Culturally-Competent Human-Robot Verbal Interaction (Special Session on Culture-Aware Robots)

Barbara Bruno[1], Roberto Menicatti[1], Carmine T. Recchiuto[1], Edouard Lagrue[2], Amit K. Pandey[2] and Antonio Sgorbissa[1]

***Abstract*— The article describes a system for culture-aware human-robot verbal interaction, that constitutes the basis for designing culturally-competent robots for health-care, i.e., robots able to autonomously re–configure their way of acting and speaking, when offering a service, to match the culture, customs, and etiquette of the person they are asstisting. The article shows how culture-aware verbal interaction is tightly related to cultural knowledge representation and acquisition, by describing the methodological and technological solutions adopted, and showing in details one of the preliminary experiments performed to design a culturally-competent robot.**

I. Introduction

The work described in this article is part of a coordinated effort involving different partners in the EU and Japan to design a cultural-competent robot for assisting older persons in a care home or in their home [13]. The main objective of the CARESSES project[1] is to design robots able to match the culture, customs, and etiquette of the person they are assisting (Figure 1) while acting and speaking. Making culturally competent robots is key to address one of the major problems in assistive robotics: how to increase acceptability by being more sensitive to the user's needs, customs and lifestyle, thus producing a greater impact on the quality of life of users and their caregivers, and improving the system's efficiency and effectiveness. From the commercial perspective, cultural customization is crucial to overcome the barriers to marketing robots across different countries.

Borrowing the term from the Nursing Literature, "culturally competent robots" have been introduced in [2], by splitting the general problem in different subproblems requiring a multidisciplinary approach:

1) How to design guidelines enabling a robot to exhibit culturally competent behaviour [12]?
2) How to encode such guidelines with formal tools for knowledge representation?
3) How to use cultural knowledge to plan and execute sequences of actions that adapt to the cultural identity of the person?

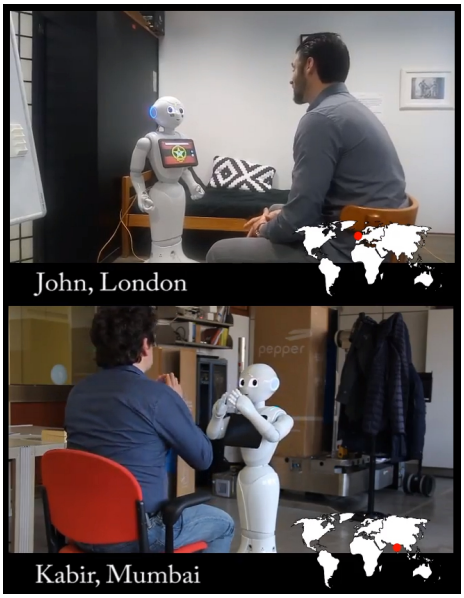

Fig. 1: SoftBank Robotics' Pepper and John: talking about Christmas. Pepper and Kabir: greeting with Namaste.

4) Which is the role of verbal interaction to acquire and satisfy the person's needs and requests, by suggesting her the options that better fit her cultural identity in terms of needs, customs, lifestyle?
5) How to acquire new knowledge by interacting with the person in order to avoid stereotyped representations?

Among the issues above, this work focuses explicitly on culturally-competent verbal interaction. To this aim, since verbal interaction is tightly related to cultural knowledge representation and acquisition, the article summarizes the basic principles adopted in order to design a Cultural Knowledge Base (CKB) that stores all the required information about the person and her cultural background.

Researchers are already investigating the role of cultural factors in robotics [17], for instance for designing robot's body movements [14], for child-robot interaction aimed at diabetes self-management [10], or for robot navigation [16]. Similarly, verbal human-robot interaction is receiving an increased attention in the last years, for example to teach robots high-level parametrized plans [5], enable them to learn high-level affordances [11], or to design robots acting as narrative companions providing access to past memories of older persons with cognitive / memory impairment [4]. Some studies such as [8], [1] aim at understanding the link between

*B. Bruno, R. Menicatti, C. Recchiuto equally contributed to this work.
[1]B. Bruno, R. Menicatti, C. Recchiuto, A. Sgorbissa are with DIBRIS, University of Genova, Via Opera Pia 13, Genova, Italy {barbara.bruno, antonio.sgorbissa}@unige.it, {roberto.menicatti, carmine.recchiuto}@dibris.unige.it
[2]E. Lagrue and A. K. Pandey are with SoftBank Robotics, 43, rue du Colonel Pierre Avia 75015 Paris France {elagrue, akpandey}@softbankrobotics.com
[1]CARESSES stands for Culture-Aware Robots and Environmental Sensor Systems for Elderly Support, caressesrobot.org

Preprint submitted to 15th International Conference on Ubiquitous Robots.
Received February 15, 2018.

the robot's dialog patterns and the cultural background of the user. However, to the best of our knowledge, there are no major investigations aimed at culturally competent verbal interaction for elderly care.

From the technical perspective, in spite of the many Cloud-based tools for Automatic Speech Recognition (e.g., Microsoft Speech Recognition, Google ASR, Nuance Freespeech[2]) and Natural Language Processing (e.g., DialogFlow – previously api.ai, Microsoft LUIS[3]), many problems still need to be solved towards the aim of achieving a natural interaction between the robot and the person. In a recent survey about verbal and non-verbal human-robot communication [9], a list of desiderata is presented:

(D1) Breaking the "simple commands only" barrier.
(D2) Multiple speech acts.
(D3) Mixed initiative dialogue.
(D4) Situated language and symbol grounding.
(D5) Affective interaction.
(D6) Motor correlates and Non-Verbal Communication.
(D7) Purposeful speech and planning.
(D8) Multi-level learning.
(D9) Utilization of online resources and services.

