2022) mentions 57 different techniques (p. 1) which are being used for decoding language, making the whole process even more complex. To name and explain each technique separately would need its own book, so to at least get a general overview, the most notable and overarching stages are going to be elucidated. The reason why the following stages have been chosen is for the reason that most AI applications make use of those stages. This does not mean they use them all and/or to the same extent, but those stages represent a logical timeline of how machine interpretation generally works.

Stages of NLP

The following stages are going to be presented in the order they are carried out, although a definite order cannot be stated here, as different authors opt for different orders of various stages. For example, while Ferilli (2011) and Arnold and Tilton (2015) place parts of speech tagging after lemmatization, Aminzadeh (2022) places it before lemmatization. As one stage does not affect the other, i. e. some stages are carried out independent of the other or at the same time, the order can vary. Furthermore, to better illustrate what is happening, an example is going to be shown after going through every stage.

The first stage is called sentence segmentation (Aminzadeh, 2022, p. 171). In this stage, the input, i. e. text, is cut up into individual sentences in order to ready the text for the next stage. Capital letters, punctuation marks and other indicators are used to determine the start and end of a sentence. In the case of the virtual assistant, sentence segmentation is not of much importance, as the conversation follows one sentence per turn from the user. Consequently, there is no need to segment it up, as there is only one sentence. The example sentence is going to be "The frog carried the scorpion over the river". The second stage is tokenization (Ferilli, 2011, p. 207). This encompasses splitting the sentence up into elementary components called tokens and putting them into specialised categories whose comprehensiveness can vary. The example sentence looks like this now: "the, frog, carried, the, scorpion, over, the, river". After that comes parts of speech tagging (Ferilli, 2011, p. 213), which determines the grammatical function of the tokens. In this stage, the result is going to be three determiners, two nouns, a verb and a preposition. After every token has been marked with a parts of speech tag, lemmatization or stemming (Ferilli, 2011, p. 210) happens. Both processes simplify the tokens for further processing, but in slightly different ways. Lemmatization takes the token and strips it of inflectional changes to get to its lemma, i. e. the form that appears as an entry in a dictionary and stands for all other forms of that lexeme (Brown & Miller, 2013, pp. 3–2). The result would then look like this: "The, frog, carry, the, scorpion, over, the, river". According to Ferilli (2011) though, stemming is a more common method of simplifying tokens (p. 210). In stemming, the inflectional changes are chopped off to get to the stem of a word, i. e. the form which affixes attach to. The next stage is called stop-word removal (Aminzadeh, 2022, p. 173). Stop words encompass all words that appear fairly frequently in texts of a specific language and that do not carry a lot of meaning. They are subsequently removed in order to simplify the amount of data the software has to process. It depends on what a certain software includes in its list of stop words, but most commonly they include articles, adverbs, conjunctions, pronouns, prepositions and auxiliary verbs (Ferilli, 2011, p. 210). The example sentence would now be "frog, carry,

scorpion, (over), river". After that follows dependency parsing (Aminzadeh, 2022, p. 174), where the structure of the sentence is established. The software checks how each word is related to the other words in the sentence and recreates the relationship. One way the example could look like is this: the frog = noun phrase; carry the scorpion over the river = verb phrase. The verb phrase would then further be segmented into even smaller constituent phrases, until all relevant relationships have been made clear. Then, the resulting phrases go through named entity recognition (Arnold & Tilton, 2015, p. 143). This process identifies tokens and categorises them into broad semantic groups like person, animal, institution etc. In the sentence example, "frog" and "scorpion" would most likely be put into the animal category, and river into a category like "natural bodies of water". The last stage is called coreference resolution (Aminzadeh, 2022, p. 175), which is a process that assesses which expression refers to what. Arnold and Tilton (2015) state, that this process establishes "semantic relationships between tokens, that may be far away within a given corpus." (p. 145) To give a simple example, in the sentences "The frog carried the scorpion over the river. They both ended up sinking," the software will know that "they" refers to the frog and scorpion.

All of the aforementioned stages belong to various levels of analysis the software has to go through in order to interpret a sentence. According to Zhao (2022), stemming, lemmatization and stop word removal belong to the morphological level of analysis; sentence segmentation, parts of speech tagging, dependency parsing and tokenization to the syntactic level of analysis. Named entity recognition is a part of the semantic level of analysis and coreference resolution a part of the discourse level of analysis.

Research

Method

Because the technology behind every virtual assistant varies, as do their expertise and abilities, and therefore the level of their language comprehension, it is important to determine which virtual assistant is going to be tested for the purposes of this paper's research.

