2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Abductive Recognition of Context-dependent Utterances in Human-robot Interaction

Davide Lanza[†], Roberto Menicatti[†], Antonio Sgorbissa[†]

***Abstract*— Context-dependent meaning recognition in natural language utterances is one of the key problems of computational pragmatics. Abductive reasoning seems apt for modeling and understanding these phenomena. In fact, it presents observations through hypotheses, allowing us to understand subtexts and implied meanings without exact deductions. For this reason in this paper, we are going to explore abductive reasoning and context modeling in human-robot interaction. Rather than a radical inferential approach, we assumed a conventional approach towards context-depending meanings, i.e, they are conventionally encoded rather than inferred from the utterances. In order to address the problem, a case study is presented, analyzing whether such a system could manage correctly these linguistic phenomena. The results obtained confirm the validity of a conventional approach in context modeling and, on this basis, further models are proposed to work around the limitations of the case study.**

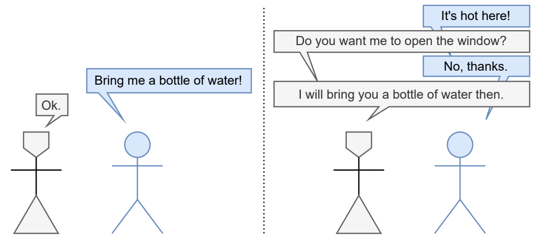

Fig. 1. Example of human robot interaction. On the left, the command is expressed through a direct speech act [9]. On the right, the desire of having a bottle of water is expressed through an indirect speech act [9]. In the latter case, the robot asks to perform actions which are close to the user's sentence, trying to infer the underlying request. By giving a certain contextual knowledge, our robot client carries out this task thanks to an abductive inference model, which is able to infer underlying requests.

I. INTRODUCTION

In order to develop artificial systems capable of human-like verbal interaction and effective results, it is fundamental to ensure a coherent model of context awareness. In human-robot interaction, the user normally expects the robot to execute his commands and orders. Meanwhile when the user relates to his kind, the requests do not necessarily have to be direct commands (as "turn the light on" or "bring me a pen") [1]. These can be expressed through implications, circumlocutions, allusions or the expression of specific needs (e.g. "it's so dark here" or "I can't see nothing" for "turn the light on") [2]. A robot that aims to react to these stimuli as an intelligent system should be able to extract from these expressions the underlying meanings. Context-awareness is fundamental to understand these non-literal meanings [3], [4].

In this paper, the problem of recognizing context-dependent meanings is solved by using pre-existing probabilistic data. Most of the current state-of-the-art in human-robot verbal interaction follows a keyword-recognition approach, collecting and categorizing specific triggering words related to certain command scopes (e.g. the CARESSES system [5]). This approach can be expanded as well through Natural Language Processing (NLP) methods, such as Gensim software framework for topic modelling [6], and tools in the Cloud that make this process easier (e.g. DialogFlow or IBM's Watson). However, most of the pre-existing approaches still require explicit commands from the users, without taking into account context modeling at all. Compared to the aforementioned techniques, we believe that a context-wise modeling approach, based on abductive reasoning formalization, can lead to more efficient solutions, needing smaller datasets and information.

Abductive reasoning has been extensively used in AI research [7], [8]. The aim of this paper is to provide a solution for dealing with indirect speech acts in human-robot interaction applications (Fig. 1) in a given case study. The purpose is served through the development of an abductive reasoning model for context-dependent meaning recognition. As shown in the experiments, this model ensures a better accuracy with a small amount of data, paving the way for abduction-oriented interactive systems. In Section II we introduce the theoretical framework which inspired our approach. Meanwhile Section III illustrates the model definition, Section IV explain how we implemented it in a case study, and in Section V the experimental results are shown and discussed. Finally in Section VI further improvements are proposed.

II. THEORETICAL FRAMEWORK

A. Conventional approach to pragmatics

Pragmatics, which is a subfield of linguistics, is the study of how context contributes to meaning in linguistic interactions. Therefore, it deals with all those utterances whose meaning cannot be limited to the literal one. An important typology of utterances are the *indexical expressions* [10], [11] which allude to a specific context. By using indexicals, one can refer to places, people, instants or objects in a specific relationship with the speaker. Contrary to context-free statements, (e.g. "2 + 2 = 4", or "Dogs are mammals'"), indexicals are difficult to be understood without knowing the context and the relations between them and the speaker (e.g.

[†] Department of Computer Science, Bioengineering, Robotics and Systems Engineering (DIBRIS) University of Genova, Via Opera Pia 13, 16145 Genova, Italy.

Deduction	*Induction*	*Abduction*
x is P	x is P	$P \Rightarrow Q$
$P \Rightarrow Q$	x is Q	x is Q
x is Q	$P \Rightarrow Q$	x is P

Fig. 2. Deductive, inductive and abductive reasoning compared. Only deduction reaches a logically certain conclusion, while inductive and abductive inferences are defeasible.