In the next Section we will show that, in spite of its present limitations, the system is able to meet some of the desiderata above. In particular, the system relies on cultural knowledge as the basis to break the "simple commands only" barrier (D1), allowing both the robot and the human to take the initiative to agree about the task to be performed or simply chit-chat about topics that may be entertaining for the person (D3). While doing this, the robot implements strategies to show attentiveness to the person's values, preferences, beliefs, and needs (D5), and to ultimately learn her individual preferences (D8). Finally, the system uses the knowledge acquired during conversation to adapt its sensorimotor behaviour during planning and execution (D6, D7 – this is not discussed in this article), and it makes extensive usage of online resources and services for speech recognition (D9).

Section II describes cultural knowledge representation. Section III introduces the solutions adopted for culturally–competent verbal interaction. Section IV describes a case-study. Conclusions follow.

II. Designing Cultural Competent Robots

A. Cultural Knowledge Base

The core of the CKB is an ontology properly structured in order to encode all elements that may play a key role in socially assistive robotic scenarios. The areas of knowledge considered at present time include:

- goals that the robot shall achieve and additional information about which goals are more likely to be relevant in different cultural contexts (e.g., does prayer or meditation plays an important role for the person? Should the robot assist the person in these activities?);
- actions that the robot shall execute and additional parameters about how to execute these actions in different cultural contexts (e.g., which is the right volume and distance while speaking to a person, or the right gesture to greet her?);
- cultural norms (e.g., are there any areas of the house that are off-limits for the robot? Does the situation change in different times of the day?);
- the environment, including furniture and objects and how they may differ in different cultural contexts;
- topics of conversation to talk with the person about her values, beliefs, habits, as well as additional information about which values, beliefs, habits, are more likely in different cultural contexts.

It shall be reminded that an *ontology* is a formal naming and definition of the types, properties, and interrelationships of the entities that exist for a particular domain of discourse [7]. The terminology defining the domain of discourse, containing general properties of concepts, is stored in the terminological box (TBox) of the ontology, whereas knowledge that is specific to individuals belonging to the domain is stored in the assertional box (ABox) of the ontology. Ontologies are interesting in that they allow non-technical users to easily[4] encode knowledge about the domain, which is a key property in cross-disciplinary contexts, such as ours.

According to the guidelines provided by experts in Transcultural Nursing, the CKB has been organized in order to deal with the necessity of representing knowledge both about cultural groups (e.g., at a national/ethnic level) and about individual persons (i.e., to avoid stereotypes). To this end, the CKB includes the following components:

- *Culture-agnostic knowledge*, a layer that stores the terminology (TBox) required to represent relevant concepts related to goals, actions, norms, the environment, etc. for all the cultures considered in the knowledge base (ideally *for all* the cultures of the world [3]);
- *Culture-generic knowledge*, a layer that stores the assertions (ABox) required to represent cultural information at national/ethnic level (i.e., the fact that an English woman is likely to celebrate Christmas and a Japanese woman is likely to have miso soup for breakfast);
- *Culture-specific knowledge*, a layer that stores the assertions (ABox) required to represent the unique cultural identity, preferences and environment of the assisted person (the fact that Mrs Smith, an English woman, loves Christmas and Mrs Yamada, a Japanese woman, prefers to have a quick continental breakfast instead of a traditional Japanese one);
- *Assessment & Adaptation*, an algorithm and a supporting Bayesian network for the discovery of culture-specific knowledge in light of culture-generic knowledge, e.g. relying on "educated guesses" to be confirmed through dialogue or autonomous robot observation (to explore the most likely hypotheses about Mrs Smith's

[2]https://msdn.microsoft.com/en-us/library/jj127860.aspx, https://cloud.google.com/speech/, https://www.nuance.com/index.html

[3]https://dialogflow.com/, https://www.luis.ai/

[4]For example using user-friendly tools such as Protégé: https://protege.stanford.edu/

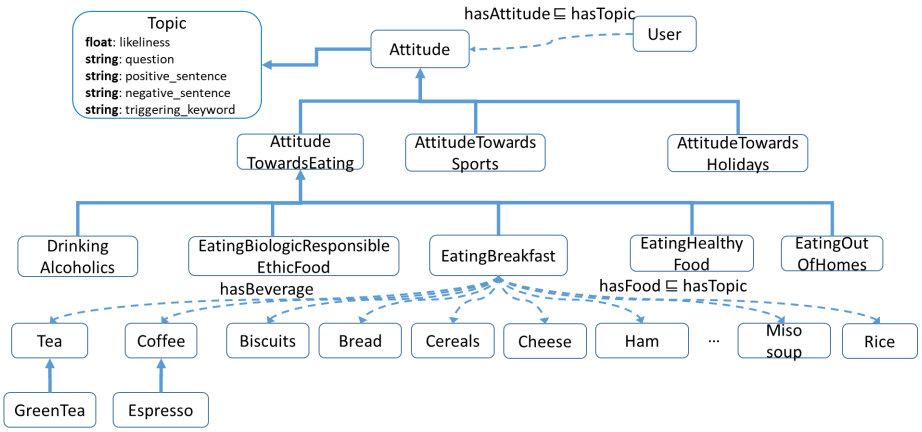

Fig. 2: The TBox corresponding to conversation topics.

and Mrs Yamada's habits, and possibly revise them).