Berdasco et al. (2019) compared the four aforementioned virtual assistants in a series of various tasks to test which virtual assistant is more correct in their answers and which had the more natural responses. Google assistant was the best in correctness and second in naturalness, only bested by Alexa by a narrow margin. Cortana's and Siri's performance was significantly lower, as Berdasco et al. (2019) state, because their focus lies more on everyday problems and conversation rather than completing requests. Another reason, why Google assistant is the best candidate for testing indirect speech acts, is the fact that it is the most recent virtual assistant that has been released. Siri launched in 2011, Cortana in 2013, Alexa in 2014, and last but not least, Google Assistant in 2016 (Hoy, 2018, p. 82). As this paper focuses on indirect speech acts with requests, and as Google's virtual assistant has the most recent technology, Google assistant has been chosen as a conversation partner for testing.

Virtual assistants primarily have two functions, completing simple tasks and requests such as opening apps, calling somebody or keeping a schedule, and answering questions. The advertised and "correct" way of making requests to virtual assistants is being blatantly direct and using the imperative form to essentially order the assistant to do something. This is considered to be the best way of communicating with the assistant, as its algorithm is built to most easily understand commands like "play [song name]; open [app]; call [contact]." This bare lexical minimum required to complete a task gives the virtual assistants its practicality and ease of use, which are becoming more sophisticated as new technologies emerge.

But that is only true when the speaker's intention is in line with what is said, in other words when the speaker uses direct speech acts to make a request or ask a question. The question then remains open, how virtual assistants are going to react when confronted with a sentence structure that differs from the speaker's intention, i. e. indirect speech acts. This research will take a look at indirect speech acts concerning requests that are made by using either an interrogative or declarative sentence form. The main goal of this discussion is to determine the degree of indirectness, where the virtual assistant is still going to understand the request and successfully complete it. The following five commands will be used to analyse the degree of

indirectness when turned into indirect speech acts, i. e. into a declarative and interrogative sentence conveying a request:

a) Play a song by [artist].
b) Open [app].
c) Show me the weather.
d) Set a reminder.
e) Define [word].

The variable in […] parentheses can vary as to further test if specific names or concepts impact the level of comprehension of the sentence by the virtual assistant. Each of the five aforementioned commands is going to get multiple forms with an interrogative structure and a declarative structure, and they are going to be viewed as two opposing categories to see if any significant changes in sentence comprehension can be observed depending on the sentence structure.

Discussion and Results

Play a song by [artist]

When the request was made with an interrogative structure, the virtual assistant has shown mostly a correct understanding of what was meant, that is the assistant started playing anything that was made by the [artist] most of the time. Interrogative sentences that concern themselves with the ability of the assistant to do an action, in this case, play a song, were almost all interpreted correctly by the virtual assistant. Sentences of that type include:

(1) Can/could you play a song by [artist]?
(2) Are you able to play a song by [artist]?
(3) Do you have the ability to play a song by [artist]?
(4) Is it within the power of your ability to play a song by [artist]?

The [artist] was replaced by world renowned singers and bands like *ACDC,* Madonna, Ariana Grande, Iron Maiden as well as some fewer known artists like Will Evans and the Croatian group *Detour* that sport a smaller amount of popularity. In examples (1), (2) and (3) the assistant would start playing a song on Spotify of the inserted artist in all cases except with the group *Detour*. Although the proposition *by* is present in each sentence which shows the relationship between the song and artist and elucidates which is which, the assistant would sometimes play something by the band *Detour*, in some cases it would understand the word

detour as referring to the song name and play the song *The Detour* by the band *The Who*, while other times it would do a google search on the song title *Detour* by different artists. The latter result can also be achieved by simply saying "Detour song" to the virtual assistant. This seemingly arbitrary occurrence of different solutions to the request shows the shifting comprehension of the virtual assistant, which in most cases comprehends the underlying intention of the sentence according to certain keywords and plays a song, while other times disregards and equals it to the literal meaning of just showing the ability to find the song or artist name and stopping there. A possible explanation for this can be found with named entity recognition. The virtual assistant could sometimes categorize *Detour* as a band name and other times as a song name. This category shifting is likely due to *Detour* not appearing as frequently in search results as, for example, *ACDC*.

Another interesting note is that in example (4) the assistant mostly just does a google search about the artist or song and does not play anything. This occurrence is mostly fixed by simply omitting parts of the sentence. In this case the illocutionary force is correctly identified by omitting *the power of* of example (4). This point will be further discussed with other examples, but it seems that the virtual assistant has more difficulties in interpreting sentences which carry more constituents, albeit attributes or adjuncts, than sentences which hold only the essential constituents. This could be an indicator that the virtual assistant, in the process of dependency parsing, incorrectly infers the relationship between words and consequently does not understand the implied request.