"Someone's ringing the doorbell" or "I forgot to pack my luggage"). Bar-Hillel [12] argued that, in natural language, more than 90% of all declarative utterances are indexical.

Not just indexicals, but many linguistic structures rely on contextual information, such as *presuppositions* [13], *conversational implicatures* [14], direct and indirect *speech acts* [9]. Understanding all the aforementioned structures is possible when the speaker and the hearer share any pre-existing knowledge, in other words, these inferences rely on premises that are not part of the content of the utterance itself [10]. Grice highlighted how communication is based on a *cooperative principle* [14] and from this hypothesis he defined implicatures as an unreliable form of inference. Though this emphasis on cooperation has been criticized [15], the notion of implicature has been widely used to explain how the speaker expresses his intentions and requests through indirect speech acts as the following:

"There's a howling gale in here!"→"Shut the window".

A *radical inferential approach* towards implicature analysis would assume that the speaker's intentions rely on the hearer's ability to make appropriate additional assumptions. We adopted instead a *conventional approach*: utterances, like the one previously mentioned, are conversational gambits conventionally encoded, and there is no need for long inferential chains to understand them on most occasions [10].

B. Abductive reasoning

Understanding natural language relative to a context involves inferences of abductive nature. Abductive reasoning seems to be very common in everyday reasoning [10]. However, it does not necessarily lead to correct conclusions (as well as induction). Peirce [16] introduced the formal notion of abduction as the third way of reasoning, along with deduction and induction (Fig. 2). A classical example is the following one:

$$\frac{\text{All the beans from this bag are white.}\quad \text{These beans are white.}}{\text{These beans are from this bag.}}$$

From the example, it is clear how abduction ends up with a defeasible hypothesis rather than a true statement. Nevertheless, solid rules which formalize previous knowledge can lead to accurate results in most cases. To express the previous knowledge[1] consider a set of Horn clauses gathered in a Knowledge Base set KB

$$KB = \{(\neg x_1 \wedge \neg x_2 \wedge ... \wedge \neg x_n \wedge x_{n+1}) \Rightarrow y\} \quad (1)$$

[1]The formalization of abductive reasoning in our model is based on [17].

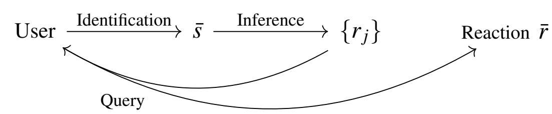

Fig. 3. Functional scheme of verbal interaction model. After receiving an utterance from the user, the robot enters in the identification processes, where it identifies the contextual situation which the input sentence is referring to. Subsequently, during the inference process, it abductively obtains the reactions which can be linked to the situation $\bar{s}$ identified. After that, the robot enters the query process, where it asks the users in order to disambiguate between the reactions, if there are more than one. Finally, the reaction process executes the action linked to the selected reaction $\bar{r}$.

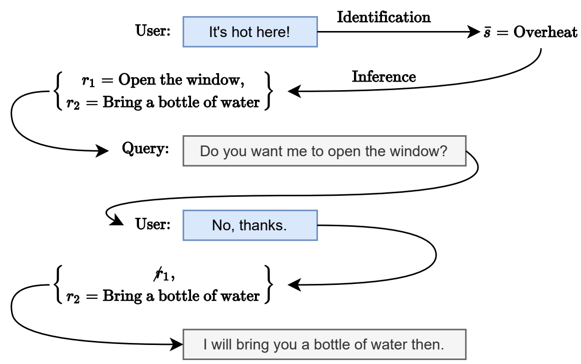

Fig. 4. Verbal interaction model processes for the verbal interaction example illustrated in Fig. 1. Here, two possible reactions are inferred, and $\bar{r} = r_2$ is performed. In our model, the querying order of the reactions is based on a probabilistic-like mapping as the one shown in Tab. II.

and given a set of atomic assumables A, define a scenario H of $\langle KB, A\rangle$ as

$$H \subset A \text{ s.t. } (KB \cup H \text{ satisfiable}) \quad (2)$$

where satisfiable means that no H subset is a conflict for KB. A proposition u from $\langle KB, A\rangle$ has an explanation H if

$$(H \text{ scenario of } \langle KB, A\rangle) \wedge (KB \cup H \models u) \quad (3)$$

It is possible to define as well a minimal explanation H_{min} as an explanation of u for $\langle KB, A\rangle$ such that

$$\nexists H \subset H_{min} \text{ s.t. } H \text{ explanation of } \langle KB, A\rangle \quad (4)$$

Given these rules, we can easily express the abductive inference process. Pragmatic phenomena, as conversational implicatures, are understood where it is possible to have a minimal explanation H_{min} which allows to interpret coherently and/or cooperatively the utterance u (for the previous example, $u =$ "These beans are white." and $H_{min} =$ "These beans are from this bag.").