For building the ontology, we adopt the OWL-2 language [6], with the usual definition of *Classes*, *object and data properties*, and *Instances*. Describing in details how the ontology has been shaped to include the required knowledge domains is out of scope of the article: consider however that the TBox and the ABox include a detailed definition of goals to be suggested and their parameters, actions and their parameters, mandatory and preferred norms, topics of conversation, and so on. On the opposite, the rest of this Section will describe the solutions adopted to represent culture-agnostic concepts (that are not related to any specific culture), culture-generic knowledge (that is related to a culture at national / ethnic level) and finally how to use culture-generic knowledge in order to infer culture-specific knowledge (i.e., beliefs, values, preferences, habits of an individual users) by making "educated guesses".

B. Avoiding stereotypes

Figure 2 shows a portion of the TBox defining the *topics of conversation* domain, intended as the collection of knowledge which is meant to keep the interest of the user and show the robot's attentiveness to the person's values, preferences, beliefs, etc. Specifically, topics of conversation have two main purposes. First, they play a key role to enable the system to acquire new knowledge about the person's preferences and attitudes and how these shall impact on the robot's behaviour. Second, they provide knowledge for "chit-chatting", under the intuition that the users might appreciate that the robot is familiar with the very same concepts that she is familiar with. Figure 2 focuses on the terms Attitude_towards_eating, Attitude_towards_sports and Attitude_towards_holidays. Specific habits and preferences are modelled with subclasses, such as Eating_breakfast, with object properties such as hasBeverage and hasFood relating the preference/habit to actual objects (e.g., drinks and food). As already stated, all concepts (e.g., drinks and food) that are typical in different cultures are represented in the TBox, whichever the nationality of the user, to avoid stereotypes.

Two important concepts should be outlined. First, the class User in Figure 2 represents the person that the robot assists, that is related to all the other concepts in the TBox by ownership (objects, furniture), preferences, habits, beliefs, etc. Second, all classes in the ontology are derived from a superclass named Topic (i.e., topics that the robot is capable to speak about) with the following properties:

- likeliness. In the culture-generic ABox layer describing the culture of the person at national/ethnic level, the data property likeliness is used to encode the priori probability that an assertion in the ABox holds for a person, given that we know that she belongs to that culture (e.g., the probability that an English person celebrates Christmas is high, as well as the probability that an old Japanese person has miso soup for breakfast, and therefore they will correspond to a high likeliness value). In the culture-specific ABox layer describing the individual attitude of Mrs Smith and Mrs Yamada, likeliness encodes the evidence acquired or the posteriori probabilities acquired through interaction (e.g., 1 if Mrs Smith or Mrs Yamada have a positive attitude, 0 if they have not a positive attitude, a value in-between in case of uncertain knowledge).
- question. In the culture-generic ABox layer, the data property question contains the questions (randomly chosen for entertainment purposes) the robot may use to ask the person about her individual attitude: e.g., "Do you celebrate Christmas?" for instances of the class Attitude_towards_Holidays (this question may be used to verify if Mrs Smith celebrates Christmas, which has a high likeliness to receive a positive answer in case of an English person). In the culture-specific ABox, a similar set of questions may be used to revise previously acquired knowledge, especially in presence of uncertain knowledge (i.e., likeliness between 0 and 1, see above).
- positive_sentence and negative_sentence. In the culture-generic and in the culture-specific ABox layers, the data properties positive_sentence and negative_sentence contain sentences that the robot can use to express, respectively, a positive or a negative

Preprint submitted to 15th International Conference on Ubiquitous Robots.
Received February 15, 2018.

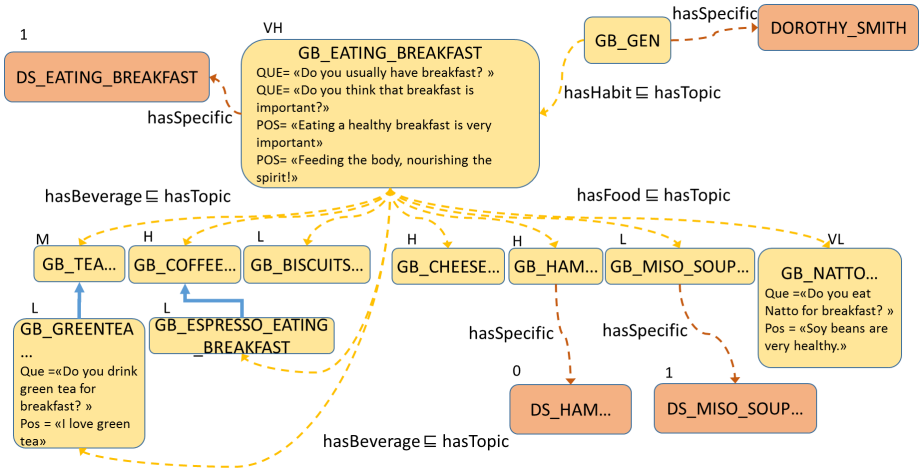

Fig. 3: The ABox corresponding to breakfast habits: British culture-generic (yellow, GB prefix); Dorothy Smith culture-specific (orange, DS prefix).