The next set of interrogative examples is made up of sentences that show the virtual assistant's intention or will to perform an action:

(5) Will/would/won't you play a song by [artist]?
(6) Aren't you going to play a song by [artist]?
(7) Aren't you going to play an awesome song by a cool artist like [artist]?

In examples (5) and (6) the results were similar as they were with the previous set of examples. For the most part, the virtual assistant did not have any difficulties in interpreting the correct illocutionary force behind the sentences, except for example (7). The assistant would either not understand the sentence at all and disregard it or do a google search of said artist. It is important to note that the only difference between example (6) and (7) are the added adjectives *awesome* and *cool*, which seem to serve as the catalyst for a new interpretation, disregarding the underlying request.

Another set of examples are sentences concerning the virtual assistant's willingness to perform an action:

(8) Would you be willing to play a song by [artist]?

(9) Do you want to play a song by [artist]?

(10) Would you mind if you played a song by [artist]?

(11) Would it be convenient for you if you played a song by [artist]?

(12) Would it be too much trouble for you if you played a song by [artist]?

This set shed light on another interesting aspect of sentence comprehension by virtual assistants. Examples (8) and (9) were interpreted by the assistant correctly, whereas examples (10) to (12) were interpreted incorrectly, resulting in more google searches on the artists and their song titles. At first, it would seem that the assistant failed to infer the illocutionary force because of the length of the sentences or the number of constituents, having too many elements to infer their relationship, but the real reason lies with the *-ed* ending of the verb *play.* Changing *played* into *play* immediately fixed the interpretation of the assistant to the correct one. The -ed ending does not automatically change the way the assistant understands a sentence, which can be proven by simply saying "Played [artist]" and it starts playing a song by the artist of choice. A possible explanation could be the large number of constituents the assistant has to analyse and connect in relation to another verb in the sentence it chooses, as it seems to disregard the verb *played* because of its ending, consequently getting "confused" in the process. This could be the case if the assistant does not prioritise lemmatization or stemming and simply just shifts its focus to a verb which is in its base form.

The last set of interrogative examples do not fall into any category in particular, but are analysed here as they offer more interesting insights into the way the virtual assistant comprehends the requests.

(13) Would it not be nice to play something by [artist] after such a long time?

(14) What do you think about playing [artist] right now?

Example (13) always results in the assistant doing a google search in two ways. The assistant either searches random songs by the inserted [artist] or it finds the songs which contain some of the words found in the example, like *Something* by *the Beatles*. This misinterpretation could also the result of the added adverbial *after such a long time*. After omitting it, the assistant plays a song of the chosen artist, correctly identifying the illocutionary force of the sentence. Even if the assistant removed stop words like *after*, *such* and *a*, the phrase *long time* seems to

affect how the assistant interprets *something*, changing it from song name to simply a placeholder for a song by the selected artist. Example (14) offers a very interesting response in some cases, i. e. with some artists. When the artist was *ACDC* the response of the assistant was:

What an interesting topic. What do you want to know about ACDC.

A possible explanation for this interpretation could be the favouring of the present simple verb *think* over the *-ing* form of *play* due to the imperative nature of communication with the virtual assistant, as well as a possible favouring of verbs in their base form, viewing it as the main idea behind the sentence. The assistant completely disregards *play* in the sentence and interprets it as a conversational inquiry into the band. The last interrogative examples concern themselves with verbs sharing the same intention as play when being used in a request:

(15) Would it not be cool to listen to some [artist] after such a long time?
(16) Do you mind putting on some [artist] music?
(17) Would you blast some [artist]?

In examples (15) to (17), the virtual assistant mostly just does a google search on the keywords concerning the inserted [artist], but there are some instances of correct interpretation. Example (16) shows the most positive results in correctly identifying the illocutionary force. This example offers the most direct way of making a request and uses a more conventionalised verb for conveying the same meaning, without using additional adverbials that seem to confuse the assistant.

With interrogative structures, the assistant understands and correctly interprets what is meant most of the time. Even when it does not automatically play a song, but does a google search instead, the virtual assistant still holds onto the logic and general idea of the illocutionary force. Requests that have a declarative structure have raised slightly different results than their interrogative counterparts. Sentences that were fairly simple in structure and did not use any additional constituents outside of the essential ones were interpreted correctly by the assistant. Those include examples like:

(18) You can/could play [artist].
(19) I want you to play [artist].
(20) I hope you'll play [artist].
(21) I wish you would play [artist].
(22) You should/ought to play [artist].

(23) You had better play [artist].