III. MODEL DEFINITION

With our model, we propose to apply the abductive reasoning formalization of Section II to human-robot verbal interactions. Given contextual information in a KB form, the robot's verbal interaction module should interpret the user's input utterances u bearing non-literal meanings. The reaction of the robot should be compliant to this process

TABLE I

CONTEXTUAL TABLE

KB		
$s_1 \Rightarrow r_1$	$s_1 \Rightarrow r_2$	
	$s_2 \Rightarrow r_2$	$s_2 \Rightarrow r_3$
	$s_3 \Rightarrow r_2$	
$s_4 \Rightarrow r_1$		

TABLE II

PROBABILISTIC-LIKE MAPPING FOR TAB. I

$P : KB \to [0,1]$		
$P(s_1 \Rightarrow r_1)$	$P(s_1 \Rightarrow r_2)$	0
0	$P(s_2 \Rightarrow r_2)$	$P(s_2 \Rightarrow r_3)$
0	$P(s_3 \Rightarrow r_2)$	0
$P(s_4 \Rightarrow r_1)$	0	0

– for example, linking an utterance as "I am freezing" with the request of either closing an opened window in the room or turning on the heating system. The case study we introduced is limited to a constrained situation, allowing to analyze the system dynamics without addressing bigger scales issues. Proposals to work around these limitations are then presented.

We considered a set S of different contextual situations s_i and a set R of available robot reactions r_j. We divided human-robot verbal interaction in four processes: identification, inference, query and reaction (Fig. 3). A verbal interaction example is reported in Fig. 4.

During identification, the user's input is processed with classical NLP techniques. A string u (the user's utterance) is returned by the speech-to-text (STT) module of the robot. Then, u is processed in order to identify a situation $\bar{s} \in S$ linked to it. This situation belongs to a situations set S which depends on the modelled context. In our implementation, identifying the relevant situation is done through the DialogFlow Cloud platform.

For context-recognition modeling, we assumed a conventional approach, as mentioned in II-A. Hence, the system should not build a complex inferential chain starting from u in order to recognize a non-literal meaning and then process in order to have coherent reactions. The system instead, has a simplified KB in the form

$$KB = \{s_i \Rightarrow r_j\} \tag{5}$$

Such a KB can be represented with a contextual table as shown in Tab. I, where each contextual situation s_i may implicate different reactions r_j. In this example, we considered a set of situations $S = \{s_1, s_2, s_3\}$ and a set of reactions $R = \{r_1, ..., r_4\}$. During inference, a set of possible reactions $\{r_j\} \subset R$ associated to the current situation $\bar{s}$ is extracted by means of the abductive reasoning model (R is the assumable set here). Because of the way we built this KB, every (minimal) explanation H will contain only one element of R. For this reason, we can identify H_j with its r_j. At the end of the inference process, the subset of reactions $\{r_j\} \subset R$ which are suitable with the observation $\bar{s}$ is returned.

With querying, given the context KB, the system disambiguates the reactions in $\{r_j\}$ trying to understand which reaction would be the most desired one in the specific situation $\bar{s}$. As previously shown, in Tab. I we have a *table of linguistic conventions* which defines the characteristics of a determined linguistic context. In order to choose a priority criteria between reactions, we decided to define a probabilistic-like mapping $P : KB \to [0,1]$ (Tab. II) which provides a way to queue the different r_j in querying phase before directly addressing the user. Consider to have a context as the one in Tab. I and suppose $\bar{s} = s_2$. The inference process would return $\{r_2, r_3\}$. Supposing $P(s_2 \Rightarrow r_2) > P(s_2 \Rightarrow r_3)$, a reasonable querying routine would ask the user to understand whether to perform r_3 and then, if not, will perform r_2. For example, for s_2 = "room too dark" triggered by u = "I can't see anything here", given r_3 = "draw the blinds" and r_2 = "turn the light on" the robot would ask the users if they want the lights turned on and, if not, he would draw the blinds.

While for the example illustrated in Tab. I and Tab. II it is not necessary to detach the KB from the probabilistic mapping (instead of using a probabilistic lookup table), it is important to keep the two concepts separated for many reasons. We introduced before a strong limitation on KB, using just Horn clauses which do not enlist any negated element, while in general, they are a disjunction of literals with at most one positive, unnegated, literal. Moreover, KB could be dynamically modified and expanded (as we proposed as further work to implement). Keeping detached KB and P allows to switch more easily between different knowledge representations paradigms, especially given an automated mapping process that assigns weights to each element of KB without human supervision. Moreover, knowledge representation for abduction does not imply probabilistic weights, and this should be kept in mind while working with such models. We can also change mapping techniques or adopt different maps for different contexts.

The reaction phase is excluded from Fig. 4 because it consists just in the execution of the robotic operative correlate to r_j. It regards the single implementation and it is independent of the model workflow.