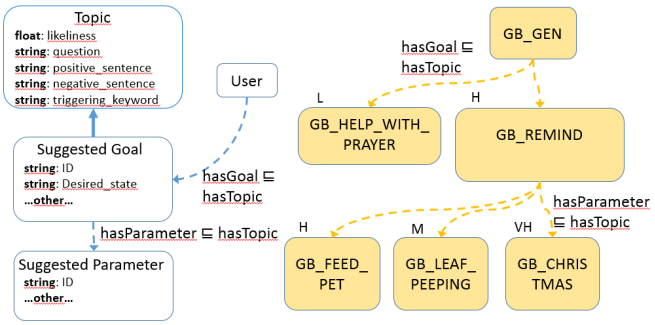

Fig. 4: The TBox (left) and ABox (right) corresponding to goals: British culture-generic.

attitude related to assertions in the ABox (e.g. a positive_sentence corresponding to an instance of the class CelebratingChristmas might be "On Christmas, people exchange gifts and share happiness with each other!", while a negative_sentence for an instance of WatchingRugby, borrowed from Oscar Wilde, could be "Rugby is a good occasion for keeping thirty bullies far from the center of the city"). All sentences, and especially negative ones, should be checked by experts, to ensure that they are ethically and culturally sound.

- triggering_keywords. It denotes one or more keywords that, if recognized by the robot while interacting with the person, trigger a conversation about the corresponding topic (e.g., if the person mentions "food" or "eating" or "cuisine", this may start chit–chatting about food; if the person mentions "Italian food", this may start chit–chatting about Italian cuisine, see Section III).

Figure 3 shows the portion of the culture-generic (yellow boxes with GB prefix and arrows) and culture-specific (orange boxes with DS prefix and arrows) ABox layers related to breakfast habits and preferences of a person named Dorothy Smith. In the Figure, boxes denote instances of classes, dotted lines denote assertions of object properties. Data properties (e.g., question) appear within the box of the instance they refer to, while likeliness values appear on the top-left corner of the instance they refer to. In the culture-generic ABox layer, likeliness values are denoted with literals instead of numbers (0.05 for Very Low, 0.1 for Low, 0.2 for Medium, 0.4 for High, 0.7 for Very High). The reason for this choice is practical: while it is very difficult (if not impossible) to obtain precise a priori probabilities from statistical analyses, it is much easier to infer approximate, qualitative values from the vast (but often inhomogeneous) corpus of information in the literature and on the web.

Notice also the property hasSpecific, that relates instances in the culture-generic ABox layer with the corresponding instances in the culture-specific ABox. The Figure shows the preferences of Mrs Smith, that may obviously differ from the preference of English people: Mrs Smith has definitely the habit of having breakfast (likeliness $= 1$), but she does not have ham for breakfast even if this appears to be quite common for English people (likeliness $= H$ for the instance GB_HAM... in the culture generic ABox layer; likeliness $= 0$ for the instance DS_HAM... in the culture specific ABox layer). Instead, the CKB tells us that Mrs Smith has miso soup, even if this choice is not very common for English people (likeliness $= L$ for the instance GB_MISO_SOUP... in the culture generic ABox layer; likeliness $= 1$ for the instance DS_MISO_SOUP... in the culture specific ABox).

Figure 4 shows a portion of the TBox and the culture-generic ABox layer related to suggested goals and their parameters. On the right, two instances of the class SuggestedGoal are shown, corresponding to possible goals the robot will suggest and is ready to receive, as well as their likeliness value for an English person, and corresponding suggested parameters. Please notice that GB_HELP_WITH_PRAYER has a lower likeliness value than GB_REMIND (the second being a very likely goal for older persons belonging to all cultures), and the parameter GB_FEED_PET (having pets is very common for older English people) has a higher likeliness value than

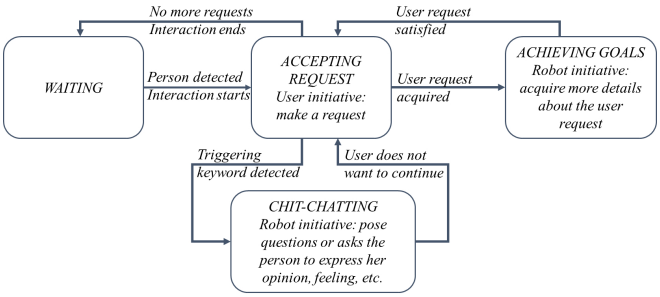

Fig. 5: Model of the overall system's behaviour.

GB_LEAF_PEEPING (an activity more common in Japan than in the UK). Notice also that the class SuggestedGoal is derived from Topic, and therefore the robot is enabled, in principle, to talk about the goals that are more or less relevant for the person, and update their likeliness value accordingly in the culture-specific layer of the ABox.

The next Section will show how these elements are sufficient to implement an algorithm for assessment and adaptation aimed at acquiring culture-specific likeliness values (i.e., evidence encoded in the culture-specific ABox layer describing the person) starting from culture-generic knowledge at national / ethnic level (i.e., a priori probabilities encoded in the culture-generic ABox layer).

III. Mixed initiative, culturally competent verbal interaction

A. General ideas

When designing culturally-competent, socially-assistive robots, there is a range of different situations in which verbal interaction between the robot and the person plays a key role. In this work we classify different typologies of dialogue patterns into the following prototypal situations:

1) the dialogue is aimed at acquiring a new request from the person or acquiring additional details to satisfy the user's request (i.e., D1 in Section I);
2) the dialogue is aimed at conversating about a given topic (under the intuition that the person might appreciate that the robot is familiar with the very same concepts that she is familiar with, D5) and learning the person's preferences, beliefs, habits (in short, acquring culture specific, individual level knowledge starting from culture generic, national level knowledge, D8).