The virtual assistant would always interpret what is meant correctly, no matter what artist would be inserted into the slot. When the present simple verb (*play*) was replaced with its *-ed* variant, the assistant would again just do google searches on songs of inserted artists. The same result was achieved with sentences that were made a bit more complex than the above examples like:

(24) I would really appreciate it if you played / would play [artist].
(25) I would be most grateful if you played / would play [artist].
(26) I'd be very much obliged if you played / would play [artist].

With constructions using *appreciate*, *grateful* and *obliged* in combination with the verb *played*, the assistant would often do a google search on how to use certain phrases like *I would appreciate it / I would be grateful,* instead of looking up songs of the inserted artist. Due to the highly conventionalised first part of the sentence used to express politeness, the assistant could have shifted the interpretation in wanting to learn how to use those phrases, rather than playing a song. The second part of the sentence, which carries the main idea of what is meant, is being disregarded in favour of the first part. This could be due to the frequency and context those phrases appear in the resources. The more people use those constructions in the context of wanting to learn how to use those phrases, the more the assistant may favour this interpretation.

Sentences that do not have the verb *play* or that replace it with others that carry the same meaning did not work most of the time for the virtual assistant. Examples like that include:

(27) Some music by [artist] would really lighten the mood right now.
(28) I haven't listened to [artist] in a very long time.
(29) I could really go for some [artist] right now.
(30) I think [artist] is our jam now.
(31) I would feel better if you put on song by [artist] right now.

All of the above examples would normally lead the hearer to infer that the speaker is not simply stating a fact of having a wish to play a certain artist, but that the speaker is actually making a request indirectly. The virtual assistant infers the artist and does a google search about them at most. It does not connect the synonymous phrases with the context of the underlying message of the sentence. The closest example that would fit the sentences above and have the assistant correctly identify the illocutionary force is the following:

(32) I sure wish to listen to [artist] right now.

The literal illocutionary force of the sentence is still a pure factual one by expressing merely the wish of listening to a specific artist and nothing more. Only after correctly identifying the implied illocutionary force of requesting to actually play the inserted artist does the sentence make sense in a communicative situation.

Moving on from the command of "Play a song by [artist]," the next four commands were analysed using the same type of sentences used in the examples above. As to not repetitively write the same sentences in a different context, only insightful and interesting sentences will be presented here for the following command. The second command fulfils one of the most basic and simple functions a virtual assistant can do, and that is opening an app.

Open [app]

Instagram was chosen for the slot of the app as it is one of the more notable and popular ones and is usually part of the default apps present when buying a new phone. Starting off with requests of interrogative structure, simple sentences were for the most part interpreted correctly by the assistant and it would immediately just open the app. Drastically different interpretations started to arise as the sentences became longer and more complex.

(33) Are you able to open Instagram?

(34) Have you got the ability to open Instagram?

(35) Won't you open Instagram?

(36) Why not open Instagram?

All of the above-mentioned examples have elicited a similar response by the virtual assistant, and that is its assumption of the speaker having a question about the functionality of Instagram. In example (33), instead of opening the app, the assistant starts giving a detailed explanation on how to open Instagram, and ultimately any app. The indirect request was interpreted as a question of "How to open Instagram?" that is the assistant disregarded the first part of the sentence, which shows the speaker asking about the ability of the hearer to perform an action, and instead just focused on the second half and automatically filled the first part of the sentence, turning it into a simple question. It is possible to interpret the sentence in that manner for the result to be an explanation of how to open an app; for example, if the hearer (i. e. the virtual assistant) infers that the speaker asks the question because they themselves do not know how to open an app, in this case Instagram, they could interpret that the speaker is

generally interested in the act of opening apps, not Instagram specifically. But this is in the virtual assistant's case rather unlikely. In example (34), the assistant does a google search on Instagram related questions, like how to change a profile, how to post a photo etc., whereas both (35) and (36) lead the assistant to give an explanation on how to restart Instagram. This is interesting, because the virtual assistant takes the constituents in (35) as *will, not, open* and *Instagram*, and interprets it as a request, although the structure is interrogative or declarative if viewed in isolation. The assistant comprehends that more was meant than said, and offers a solution on how to fix the inability to open Instagram, namely by restarting it. Example (35) could be also viewed in terms of broken English for the assistant and turned into a sentence like "Why won't Instagram open?" or "Why can't Instagram be opened?" so the assistant's interpretation would make sense in that regard.

Another interesting instance was with the examples containing *like*:

(37) Aren't you going to open a cool app like Instagram?
(38) Would you be willing to open a cool app like Instagram?

