Robot Experience Stories: first person generation of robotic task narratives in SitLog

Jorge Garcia Flores, Iván Meza, Émilie Colin, Claire Gardent, Aldo Gangemi, Luis Pineda

Undefined 1 (2009) 1–5
IOS Press

Robot Experience Stories: first person generation of robotic task narratives in SitLog

Jorge Garcia Flores [a], Iván Meza [b], Émilie Colin [c], Claire Gardent [c], Aldo Gangemi [a] and Luis Pineda [b]
[a] *LIPN-CNRS-Université Paris 13, 99 av. Jean-Baptiste Clément, 93430, Villetaneuse, France*
E-mail: {garciaflores,gangemi}lipn.univ-paris13.fr
[b] *IIMAS-UNAM, Circuito Escolar 3000, Cd.Universitaria, 04510 CDMX, Mexico*
E-mail: ivanvladimir@turing.iimas.unam.mx, lpineda@unam.mx
[c] *CNRS-LORIA, Campus Scientifique BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France*
E-mail: {claire.gardent,emilie.colin}@loria.fr

Abstract. What if service robots could tell the story of the task they've just realized like a story? The aim of our work is to provide service robots with natural language capabilities to produce a Robot Experience Story for its human interlocutors. *Robxp stories* are narratives composed of the robot's holistic perception of a recently performed task: navigation, visual perceptions and action descriptions. We contribute with a *narrate* dialog model specifying the composition of situations necessary for a service robot to transform its task history record into a narrative knowledge representation. We provide SitLog algorithms allowing to analyze the robot's situation and behaviors sequence in order to generate a *robxp story* of the task. Both the dialogue model and the algorithms can be embedded as compositional behaviors in any other SitLog task structure. We instantiated our model into the Golem service robot framework on an experimental task. We believe Robxp stories generation could be integrated as a standard behavior for more complex service robot tasks.

Keywords: robot experience stories, SitLog, service robot, narrative generation

1. Introduction

When humans get back home, there is usually another human there to ask: *"How was your day?"* Our work aims at providing robots with the natural language (NL) capabilities to answer a similar question, not exactly about their day, but about their experience performing a certain task. From the point of view of human-robot interaction, a desirable skill for a service robot would be to provide its users with a natural language summary of any task it has recently performed, talking about the actions performed, the locations where the task was executed and its results. Up to date there are only two robots [27,16] able to narrate in natural language the result of their activity, the usual approach being a record of the robot's actions into a set of log files created for debugging purposes, and not well adapted to tell the story in a friendly way. This narrative behavior could be applied to the surveillance of robot's activity in an autonomous environment or to improve the communication quality with the robot's interlocutors.

In this work we introduce *Robot Experience Stories* (or *robxp stories*) as the first person natural language tale of a task performed by a robot. Robot experience storytelling is a higher level extension of the *verbalization* [27] process beyond the spatial domain, integrating visual, cognitive and narrative perception of the robot into the familiar, user friendly form of a story. Instead of a natural language enumeration of the robot's locations, we generate a story of the robot's task resolution based on the robot's visual perception, action experience and navigation history. We define robxp stories based on three types of actions:

Dynamic Aations performed by the robot while executing a specific task

Visual actions which senses the surrounding objects, persons and reads textual messages in the scene
Spatial actions related to the navigation and spatial movement of the robot. Robxp stories discursive structure is highly based in spatial actions.

Our work is developed in the context of the Golem project: a three generation family of service robots based on the evolution of the same conceptual model and functional structure [21]. The Golem robot performs several tasks such as being a clerk in a super market [23], being a waiter in a restaurant [25] or playing Marco-Polo [17]. The core of the robot consists on a set of compositional behaviors defining the structure of the task that the robot must perform, which is specified using the SitLog robot programming language [22]. Golem has a large library of behaviors, some of them are simple such as recognizing an object, grabbing an object with its arm, identifying a person or locating a sound source, while others are more complex like looking for an object or a person inside the house, navigating to a point or follow a person through the house.

Other than the concept of Robot Experience Story, this work contributes with a definition of a new *narrate* behavior specified in SitLog, which translates the robot's task history into a coherent story describing what the robot has just done, by means of a narratology inspired model [24,6] of a robxp story, taking into account the robot's movements, activities and perceptions in a holistic way. Unlike prior work on robot verbalization, which focuses exclusively on the robot's spatial navigation [27], our proposal considers navigations as the main discourse structure axis, to be enriched with visual behavior context, dynamic behaviors narrative of the tasks performed by the robot and a *read* behavior, where the robot is also sensible to the textual messages found in its trajectory.