Dialogue patterns belonging to both classes are possibly composed of different speech acts [15] in which either the robot or the person take the initiative (D3). Specifically, dialogue patterns belonging to class (1) are ultimately lead by the person giving commands to the robot; the robot may ask questions but with the only purpose of acquiring contextual information enabling it to better meet the person's requests. Dialogue patterns belonging to class (2) are ultimately led by the robot that uses existing cultural knowledge to acquire new knowledge, and are automatically composed starting from the values of data properties question, positive, and negative of instances in the CKB (Section II-A). The role of the persons is mainly to reply to the robot's questions, or express her opinion about what the robot says. Needless to say, a seemless integration between either classes of dialogue patterns is required, if we want to avoid an overall fragmentary behaviour that the person may find boring, puzzling or bizarre.

According to this rationale, we model the overall system's behaviour as a state machine with four states (Figure 5).

1) *WAITING*: The robot is turned on and waiting for somebody to enter the room. When the robot detects a person in the room and/or a person asks the robot that she wants to start the interaction, the robot switches to the state *ACCEPTING REQUESTS*.
2) *ACCEPTING REQUESTS*: The person leads the dialogue: the robot is waiting for a request (possibly after providing a list of possible options), that can either be given through free speech or through a tablet.
   - If the person makes a request that can be satisfied by the robot, the robot switches to the state *ACHIEVING GOALS*.
   - If the person pronounces a triggering keywords, the robot jumps to the corresponding topic of conversaton, by switching to the state *CHIT-CHATTING*.
   - If the person does not make any request for a long time, or says explicitly that she wants to terminate the interaction, the robot switches to the state *WAITING*.
3) *ACHIEVING GOALS*: If required, the robot leads the dialogue to acquire details about what the person actually wants, i.e., the exact identity of another person to video-call, the exact name of a song to be played, and so on. Then the robot performs the sensorimotor actions required to achieve the goal (not relevant to the present discussion). After achieving the goal, the robot goes back to the state *ACCEPTING REQUESTS*[5]
4) *CHIT-CHATTING*: The robot leads the dialogue by starting a conversation about one of the topics that are associated with the triggering keywords, by posing questions to the person and giving her the opportunity to express her opinions and needs. The robot periodically asks the person if she wishes to continue chit-chatting: if the answer is negative, the robot switches back to the state *ACCEPTING REQUESTS*.

The reader may be confused by the fact that the dialogue for acquiring additional details about the person's request is part of the state *ACHIEVING GOALS* instead of being part of *ACCEPTING REQUESTS*. This depends on how the behaviour of the robot is modelled as a sequence of elementary actions for the sake of planning and real-time execution: in our system, the state *ACCEPTING REQUESTS* is implemented in software as a specific action whose

[5]In principle, it is possible to enter the state *ACHIEVING GOALS* asynchronously, i.e., depending on some external events (e.g., the robot needs to remind the person to take a medicine at a given time). Event-based and time-based goals are not considered for sake of brevity.

purpose is to acquire the intentions of the person among a set of possible options, whereas the state *ACHIEVING GOALS* is ultimately implemented by picking out one among a set of actions, each including a contextual dialogue that is purposely aimed at acquiring details about a specific goal. For instance, suppose that the system is initially in the state *ACCEPTING REQUESTS*: after the person has expressed her intention to set a reminder to call the doctor in the afternoon, the system switches to the state *ACHIEVING GOALS*, and a dedicated action is executed that first acquires details about the activity to be reminded and the time of the reminder, and then sets the reminder accordingly. This solution is very efficient to drive the dialogue along a predefined path to acquire all the necessary information to achieve a goal: for instance, when the robot is talking with the user in the state *ACHIEVING GOALS*, it is not possible to jump to *CHIT-CHATTING* even if a triggering keyword is detected (which increases reliability but at the price of a reduced versatility).

B. The role of cultural knowledge

Cultural knowledge encoded in the CKB plays a key role in the different states depicted in Figure 5 as follows.

ACCEPTING REQUESTS. When in this state, cultural knowledge provided by the CKB enables the robot to suggest to the user which goals are more suitable to the present situation, and/or to visualize on the tablet the most likely options. Giving the right suggestions is important, since the goals that the robot may accept are potentially infinite (the only limit being the creativity of the developers). There is nothing that prevents an English woman to ask the robot to perform a task that (in the developers' intentions) has been introduced to fit the needs of a different culture (e.g., showing the Qibla, the direction that should be faced when a Muslim prays during salah prayers). Suggestions are particularly relevant since the tablet's screen can only display a limited number of options at the same time: it is therefore desirable that the options displayed first are those more likely to be selected, i.e., those which are more coherent with the cultural profile of the person and hence have a higher likeliness in the CKB (see Figure 4 as an example).