Those examples resulted in the assistant conducting a google search about apps which are similar to Instagram. It interpreted the sentences as simply being *app like Instagram.* A possible explanation for this interpretation could be the conventionality of the phrase *app like,* as it is fairly commonly used in the functionality of Google's Play store, where after an app is downloaded, similar apps are displayed.

When it comes to request made with a declarative structure, the assistant came across mainly one problem.

(39) I would appreciate it if you opened Instagram.
(40) I would be most grateful if you could open Instagram.
(41) It wouldn't hurt if you opened Instagram.

Examples (39) to (41) would all result in the assistant searching for Instagram posts that contained the first half of the sentence, mainly focusing on *appreciate, grateful* and *hurt* as hashtags in the app. This case differs from the first command, where the interpretation mismatch arose because of the *-ed* ending of the verb. As Instagram heavily relies on keywords (hashtags) for finding certain posts, it stands as a valid possibility the assistant could follow the same logic when having requests dealing with Instagram.

Increasing the degree of indirectness or complexity of the sentence for the assistant results in completely disregarding the possibility of opening Instagram, and instead of either disregarding the request completely or doing random google searches vaguely relating to the words in the sentences. Examples like that include:

(42) I haven't been on Instagram today yet.

(43) It would be better for if you went and opened Instagram right now.

(44) It would be awesome to go to Instagram.

Show me the weather

This command was met with the most positive response when it comes to the correct interpretation of the request. All of the sentences with an interrogative structure yielded the correct result of the assistant showing the weather forecast for the general area, except for one.

(45) Is it within the power of your ability to show me the weather?

The assistant seemed to have taken the word power as a keyword, which resulted in a google search about the possibility and chance of having the power to control the weather. Save for this example, the assistant shows the forecast even within a correct timeframe if a temporal adverbial like *tomorrow*, *in three days* or *now* is used. When asked about a specific weather condition, the assistant would also start off with addressing the specific condition first and then continuing on with the general forecast. It is interesting to note, that in those cases the priority of stop words has changed as adverbs play an important part in weather forecasts.

Sentences with a declarative structure were also met with a streak of correct interpretations by the virtual assistant, yielding only three interesting occurrences of getting to the wrong interpretation.

(46) I still don't know the weather.

(47) I wish you wouldn't give me the news, but the weather instead.

(48) I wish I knew the weather.

Example (46), after multiple tries, results always in a google search of the song *I don't know what the weather will be* by Laura Mvula. The virtual assistant seems to take the whole sentence as one fixed phrase and equates it to the song name, never entertaining the possibility of doing a different interpretation of the sentence. In example (47), the assistant either did not understand the request and disregarded it or did a google search on practicing grammar. It is interesting to

observe here how the assistant possibly took the structure of the sentence and connected the very structure with the ones used to practice grammar. By simply using a certain structure, no matter the content, the assistant would comprehend the user's sentence as wanting to practice grammar. This example shows how sometimes the assistant can favour certain structures and keywords to the underlying message of what is meant and disregarding it completely. Example (48) seems as a fairly simple one to understand at first, but the assistant has a variety of different interpretations which it can use to make sense of this sentence. In some instances, it disregarded the whole sentence as being invalid, while other times it did a google search on songs about the weather, and sometimes on weather phenomena. An interesting note is that it would underline the word *knew* when disregarding the request. If it were replaced with the present simple verb *know*, the assistant interprets the sentence as intended and gives a weather report. Stemming or lemmatization does not seem to carry much priority in requests like that as a difference should not be noted if that was the case.

Set a reminder

This command was also mostly understood in most types of sentences with an interrogative structure by the assistant. As long as the keyword *reminder* or *remind* would appear in the sentence, the assistant would infer the illocutionary force behind the questions and most of the time opt for setting a reminder. This is evident by simply saying *remind?* to the assistant, which results in a reminder being set. However, there were a few cases where the assistant did not interpret what is meant correctly. The following examples confused the assistant:

(49) Would it be too much trouble for you if you set a reminder?

(50) Why not set a reminder?

In both cases the response of the assistant was "Sorry, I don't understand," prompting the user to either repeat the request or ask about the functionality of the assistant. It is intriguing how the structure of (49) works with other commands but not with this one specifically. It contains the main keyword *reminder* paired with the verb *set*, and yet does not even offer a google search as with other examples, but the assistant simply does not understand the input. Only after omitting *too much* and *for you* does the sentence get correctly interpreted by the assistant, and ultimately understood. It seems that it varies from request to request and their keywords, which NLP techniques are prioritised, as dependency parsing could connect words differently based on this, albeit having the same adverbials and prepositional phrases. Example

(50) could be interpreted as the previous one with the same structure, where the interpretation could go as wanting to know the reason why a reminder cannot be set, and the possible lack of understanding could arise from the usual impossibility of that happening, thus not having a solution for that problem. Favouring of a certain interpretation could also be due to frequency of certain user inputs, as the more frequent inputs would be a go-to for the assistant.