IV. IMPLEMENTATION

A. Use case definition

The use case presented here is based on three available reactions

$$R = \{r_1, r_2, r_3\} \tag{6}$$

with some situations s_i in common. We imagined a robot interacting with a user in his/her flat. The robot is able to perform three actions, namely, opening windows (r_1), bringing the user a bottle of water (r_2) and turning on the TV (r_3). Intuitively r_1 and r_2 can be linked by a strong common cause (e.g. overheating) but they present as well different characteristics (e.g. one may want to open a window to check something outside). Moreover, r_3 is intuitively less linked

TABLE III

FIRST FORM RESULTS

S			r_1	r_2	r_3
s_1	$\triangleq$	Overheat	10%	2%	
s_2	$\triangleq$	Bug in the room	7%	2%	
s_3	$\triangleq$	Plants outside need water	4%	8%	
s_4	$\triangleq$	Check the weather	3%		3%
s_5	$\triangleq$	Air the room	10%		
s_6	$\triangleq$	Noise/stimulus from outside	5%		
s_7	$\triangleq$	Unpleasant smell	4%		
s_8	$\triangleq$	Smell of gas	2%		
s_9	$\triangleq$	Feeling suffocated	2%		
s_{10}	$\triangleq$	Being thirsty		12%	
s_{11}	$\triangleq$	Workout, do exercise		5%	
s_{12}	$\triangleq$	No water at the dining table		4%	
s_{13}	$\triangleq$	Open fire in the house		4%	
s_{14}	$\triangleq$	No current water		4%	
s_{15}	$\triangleq$	Cooking	3%		1%
s_{16}	$\triangleq$	Keep up with current events			14%
s_{17}	$\triangleq$	Want to watch a film/TV series			10%
s_{18}	$\triangleq$	Need to cheer up or relaxing			9%
s_{19}	$\triangleq$	Being bored			7%
s_{20}	$\triangleq$	Not being able to sleep			3%
s_{21}	$\triangleq$	The house is too quiet			2%
		Other situations (discarded)	50%	59%	51%

to the first two. Hence, this choice of thee reactions may show all the different dynamics of the system, e.g. excluding r_3 and disambiguating between r_1 and r_2 during the query process.

B. Data collection

We collected a dataset specifically for our use case. This choice was made to see if such a system could obtain good accuracy even with a small, noisy dataset which is not suitable for classical NLP techniques. Two survey forms have been made, the first one to collect the biggest number of situations s_i linked to $r_i \in R$ and the second one to build a corpus of sentences associated with each situation s_i.

The first form was filled by 18 people between 19 and 55 years old (85% about 20 years old and 15% about 50 years old). We asked the subjects to picture themselves in a certain situation which required one of the three robot reactions $r_i \in R$. In the form, they had to enlist ten causes to that particular desire. The results are enlisted in Tab. III. As reported, we selected only the best situations with respect to the number of reactions linked to them and their frequency in the form answers (i.e. their "probability"). From the results we can see how, for example, users suggested that opening the window may be a consequence of overheat, the presence of a bug, or a bad smell, just to name a few. Bringing a bottle of water, instead, could maybe be a consequence of either the need to water the plants or the fact that the user is thirsty. Notice that, with a number of samples sufficiently high, the percentage in Tab. III can be straightforwardly used in order to compute the probability in Tab. II, as the frequency of the event $s_i \Rightarrow r_j$ converges to the probability of the event.[2] So we assumed that the number of answers was high enough to proceed in this way.

[2]Recall that the probabilistic-like mapping does not have a probabilistic connotation by itself. Hence, "probabilities" in Tab. II do not have to sum up to one. They are rather weights based on first form's probabilities.

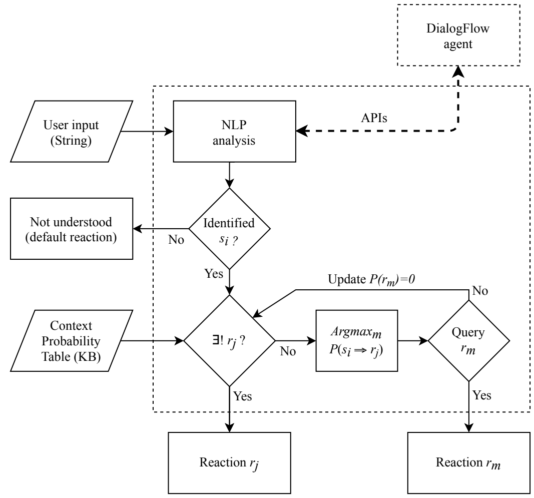

Fig. 5. Workflow of the client implemented starting from the functional scheme illustrated by Fig. 5.