ACHIEVING GOALS. The role of cultural knowledge is very similar as in the previous case. Suppose that the person has just made the request of listening to some music: after the main goal has been selected, an additional dialogue between the robot and the person may be required to acquire details about how the request shall be satisfied (which kind of music?). As usual, there is nothing that prevents an English woman to listen to traditional Arab music, even if she has never expressed her interest in that kind of music before: however, cultural knowledge enables the robot to suggest to the user the most likely options according to his/her cultural profile, i.e., those which have a higher likeliness in the CKB (Figure 4). For sake of completeness, it shall be mentioned that – when in this state – cultural knowledge plays also a key role to decide which actions to execute in order to achieve a goal (e.g., should the robot greet the person by waving hands, bowing or doing Namaste?), which are the parameters of these actions (e.g., which are the optimal volume and distance from the listener during interaction?), and which are the norms to be respected (e.g., are there any areas of the house that are off-limits for the robot? Does the situation change in different times of the day?). Since this article deals only with verbal interaction, the role of cultural knowledge in choosing the right actions, parameters, and norms is not considered in the following.

CHIT-CHATTING. The role of cultural knowledge is to drive the conversation to topics that the person is more familiar with and, at the same time, to explore those topics for which individual preferences are not known yet, by giving priority to topics for which a positive attitude is expected: e.g., the robot may not know, initially, if Mrs Smith likes cheese and ham for breakfast (Figure 3) and celebrating Christmas (Figure 4), but it will investigate these possibilities before exploring if she likes miso soup and she celebrates the Japanese Birthday of the Emperor. In short, when in this state, the likeliness of instances in the CKB is the basis for acquiring new cultural knowledge through "educated guesses", and update the likeliness of the corresponding instances in the culture-specific layer of the ABox accordingly.

The reader may notice that updating likeliness values (when in the *CHIT-CHATTING* state) will ultimately have an impact on the most likely options to suggest in the *ACCEPTING REQUEST* and *ACHIEVING GOALS* state (as well as in the actions to be performed to achieve such goals, their parameters, and cultural norms). This approach allows for avoiding *stereotyped representations* of cultures: culture-generic, national level knowledge is not straightforwardly applied to the person by assuming that she has the same traits of the cultural group she belongs to. On the opposite, culture-generic knowledge is only used to speed-up the acquisition of culture-specific knowledge about the individual through dialogue (or observation, not discussed here).

IV. Case-study: Pepper and Kabir

A. Technological solutions

The methodological approach described in the previous Sections has been adopted in order to embed with cultural-competent behaviour the humanoid robot Pepper developed by SoftBank Robotics (Figure 1). Verbal interaction is managed by a module referred to as *Conversation Module*. Depending on the state of the system (i.e., *ACCEPTING REQUESTS*, *ACHIEVING GOALS*, or *CHIT-CHATTING*) the modules acquires speech or text from the person and parses it using different technological solutions, and ultimately returns outputs which have different meanings:

1) when the system is in the state *ACCEPTING REQUESTS* or *ACHIEVING GOALS*, the output is interpreted as a *User request*, i.e., short-term information describing goals to be achieved by the robot;
2) when the system is in the state *CHIT-CHATTING*, the output is interpreted as a *User statement*,i.e., long-term information about the person's preferences, beliefs, habits, that shall be used to update the CKB.

Preprint submitted to 15th International Conference on Ubiquitous Robots.
Received February 15, 2018.

In order to meet the requirement introduced in the previous Section, the base component of the Conversation Module is a tablet-based HRI interface used by the robot to display questions and collect answers through its touch-screen. In spite of its simplicity, the interface is meant to be a very robust and reliable method for the user to communicate with the robot, that is guaranteed to work even in those conditions when speech-based interfaces are likely to fail (e.g., in noisy environments). Even if most of the persons will prefer free-speech rather than using the tablet, still the questions and answers that are visualized on the tablet's screen (a set of options / keywords to choose from) may play a very important role to guide the user to make reasonable requests, and using the most appropriate keywords to place such requests (even when the person is speaking instead of using touch). Finally, this solution may be very important for persons with auditory problems. As already stated, the tablet's screen can only display a limited number of options at the same time, and therefore cultural knowledge about the person plays a key role to display those options that are more likely to be selected (the person can browse a menu if the desired option is not among the suggested ones).

The second component of the Conversation module is a tool for speech and natural language processing used by the robot to collect requests, statements, and answers based on free-speech interactions with the person. The component relies on the usage of Cloud-based solution for Automatic Speech Recognition (either Google ASR or Nuance Freespeech) to convert an audio file to text, and then feeds it to an on-board system for keyboard extraction and processing (which constitutes the basis for guessing the person's intentions, and to start a corresponding dialogue). This tool is mainly used when the robot is in the state *ACCEPTING REQUEST*, to the end of detecting triggering keywords that may cause a transition to the *ACHIEVING GOALS* state or to the *CHIT-CHATTING* state. Concretely, the keywords to be extracted are defined in a JSON file automatically created by querying the CKB (where the triggering keywords are stored, see Section II-A). The format of the file allows for specifying multiple keywords that belongs to an equivalence set (i.e., keywords that shall be interpreted in the same way by the robot) and for defining both a main keyword and a second-level keyword. For instance, an equivalence set may include all actions that describe motion from one place to another $\{go, come, accompany, \ldots\}$, or even arbitrarily define different kitchen features as synonyms, such as $\{kitchen, dining\ table, fridge \ldots\}$. Then, if the person pronounces the sentence "I need to clean the dining *table*, would you *come* with me?", the robot will be able to recognize a main keyword in the first set and a second-level keyboard in the second set, and respond properly by heading to the kitchen[6]. In case that the robot cannot understand what the person says, the system proposes to switch to a tablet-guided dialogue (improving reliability at the price of a reduced naturalness and versatility).