Requests with a declarative structure are also well understood, but not without a few interesting cases:

(51) I would be very much obliged if you set a reminder.

(52) It wouldn't hurt if you set a reminder.

(53) I hope I don't forget to buy eggs tomorrow.

(54) I sure hope someone could remind me that I have an appointment tomorrow.

Examples (51) and (52) fulfil their role and do elicit a response of the assistant setting a reminder, but the assistant automatically fills in the text of the reminder with the first half of the respective sentences. So a reminder with the text *I would be very much obliged* and another one with the text *wouldn't hurt* are made. It is unclear as to why the assistant fills in the text explicitly with these sentences, as other ones of similar structure set a reminder without the text. A possible explanation would be because of the *if* in the sentence. Through stop word removal, the assistant might interpret the sentence the following way: "Set a reminder: it wouldn't hurt". Example (53) elicits a google search on the use of *future simple*, which is most likely motivated by the keywords *forget* and *tomorrow*, although the leap in logic is humorous. The first results on exercises for *future simple* also contain the part *to buy*, which further establishes this interpretation. The last example (54) is not mentioned here because of a false interpretation, but rather that it works despite the length of the sentence. Most other commands do not work if they were packed in a large sentence like that, but the possible reason this sentence works is because of two keywords, one being *remind*, and the other one being *appointment*, which solidifies what is meant in the sentence, overshadowing the rest of the added content. This can be tested by simply saying *appointment*, prompting the assistant to set a reminder, showing which relationships between words are most prevalent for a correct interpretation with the virtual assistant.

Define [word]

The last command was chosen because Google assistant, when asked about its functionality, advertises how it can define a word and prompts the user to try it. Out of the five commands, defining a word has proven to be the most difficult task for the virtual assistant to understand. The word that was chosen to be defined by the assistant was *poignant*. The interpretation of the sentence was considered correct if the assistant started explicitly defining via its voice function what *poignant* means. Google searches where the results led to a dictionary containing the word definition were not considered correct because this could have been the result of simply tagging keywords.

Interrogative structured requests worked without problems either with sentences concerning the assistant's ability to do an action, that is sentences with can/could as with examples (1) to (4), or shorter sentences expressing the assistant's willingness to do an action, like examples (5) to (7). All other examples were met either with google searches on the word *poignant* or showing synonyms and antonyms of *poignant*. An example that was particularly interesting is the following:

(55) Would it be convenient for you if you defined poignant?

Following this request, the first result the virtual assistant came up with was the definition of *convenient*, followed by the definition of *poignant*. The logic for this one is found in the prompts of the assistant after doing the request. The assistant suggests similar requests and questions that relate to the user's original one. The prompt reads:

Would it be convenient for you meaning?

This was the interpretation the assistant opted for in this instance, ignoring the second word which was actually the target of the sentence. *Defined* was recognised but not in relation to *poignant*, leading to the wrong interpretation.

Declarative structured requests faced similar problems as the interrogative ones. Shorter and simpler sentences expressing the sole fact that the assistant is able to define *poignant*, led to the assistant actually defining it. It was surprising how some sentences that looked as just an expression of opinion elicited the correct interpretation for the assistant, like:

(56) I think I don't know what poignant means.

(57) I don't think poignant is a word.

(58) Poignant is a word I don't know the meaning of.

(59) I never learned what poignant meant.

The assistant was surprisingly accurate in responding to every one of those examples with the correct interpretation of what is meant and started explaining the word. This could be due to the fact that all sentences are relatively simple and contain important keywords such as *poignant, means, meaning, word* and *learned*, which simplified determining the relationship between the words and correctly determining the intent.

Conclusion

The goal of this paper was to identify the degree of indirectness a virtual assistant could comprehend using indirect requests as the main point of the analysis. The virtual assistant that was chosen for an analysis was Google assistant. Employing traditional pragmatics, mainly Austin's speech acts (1975), and contemporary literature on natural language processing, a framework was constructed to offer the necessary tools to put an interaction with a virtual assistant into context. The interaction consisted of five frequent commands that were turned into indirect requests, each command getting a number of different forms with varying degrees of indirectness. In a normal conversational analysis (one with two human speakers), all of the reviewed concepts could be employed to interpret the flow of a conversation and point out how an intention is recognized and what is implicated. The same did not generally apply to a virtual assistant.