The second form was filled by 17 people between 20 and 50 years old who did not fill the first form (94% about 20 years old and 6% about 50 years old). The form asked the subjects to picture themselves in their house in three different contexts (with a relative, with some guests or with their butler, assuming that different degrees of relationship would imply different wordings of the same request). In the form, they had to enlist five sentences with which they would express being in the situations $s_i \in S$. The total corpus consisted of a small dataset of 1771 sentences expressed in Italian[3], from 80 to 85 for each of the 21 situations. For example, a user suggested that, to address the situation $s_3 =$ "plants outside need water", he would say to a relative "our plants are not looking healthy", to a guest "I guess I forgot to water the plants yesterday", and to their butler "you have to water the plants".

C. Client development

To manage NLP for situation identification we used the DialogFlow platform. In the DialogFlow Agent we defined an agent whose intents were the 21 different situations s_i.[4] Each intent has been trained with the corresponding 85 sentences of the collected corpus, which - as already explained - represent different ways in which different users may express their desires in that specific situation when talking with a relative, a guest, or their butler.

The client was developed in Python and accessed DialogFlow through the provided APIs. [5] The input is taken from the user's string and the context probability table based on the KB. To perform identification, the client relies on DialogFlow, and it does not involve any of the actual implementations of the reaction (being a console-based program). If a situation s_i is detected, and for that situation it exists only

[3]The model was implemented for Italian-speaking users.

[4]For further information about DialogFlow's Intents see https://cloud.google.com/dialogflow/docs/intents-overview

[5]The GitHub repository related to this paper is available at https://github.com/Davidelanz/nlp-contextual-meaning.

TABLE IV

DATASET'S SEMANTIC OVERLAPS

Situation expected	Test sentence	Situation identified
s_1 = Overheat	I can't breathe.	s_5 = Air the room
s_5 = Air the room	Open the window!	s_7 = Unpleasant smell
s_5 = Air the room	It smells like gas in the kitchen.	s_8 = Smell of gas
s_5 = Air the room	There's stale air in the room.	s_7 = Unpleasant smell
s_5 = Air the room	There's a bad smell.	s_7 = Unpleasant smell
s_5 = Air the room	I can't breathe.	s_9 = Feeling suffocated
s_6 = Noise/stimulus from outside	Anybody home?	s_{21} = The house is too quiet
s_8 = Smell of gas	I can't breathe.	s_9 = Feeling suffocated
s_{12} = No water at the dining table	There's no water.	s_{14} = No current water
s_{12} = No water at the dining table	There's no more water.	s_{14} = No current water

one reaction, it is immediately executed. Otherwise, based on the context probability table, the system prioritizes reactions according to their probabilistic-like mapping $P(s_i \Rightarrow r_j)$, and queries the user. The client's workflow is shown in Fig. 5).

V. VALIDATION AND RESULTS

A. Identification validation

The DialogFlow agent accuracy for the identification task has been tested with k-fold cross validation [18], choosing $k = 5$. The accuracy results are reported in the confusion matrix [19] at Tab. V and synthetically in Tab. VI.[6] The accuracy index has been defined as the correct/total ratio for the singles expected situations s_i and for the total matrix. In Tab. V the additional column D indicates a *default fallback* intent returned by DialogFlow, i.e. none of the 21 situations in S identified. This shows that DialogFlow is accurate enough to match users utterances with situations, which is the preliminary step for ultimately producing a reaction. Moreover, since the 21 chosen situations overlap semantically with each other, it is possible to have in the dataset similar sentences for different situations. As we can see from Tab. IV, these overlaps lead to smaller accuracy results, but they are not as bad as actual identification errors.

B. Experimental results

The system's total accuracy depends on two main factors: the effectiveness of the identification process and the quality of KB modeling. For the final testing, we asked to a test group of people (different from the two used for collecting data) to picture themselves in some predefined situations. Every user was instructed with the three actions the robot can perform, a set of "forbidden words" to avoid direct addressing of the desired reaction and seven situations from S in which picture himself in. An example of the testing module they received is reported in Fig. 6. Each user interacted with the client 5 times for each situation and annotated whether the output action was the one he was implying. The use of forbidden words and the directive of not directly requesting the desired action allowed us to have a set of input sentences that resembled the use case real-world situation

```
The robot you are interacting with can:
  - bring you a bottle of water,
  - turn on your television,
  - open the window of the room you are in.

Forbidden Words:
  - turn on, open, take, bring, water, bottle,
    window, television

 -------------------------- Example -------------------------

The given situation is "There is a cake in the oven". You
want the robot to turn off the oven. You have five attempts.
At the first attempt you input "The cake is burning!" and
the robot turns off the oven. Hence you complete this form as
follows:

Situation 0. There is a cake in the oven

   Attempt no. 1:
   - I wanted the robot to turn off the oven
   - Got it? [Y]

 -------------------------------------------------------------

Situation 1. There is no water at the dining table.

   Attempt no. 1:
   - I wanted the robot to _________________
   - Got it? [ ]

   Attempt no. 2:
   - I wanted the robot to _________________
   - Got it? [ ]

(...)

Situation 7. (...)
```