The third component of the Conversation Module is the off-the-shelf tool for keyword recognition included in the standard NAOqi API. This tool has limited functionalities with respect to Google ASR and Nuance Freespeech, but it turns out to be sufficiently reliable when the person replies by picking out a keyword in a small set. However, it does not require an Internet connection and therefore it can be used under any network condition. This tool is used when the robot is in the state *CHIT-CHATTING*, i.e., when the robot is leading the dialogue and the person is given the possibility to express her opinions or needs using multiple choice answers (e.g., *yes*, *no*, *sometimes*, *never*, *I love it*, ...).

B. Verbal Interaction Step-by-step

The system has been extensively tested and more than 12 videos have been recorded, showing the behaviour of the system when interacting with an actor (experiments with older persons living in care homes will be performed in the third year of the project). Figure 1 shows the robot Pepper interacting with John, London, and Kabir, Mumbai.

Table 6 reports transcripted sentences[7] during the interaction between Pepper and Kabir. White rows correspond to Pepper speaking or performing a task; gray rows correspond to Kabir speaking. Pepper is initially in the state *WAITING* (W: line 1). After greeting the person, it switches to the state *ACCEPTING REQUESTS* (AR: 3), and it suggests a goal (SG: 3) which may be very relevant in the Indian culture (i.e., helping the person to choose dresses), but the person is not interested. Instead, Pepper detects the keywords "feel" and "well" (KD: 5), switching to *CHIT-CHATTING* (7: CC): when in this state, Pepper asks questions (Q: 7 and 11), expresses positive (P: 9, 13, and 15) or negative sentences (N, not shown), and - occasionally - leaves Kabir the possibility to speak (W: 9 and 15).

In the second part, Pepper detects the keyworkd "breakfast" (KD: 19), and starts asking Kabir about his breakfast habits. When asking questions, Pepper explores possibilities that are associated with a higher likeliness value in the culture-generic ABox layer, e.g., by asking "Do you usually have Indian food for breakfast?" (Q: 21). In this case, Pepper receives a positive answer from Kabir, and this answer is used to update the culture-specific ABox layer accordingly (U: 25). Finally, Kabir expresses the desire to listen to music, therefore switching to *ACHIEVING-GOALS* (AG: 33): Pepper acquires additional information about the music genre that Kabir prefers (AP: 33 and 35), and starts playing it.

Finally notice that, when Kabir is given the opportunity to talk freely (W: 9 and 15), Pepper waits for him to end talking, but it does not parse his speech. Updating cultural knowledge in the CKB is a process that needs reliable information, which can only be achieved by posing a direct question to the person and collecting a direct answer.

[6]Cloud-based systems for Natural Language Processing such as DialogFlow are tested as well as alternative solutions. However, this does not have an impact on the role played by cultural knowledge to determine dialogue patterns, and therefore they are not discussed here.

[7]See a video showing the full experiment with additional dialogues and actions: https://youtu.be/Nz7vHCjcJfE. A more complex experiment with persons simulating different cultural groups: https://youtu.be/RlKtrkmP5us

Preprint submitted to 15th International Conference on Ubiquitous Robots.
Received February 15, 2018.

N.	Sentence	St.	Ev.
1	Namaste Kabir.	W	
2	Namaste Pepper.		
3	If you need me I am here. As an example, I may help you to get dressed.	AR	SG
4	No, I don't feel very well this morning, Pepper.		
5	I heard "feel" "well". Let's talk for a while	AR	KD
6	Ok		
7	I understand that you do not feel very well today. Is it true?	CC	Q
8	Yes it's true		
9	It may happen that we do not feel very well. It also happens to robots, sometimes. But today I am feeling well, so maybe I can help you.	CC	P, W
10	Thank you. How can you help me? Let me understand. You are very nice.		
11	I understand. Do you feel a little sick today?	CC	Q
12	Yes, I feel a little sick. Yes.		
13	I see. Today you do not feel very well.	CC	P
14	No.		
15	But please do not worry. You can explain everything to the nurse or the doctor during their visit. In the meanwhile, you can talk with me about what you feel, if you wish.	CC	P, W
16	Okay, I feel a little pain here, here in my stomach. Maybe I ate to much yesterday, in the evening. You know?		
17	I understand, do you want to keep talking about this?	CC	
18	No, no. I want to have breakfast now.		
19	I heard breakfast. Let's talk together.	AR	KD
20	(Kabir does not say anything)		
21	Having a healthy breakfast is very important to start your day. Do you usually have Indian food for breakfast?	CC	P, Q
22	Usually.		
23	Sorry I do not understand. Do you usually have Indian food for breakfast?	CC	Q
24	Always, I have to say.		
25	Indian food is very healthy for breakfast. Do you usually have chapati for breakfast?	CC	U, P, Q
26	I always have chapati for breakfast. I love it!		
27	Chapati is perfect for breakfast. I would like to be able to cook it myself.	CC	U, P
28	That is not easy.		
29	Do you want to keep talking about this?	CC	
30	No, no thank you. Do you know what? I would like to listen to some music now.		
31	Okay, please just give me a second. I am happy to play some music.	AR	KD
32	Okay, I am here. I am not going anywhere.		
33	Which is your favourite music genre?	AG	AP
34	I think blues.		
35	Blues.	AG	AP
36	Yes, blues.		
37	(loading a YouTube Video on the screen)	AG	
38	Is it happening something here? Oh, I think... I think I have to press here, right? (Kabir touches the screen)		
39	(music plays)	AG	
40	I like this song, Pepper. But maybe I don't like it so much. Thank you. You know what? I would like to see the weather forecast, now.		