Speech devices argued by Austin (1975), Searle (1965) and Yule (1998), that usually indicate what the illocutionary force behind an utterance could be, did not seem to have helped the virtual assistant interpret the sentence. Mood, adverbs and connecting particles, the devices thought the assistant will take into consideration, did not appear to do anything for it, as Google assistant interpreted the user inquiries either as a pure request or a question, without reading into the added distinctions. The assistant even may have removed phrases like that entirely with stop word removal, as they did not carry the main point of the sentence. Grice's maxims (1968) and the cooperative principle also do not give any interpretational insights for the virtual assistant, inasmuch that Google assistant always takes every user inquiry as being fully in accordance with the cooperative principle. It does not seem to consider the possibility of a maxim being flouted, which in the case of the maxims of quantity and manner does happen in order to test the actual degree of indirectness, incorporating a certain degree of wordiness and vagueness to make a request more indirect. Examples that were more wordy or ambiguous in their phrasing were often interpreted incorrectly by the assistant. This indicates that conversational implicature does not play a significant role for the assistant's interpretation, but conventional implicature does. The conventional meanings of the words themselves are used, after they are separated into tokens, to infer the intent of the user inquiry, which falls in line with the principles of natural language processing.

To further elaborate, a concept that seemed to have a big impact on the interpretations was convention. Google Assistant was able to infer certain intentions correctly, which could be attributed to the conventionality of the expression used. As more conventional expressions can

be used to say more than what was meant, the virtual assistant had a higher chance of interpreting indirect intention from examples with conventional expressions. This does not mean that the assistant understands what is conventional and what is not, but it stands to reason that more conventional expressions are listed in the resources the assistant uses to simplify the processing of requests the user gives. This would fall in line with the conventional approach of interpreting speech acts argued by Morgan (1978). The frequency of user inputs in a certain context and with certain constructions seems to play a part in the favouring of certain interpretations, as do certain keywords being used for specific requests. The other side of interpretation lies with NLP.

It is evident from the results that due to interpretation discrepancies in some natural language processing stages like stemming/lemmatization, dependency parsing and named entity recognition, the virtual assistant inferred some requests incorrectly, either not recognising the indirect speech act or recognizing it, but with a different implicature. Sentences that were fairly simple in structure were generally interpreted correctly, which could suggest that the number of constituents is one of the most prominent factors for a correct interpretation. In examples where two words were semantically connected, but were separated by adverbials or prepositional phrases, Google assistant often opted for a wrong interpretation, incorrectly connecting the relations between certain words. Another significant factor could be stop word removal, as some examples yielded wrong interpretation due to some words being ignored, as if they were deemed by the virtual assistant as unnecessary for the intention of the message. The difference in interpretation due to inflectional endings could be because of lack of stemming/lemmatization during the processing of the request, as a notable correlation between correct interpretations and verb forms without inflectional endings was observed. A possible reason is the favouring of verbs in their base form if stemming/lemmatization is not prioritised.

It is clear that to further this discussion, a more extensive computational framework and knowledge is needed to pinpoint the exact processes that lead to certain interpretations and no clear line can presently be drawn as to the degree of indirectness the virtual assistant can understand without relying heavily on speculation. As semantics and pragmatics deal with the deepest and most complex aspects of analysis that a virtual assistant has to undertake, a more suitable starting point for this topic could be from the syntactic level because the results indicated that dependency relations played a big part in determining the intent behind the requests. This paper offered an insight into which processes might influence the determination of intent in indirect requests and where in the processing line an extensive analysis could offer

more in order to determine the degree of indirectness and a gain a better understanding of how a virtual assistant could comprehend natural speech better. It will be exciting to see what new technologies hold in store in terms of AI-based natural speech generation, since in 2018 Google has announced its new AI project called Google Duplex, a "new technology for conducting natural conversations to carry out 'real world' tasks over the phone" (*Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone*, 2018).

Works cited

Aminzadeh, F., Temizel, C., & Hajizadeh, Y. (2022). *Artificial Intelligence and Data Analytics for Energy Exploration and Production*. Wiley.

Arnold, T., & Tilton, L. (2015). *Humanities Data in R: Exploring Networks, Geospatial Data, Images, and Text (Quantitative Methods in the Humanities and Social Sciences)* (1st ed. 2015). Springer.

Asher, N., & Lascarides A. (2001). Indirect Speech Acts. *Synthese*. pp. 183-228. https://www.jstor.org/stable/20117151

Austin, J. L. (1975). *How To Do Things With Words*. Harvard University Press

Berdasco, A., López, G., Diaz, I., Quesada, L., & Guerrero, L. A. (2019). User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. *13th International Conference on Ubiquitous Computing and Ambient Intelligence UCAmI 2019*. https://doi.org/10.3390/proceedings2019031051

Brill, T. M., Munoz, L., & Miller, R. J. (2019). Siri, Alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications. *Journal of Marketing Management*, *35*(15–16), 1401–1436. https://doi.org/10.1080/0267257x.2019.1687571

Brown, K., & Miller, J. (2013). Dictionary. In *The Cambridge Dictionary of Linguistics* (pp. 3-2). Cambridge: Cambridge University Press.