Fig. 6. Example of the module given to the users in order to test human-robot interaction performances in our model. Every module was presenting 7 different situations, for a total of 35 attempts for each user.

here described. To evaluate these experimental results we defined two performance indexes:

$$\alpha = \frac{\text{correct runs}}{\text{total runs}} \qquad \beta = \frac{\text{average runs for the}}{\text{first correct result}} \tag{7}$$

The results obtained are reported in Tab. V. The final result allows to obtain a value for KB modeling accuracy. In fact, the accuracy of the system $\alpha = 82\%$ depends on the one of the identification process ($\alpha_{id} = 91.5\%$) and on the accuracy of the knowledge base (α_{KB}), hence:

$$\alpha := \alpha_{KB} \cdot \alpha_{id} \tag{8}$$

Given this and the data we got from the tests, we can obtain:

$$\alpha_{KB} = \frac{\alpha}{\alpha_{id}} = \frac{82\%}{91.5\%} = 89.6\% \tag{9}$$

It needs to be said that the small amount of users used to collect data, in addition to the testing methods adopted and the limited period of time in which the project was carried out, have surely affected the overall system's accuracy. Moreover, the client was evaluated with a human-computer interaction test, rather than an actual human-robot interaction, therefore, this could have provoked the enforcement of unnatural and not spontaneous inputs, which concurred in the precision of the system. In the next Section we describe our proposals to work around these limitations as well as the constraints of the use case scenario.

VI. DISCUSSION AND FURTHER WORK

A. Horn clauses extension

In our model, we used Horn clauses made by the single unnegated literal. An extension to actual Horn clauses which

[6]For the 5-fold cross validation, each training fold was composed by 16 test sentences, for a total of 80 test sentences for every intent (the smallest number of sentences associated to an intent in the dataset). In Tab. V the expected intents are reported on the rows and the predicted ones on the columns. The additional column D represents the default intent response.

TABLE V

5-FOLD CROSS VALIDATION CONFUSION MATRIX

S	Output s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}	s_{13}	s_{14}	s_{15}	s_{16}	s_{17}	s_{18}	s_{19}	s_{20}	s_{21}	D
s_1	72	2	.	1	2	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	1
s_2	.	73	3	.	.	.	3	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
s_3	.	.	76	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
s_4	.	.	.	77	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
s_5	.	.	.	.	62	2	7	1	6	.	.	1	.	.	.	.	.	1	.	.	.	.
s_6	.	.	.	.	.	73	2	.	.	.	.	.	.	.	.	1	.	.	1	.	1	2
s_7	.	.	.	.	3	.	72	2	2	.	.	.	.	.	.	.	.	.	1	.	.	.
s_8	.	.	.	.	1	.	3	72	4	.	.	.	.	.	.	.	.	.	.	.	.	.
s_9	.	.	.	.	1	.	1	.	75	2	.	.	1	.	.	.	.	.	.	.	.	.
s_{10}	.	.	.	.	.	.	.	.	.	76	2	2	.	.	.	.	.	.	.	.	.	.
s_{11}	.	.	.	.	.	.	.	.	.	.	76	2	.	.	1	1	.	.	.	.	.	.
s_{12}	.	.	.	.	.	.	.	.	.	3	.	65	3	9	.	.	.	.	.	.	.	.
s_{13}	.	.	.	.	.	.	1	.	.	.	.	.	73	2	.	.	.	.	.	.	.	4
s_{14}	.	.	.	.	.	1	.	.	.	.	.	3	.	73	2	.	.	.	.	.	.	1
s_{15}	.	.	.	.	.	.	1	.	.	.	.	.	.	.	76	2	.	.	.	.	.	1
s_{16}	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	77	2	.	.	.	.	.
s_{17}	.	1	.	.	.	.	.	.	.	.	.	.	.	.	1	1	75	2	.	.	.	.
s_{18}	1	.	.	.	1	.	.	.	.	.	1	.	.	.	.	.	.	72	4	.	.	1
s_{19}	1	.	1	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	73	2	.	2
s_{20}	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	74	3	1
s_{21}	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	75	1

TABLE VI

SINGLE SITUATION ACCURACY AND TOTAL ACCURACY

Identification accuracy (%)							
s_1	90	s_2	91.25	s_3	95	s_4	96.25
s_5	77.5	s_6	91.25	s_7	90	s_8	90
s_9	93.75	s_{10}	95	s_{11}	95	s_{12}	81.25
s_{13}	91.25	s_{14}	92.25	s_{15}	95	s_{16}	96.25
s_{17}	93.75	s_{18}	90	s_{19}	91.25	s_{20}	92.5
s_{21}	93.85						
Overall accuracy: 91.48							

TABLE VII

USER-CLIENT INTERACTION RESULTS

S	α	β	S	α	β
s_1	90%	1	s_{11}	84%	1
s_2	86%	1	s_{12}	82%	1
s_3	86%	1	s_{13}	82%	1,5
s_4	84%	1	s_{14}	82%	1
s_5	90%	1	s_{15}	74%	1
s_6	70%	1	s_{16}	82%	1
s_7	82%	1	s_{17}	84%	1
s_8	86%	1	s_{18}	70%	1
s_9	82%	1,5	s_{19}	86%	1
s_{10}	84%	1	s_{20}	70%	1
s_{21}	86%	1			
TOT	**82%**	**1.05**			

are a disjunction of negated literals with an unnegated one has to be studied. If it is true that many contextual reactions to some situations rely on other environmental conditions (and then a disjunction would be ideal for store these kinds of information), the complexity of the model may drastically increase.