Fig. 6: Pepper (white rows) and Kabir (gray rows) talking.

The analysis of transcripts (and the related videos) allows for observing all the key principles discussed throughout the article in action, showing the natural flow of the discourse generated by the interaction, even when the system switches between different states in Figure 5.

V. Conclusions and Future work

The article has described preliminary work towards culture-aware verbal interaction, by showing a system meeting some of the desiderata for verbal human-robot communication. The system have many limitations: among the others, we need to extend the system to give more importance to contextual information in the dialogue, i.e., enabling the robot to talk about what is happening "here" and "now" (either the information is acquired through observations or dialogues) in a culturally competent way.

Also, no systematic study aimed at user-evaluation has been performed yet: a test study with older persons living in care homes will be performed starting from April 2019.

Acknowledgement

This work has been supported by the European Commission H2020 Programme under grant agreement No. 737858, and from the Ministry of Internal Affairs and Communication of Japan. The authors wish to acknowledge all CARESSES researchers and their teams, including Prof. I. Papadopoulos (Middlesex Univ.), Prof. A. Saffiotti, Prof. F. Pecora (Örebro Univ.), Prof. N. Y. Chong, Prof. Y. Lim (JAIST), Prof. H. Kamide (Nagoya Univ.), Dr. J. Lee (Chubu Univ.), Dr. C. Papadopoulos, Prof. G. Randhawa (Univ. of Bedfordshire), Dr. S. Kanoria, Mr. L. Merton (Advinia Health Care).

References

[1] S. Andrist, M. Ziadee, H. Boukaram, B. Mutlu, and M. Sakr, "Effects of culture on the credibility of robot speech: A comparison between english and arabic," in *HRI 2015*, 2015, pp. 157–164.

[2] B. Bruno, N. Y. Chong, H. Kamide, S. Kanoria, J. Lee, Y. Lim, A. K. Pandey, C. Papadopoulos, I. Papadopoulos, F. Pecora, A. Saffiotti, and A. Sgorbissa, "Paving the way for culturally competent robots: A position paper," in *Proc. 26th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN 2017)*, Aug 2017, pp. 553–560.

[3] M. Carrithers, M. Candea, K. Sykes, M. Holbraad, and S. Venkatesan, "Ontology is just another word for culture, motion tabled at the 2008 meeting of the group for debates in anthropological theory," *Critique of Anthropology*, vol. 30, no. 2, pp. 152–200, 2010.

[4] P. F. Dominey, V. Paléologue, A. K. Pandey, and J. Ventre-Dominey, "Improving quality of life with a narrative companion," in *Proc. 26th IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN2017)*, 2017, pp. 127–134.

[5] G. Gemignani, E. Bastianelli, and D. Nardi, "Teaching robots parametrized executable plans through spoken interaction," in *Proc. 2015 Int. Conf. on Autonomous Agents and Multiagent Systems*, ser. AAMAS '15, 2015, pp. 851–859.

[6] W. O. W. Group *et al.*, "{OWL} 2 web ontology language document overview," 2009.

[7] N. Guarino *et al.*, "Formal ontology and information systems," in *Proceedings of FOIS*, 1998, pp. 81–97.

[8] M. Makatchev, I. Fanaswala, A. Abdulsalam, B. Browning, W. Ghazzawi, M. Sakr, and R. Simmons, "Dialogue patterns of an arabic robot receptionist," in *Proc. 5th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*, 2010, pp. 167–168.

[9] N. Mavridis, "A review of verbal and non-verbal humanâĂŞrobot interactive communication," *Rob. Aut. Syst.*, vol. 63, pp. 22 – 35, 2015.

[10] A. Neerincx, F. Sacchitelli, R. Kaptein, S. van der Pal, E. Oleari, and M. A. Neerincx, "Child's culture-related experiences with a social robot at diabetes camps," in *Proc. 11th ACM/IEEE Int. Conf.e on Human Robot Interaction*, ser. HRI '16, 2016, pp. 485–486.

[11] A. K. Pandey and R. Gelin, "Human robot interaction can boost robot's affordance learning: A proof of concept," in *2015 International Conference on Advanced Robotics (ICAR)*, 2015, pp. 642–648.

[12] I. Papadopoulos, *Transcultural health and social care: development of culturally competent practitioners*. Elsevier Health Sciences, 2006.

[13] I. Papadopoulos, A. Sgorbissa, and C. Koulouglioti, "Caring robots are here to help," *Nurs Stand.*, vol. 31, no. 51, pp. 18–20, 2017.

[14] M. Rehm, "From multicultural agents to culture-aware robots," in *Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 431–440.

[15] J. R. Searle, *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press, 1969.

[16] X. T. Truong, Y. S. Ou, and T. D. Ngo, "Towards culturally aware robot navigation," in *Proc. IEEE Int. Conf. on Real-time Computing and Robotics (RCAR2016)*, 2016, pp. 63–69.

[17] S. Šabanović, C. Bennett, and H. Lee, "Towards culturally robust robots: A critical social perspective on robotics and culture," in *Proc. Workshop on Culturally Aware Robots, 9th Int. Conf. on Human-Robot Interaction*, ser. HRI '14, 2014.