Clark, H. H. (1979). "Responding to Indirect Speech Acts." In *Pragmatics: A Reader* (1st ed.) (pp. 199-230). Oxford University Press.

Ferilli, S. (2011). *Automatic Digital Document Processing and Management: Problems, Algorithms and Techniques (Advances in Computer Vision and Pattern Recognition*) (2011th ed.). Springer.

Google Assistant on your phone. (n.d.). Assistant. Retrieved September 17, 2022, from https://assistant.google.com/platforms/phones/

Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. (2018, May 8). Google AI Blog. Retrieved October 2, 2022, from https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

Grice, H. P. (1968). "Logic and Conversation". In *Pragmatics: A Reader* (1st ed.) (pp. 305-315). Oxford University Press.

Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, *37*(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391

Joshi, A. K. (1993). Natural language processing: Some recent trends. *Current Science*, 393–406. https://www.jstor.org/stable/24098875

Levinson, S. C. (1995). *Pragmatics*. Cambridge University Press.

Mey, J. L. (2000). *Pragmatics: An Introduction*. Blackwell.

Morgan, J. L. (1978). "Two Types of Convention in Indirect Speech Acts". In *Pragmatics: A Reader* (1st ed.) (pp. 242-253). Oxford University Press.

Sarikaya, R. (2017). The Technology Behind Personal Digital Assistants: An overview of the system architecture and key components. *IEEE Signal Processing Magazine*, *34*(1), 67–81. https://doi.org/10.1109/msp.2016.2617341

Searle, J. R. (1965). „What Is a Speech Act?" In *Pragmatics: A Reader* (1st ed.) (pp. 254-264). Oxford University Press.

Searle, J. R. (1975). „Indirect Speech Acts." In *Pragmatics: A Reader* (1st ed.) (pp. 265-277). Oxford University Press.

Searle, J. R. (1976). A Classification of Illocutionary Acts. *Language in Society, 5(1),* pp. 1-23. https://www.jstor.org/stable/4166848

Strawson, P. F. (1974). „Intention and Convention in Speech Acts." In *Pragmatics: A Reader* (1st ed.) (pp. 290-302). Oxford University Press.

Wunderlich, D. (1986). WIE KOMMEN WIR ZU EINER TYPOLOGIE DER SPRECHAKTE? *Neuphilologische Mitteilungen,* pp. 498-509. https://www.jstor.org/stable/43343770

Yule, G. (1998). *Pragmatics*. Oxford University Press.

Zhao, L., Alhoshan, W., Ferrari, A., & Letsholo, K. J. (2022). Classification of Natural Language Processing Techniques for Requirements Engineering. Retrieved from arxiv.com. https://arxiv.org/abs/2204.04282

Abstract

As artificial intelligence technology is becoming increasingly advanced, the line between natural speech and computer-generated speech is starting to blur. Machines are starting to comprehend nuances in conversation and read between the lines. One of the more complex abilities of natural language is the possibility of saying one thing, but mean something else entirely. As AI assistants, i. e. personal virtual assistants, are made to help users with increasingly advanced tasks and speak to them more naturally, it is important that they are able to understand possible underlying intentions in that regard to fulfil their duty. This paper offers an insight into the degree of indirectness a personal virtual assistant can comprehend without losing the main focus of the sentence. Using traditional pragmatic concepts like speech acts, implicature, convention and the cooperative principle, a theoretical framework is constructed in order to shed some light on the human side of the interaction. As for the machine side, contemporary literature from computer science with an emphasis on natural language processing is reviewed and a rudimentary overview of the main concepts is given, which should help elucidate some of the processes a virtual assistant goes through in order to correctly interpret a sentence. The analysis consists of five basic and frequently used direct commands, which are turned into multiple indirect requests using a different sentence structure, i. e. in the form of a declarative and interrogative sentence with varying degrees of indirectness. The analysis will try to clarify the ways a virtual assistant processes a specific request and, if interpreted incorrectly, where the interpretation went wrong. The importance of a pragmatic framework will also be tested to see to what extent can it be applied in a conversational analysis, where one of the participants is a machine and how it consequentially correlates to the programming of the virtual assistant.

Keywords: virtual assistant, pragmatics, indirect speech acts, natural language processing, conversation analysis