B. Dataset automated extension

In this work, it has been demonstrated how it is possible to achieve effective accuracy results with a small noisy dataset. Since building a bigger dataset can be an arduous task that takes too much time, an automatic expansion algorithm could be an alternative. In fact, starting from the sentences in the dataset, the algorithm can process them in order to return most of the common variations for each sentence, according to a syntactical and semantical point of view [20], [21], [22].

C. Adaptive learning

An interesting improvement could be a system able to start with an empty KB (or with a minimal general-purpose one) and so able to learn through a simple command. The system would add new situations when told by the user, enabling connections with unrecognized sentences. As shown in Fig. 7, if $\bar{s}$ can't be identified, the system would ask the user if the situation already exists in S ("$\exists s_i$?"). If not, it should add it to KB. A correctness check is added as well after the reaction, to understand if the robot acted accurately. If not, the robot would be able to decrease the correspondent weight mapping P.

This system would likely have a slow learning curve, but it could be a good solution to customize robots in various contexts not already modeled. Moreover, the KB from different robots could be integrated into a cloud system that would fasten the learning process, and so the knowledge sharing between the robots.

A problem which this extension would address is related to S increasing size. In fact, the more the sets are composed by different situations, the more the P weights might be reduced and be close in value so that selecting the wrong reaction from R would be more probable. This phenomenon would be even worse with real-world cases with a high amount of available reactions.

One could argue that the number of reactions is, at a certain extent, limited to the scope of the robot, and so the number of situations can intensify, but this does not necessarily produces a more homogeneous table. This problem should be addressed with use cases on a larger scale, trying to improve the model if needed. For example, by adding an intermediate query process, where the robot asks something like "what do you mean?", in order to have multiple identification results and reduce uncertainty by intersecting them.

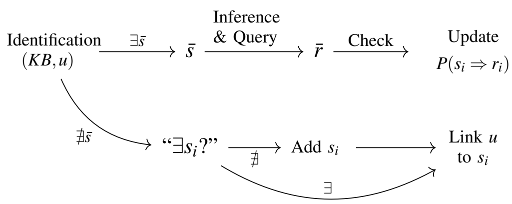

Fig. 7. Functional scheme integrating adaptive learning

D. Improved identification

In this project, we demanded NLP situation identification to DialogFlow. This choice was made because the aim was to focus on the inference modeling part. Identification can be improved using classical NLP literal-meaning recognition techniques. Tools as Gensim [6] and state-of-the-art methods can enhance the identification accuracy.

VII. CONCLUSIONS

In conclusion the study aims to make a first step towards abductive reasoning modeling for a human-robot verbal interacting case study. By employing abduction, the robot is able to extract, through contextual information, the underlying context-dependent meanings borne by utterances: including indexicals, conversational implicatures, and others. The study, therefore, has analyzed the effectiveness of a conventional approach to linguistic forms decoding. The agent responds correctly to statements of these types, by demonstrating a good accuracy, considering the limited toy model. On this basis, further developments have been proposed in order to overcome these limitations.

Being the problem extremely complex, there is no presumption here to claim that we designed a system capable of reacting to contextual utterances correctly. However, it has been shown that, especially in limited contexts, the conventional approach - combined with probabilistic modeling of the abductive hypotheses - leads to remarkable results. The further work proposal has to be read as the next step to approach coherent modeling of the phenomenon in its entirety.

ACKNOWLEDGMENT

Many thanks to Marcello Frixione, full professor of the philosophy department at the University of Genoa, for his help and advice as regards the conceptual framework of the project.

REFERENCES

[1] N. C. Krämer, A. von der Pütten, and S. Eimler, “Human-Agent and Human-Robot Interaction Theory: Similarities to and Differences from Human-Human Interaction,” in *Human-Computer Interaction: The Agency Perspective*, ser. Studies in Computational Intelligence, M. Zacarias and J. V. de Oliveira, Eds. Berlin, Heidelberg: Springer, 2012, pp. 215–240. [Online]. Available: https://doi.org/10.1007/978-3-642-25691-2_9

[2] G. Briggs, T. Williams, and M. Scheutz, “Enabling robots to understand indirect speech acts in task-based interactions,” *Journal of Human-Robot Interaction*, vol. 6, no. 1, pp. 64–94, May 2017. [Online]. Available: https://doi.org/10.5898/JHRI.6.1.Briggs

[3] K. Bach, “Literal Meaning.” *Philosophy and Phenomenological Research*, vol. 75, no. 2, pp. 487–492, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1933-1592.2007.00088.x

[4] K. DeRose, “Assertion, Knowledge, and Context,” *Philosophical Review*, vol. 111, no. 2, pp. 167–203, 2002, publisher: Duke University Press.

[5] C. T. Recchiuto, C. Papadopoulos, T. Hill, N. Castro, B. Bruno, I. Papadopoulos, and A. Sgorbissa, “Designing an Experimental and a Reference Robot to Test and Evaluate the Impact of Cultural Competence in Socially Assistive Robotics,” in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, Oct. 2019, pp. 1–8, iSSN: 1944-9437.

[6] R. Rehurek and P. Sojka, “Software Framework for Topic Modelling with Large Corpora,” in *In Proceedings of the Lrec 2010 Workshop on New Challenges for Nlp Frameworks*, 2010, pp. 45–50.

[7] G. Paul, “AI Approaches to Abduction,” in *Abductive Reasoning and Learning*, ser. Handbook of Defeasible Reasoning and Uncertainty Management Systems, D. M. Gabbay and R. Kruse, Eds. Dordrecht: Springer Netherlands, 2000, pp. 35–98. [Online]. Available: https://doi.org/10.1007/978-94-017-1733-5_2

[8] T. Menzies, “Applications of abduction: knowledge-level modelling,” *International Journal of Human-Computer Studies*, vol. 45, no. 3, pp. 305–335, Sept. 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581996900543

[9] J. L. Austin, *How To Do Things With words: The William James Lectures Delivered at Harvard University in 1955*. Oxford, New York: Oxford University Press, Sept. 1975.

[10] H. Bunt and W. Black, “The ABC of Computational Pragmatics,” in *Abduction, Belief and Context in Dialogue: Studies in computational pragmatics*, ser. Natural Language Processing, H. Bunt and W. Black, Eds. Amsterdam: John Benjamins Publishing Company, 2000, vol. 1. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.22.4842

[11] D. Kaplan, “Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals,” in *Themes From Kaplan*, J. Almog, J. Perry, and H. Wettstein, Eds. Oxford University Press, 1989, pp. 481–563.

[12] Y. Bar-Hillel, “Logical Syntax and Semantics,” *Language*, vol. 30, no. 2, pp. 230–237, 1954, publisher: Linguistic Society of America. [Online]. Available: https://www.jstor.org/stable/410265

[13] F. Domaneschi and C. Penco, “Presupposizioni,” *APhEx Portale Italiano di Filosofia Analitica*, no. 15, 2017. [Online]. Available: aphex.it/index.php?Temi=557D030122027403210403027773 27

[14] H. P. Grice, “Logic and Conversation,” in *Syntax and Semantics*, academic press ed., P. Cole and J. Morgan, Eds., 1975, vol. 3, reprinted as ch.2 of Grice, H.P. (1989). “Studies in the Way of Words”. Harvard University Press, pp. 22-40.

[15] D. Wilson and D. Sperber, “On Choosing the Context for Utterance Interpretation,” in *Foregrounding background*, J. Allwood and E. Hjelmquist, Eds. Doxa, 1981.

[16] C. S. Peirce, “On the Natural Classification of Arguments,” *Proceedings of the American Academy of Arts and Sciences*, vol. 7, pp. 261–287, 1867.

[17] D. Poole and A. Mackworth, *Artificial Intelligence. Foundations of Computational Agents*, 2nd ed. Cambridge University Press, 2017. [Online]. Available: https://artint.info/2e/

[18] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York: Springer-Verlag, 2013. [Online]. Available: https://www.springer.com/gp/book/9781461468486

[19] R. Kohavi and F. Provost, “Glossary of Terms,” *Machine Learning*, vol. 30, no. 2, pp. 271–274, Feb. 1998. [Online]. Available: https://doi.org/10.1023/A:1017181826899

[20] M. Fadaee, A. Bisazza, and C. Monz, “Data Augmentation for Low-Resource Neural Machine Translation,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vancouver, Canada: Association for Computational Linguistics, July 2017, pp. 567–573. [Online]. Available: https://www.aclweb.org/anthology/P17-2090

[21] Z. Xie, S. I. Wang, J. Li, D. Lévy, A. Nie, D. Jurafsky, and A. Y. Ng, “Data noising as smoothing in neural network language models,” Jan. 2019. [Online]. Available: https://collaborate.princeton.edu/en/publications/data-noising-as-smoothing-in-neural-network-language-models

[22] S. Yu, J. Yang, D. Liu, R. Li, Y. Zhang, and S. Zhao, “Hierarchical Data Augmentation and the Application in Text Classification,” *IEEE Access*, vol. 7, pp. 185 476–185 485, 2019, conference Name: IEEE Access.