The article is organized as follows: Section 2 discusses previous work on robotics, verbalization and storytelling; Section 3 defines the concept of robxp stories and proposes a *narrate* dialogue model in the framework of Golem's concept and functional structure; Section 4 presents the generation algorithms in Sitlog; Section 5 introduces two experimental implementations of robxp stories generation, including the following robxp generated story generated by an in vitro robotic task execution[1]:

> *"I woke up at the hall there was a sign on the door there it said COMPUTER SCIENCE DEPARTMENT. After that I wanted to walk towards the office of Ernesto but couldn't, after that I walked towards the office of Ivan to a desk there it said SOFA: Random Forests Regression for the Semantic Textual Similarity task I took a bottle then I moved to a desk I leaved the bottle I noticed a cluttered desk with a laptop there it said "sausenc &aceta para consejoros - ASO". Afterwards I moved to the meeting room to a table I saw a man standing in a room then I went to a door there it said GOLEM. After that I moved to the laboratory to a desk I noticed a desk with a computer and a chair afterwards I finished."*

Finally sections 6 and 7 respectively provides some perspectives and our conclusions.

2. Related work

Although there has been a great interest in providing robots with language capabilities [15], it is until recent years that the community has focused on producing a natural language narrative of a robot's activity. Experience storytelling behavior was introduced simultaneously by Rosenthal et al. [27] and Meza et al. [16], the former calling it "*verbalization*" while we then called it "*spatial experiences*". However, we believe that "*robot experience stories*" meaning conveys better the high degree of structure, coherence and linguistic skills necessary to summarize a robot's performance around the complex discursive structure of a story, and furthermore, it follows an active research line in storytelling with robots [7,18,11,12,13,9].

The main distinction between Rosenthal et al. [27] work and and the present work is that in their approach, the robot generates a *verbalization* text describing spatial perceptions only (location, navigation, position) while our method includes not only the spatial perception description, but visual perceptions (objects and persons in a room, for instance), action descriptions (success or failure of the tasks performed by the robot) and textual message decoding (the robot can read textual messages like posters or signs found in the visual environment of its navigation path). While Rosenthal's descriptions are focused on the space navigated by the robot, ours prefer to concentrate in producing a story out from the robot's holistic perceptions of recently performed tasks.

Jensen et al. [10] generate raw phenomenological command-like descriptions from sensor robots distributed in a closed space. Their robots are able to track person's movement in the space and produce NL command-like phrases like "*person1 is coming close*

[1] Notice that the narration model does not includes a human level punctuation, and here it is presented as outputted by the system.

to person2", but they don't pretend to get them together in a text. Recent work on NL processing and robotics is more focused on the grounding problem, which is concerned with grounding NL instructions to the right symbolic representations in the robot's knowledge base. Paul et al. [19] present a grounding approach based on probabilistic predicates, where NL instructions are associated to first-order logic like expressions that can take into account past temporal associations to simulate context understanding by the robot. Walter et al. [30] improves previous approaches with NL information about the accuracy of a map's semantic descriptions. Marge et al. [14] present a linguistic study of human-robot dialogues where they show how the users tend to be very precise about space details, including metric informations in their first dialogues, but then decrease and choose landmarks over time. While all these works focused on a different task (from NL instructions to KR), some grounding methods could be used to walk the inverse path (from the robot's KR to NL). The interesting thing here is that most of these works base their proposal on the role of the spatial information plays, which follows Rosenthal et al. [27] intuition that space is crucial in a robot's discourse. However, none of these prior works give any attention to non spatial perceptions, like we do in the present work.

Apart of the robotic-NL interaction, there is work from the image and video processing field worth mentioning here because of the crucial role that visual perceptions have in our robxp stories generation approach. As it is described in section 5, the robot in which we implement and test, follows a trajectory and stops at certain points, where a picture is taken and described in order to include its description in the robxp story. Recent work on computer vision and image description generation makes available these kind of descriptions [29,26]. However, even if for the moment our work's scope is limited to include photographs of descriptions, a promising perspective would be to include video summarization methods to the overall robot trajectory video [28,32].

3. Robot Experience Stories

We define a robxp story as a narrative $N = \langle S, E, A \rangle$ binding a set of situations $s = \langle t, l, c_1 \dots c_n \rangle \in S$ where t represents a state of a set of characters $(c_1 \dots c_n)$ at location l, linked by a set of arcs $a = \langle s_1, s_2, e_1 \dots e_n \rangle \in A$ connecting two situations (s_1, s_2) through a set of events:

$$e = \langle a, l, c_1 \dots c_n, o_1 \dots o_n, m_1 \dots m_n \rangle \in E$$

where a represents an action performed at location l by characters $(c_1 \dots c_n)$ involving objects $(o_1 \dots o_n)$ and messages $(m_1 \dots m_n)$.

We instantiate robxp stories following Brémond's [2] story model, where there is:

- An initial situation I
- A final situation F
- A set of arcs $a \in A$, each one associated to a main location (or *landmarks* in verbalization's notation [27], see section 4).
- An ordered list of events $e \in E$ in each arc

We propose the following list of actions to narrate: `depart_from(l)`, `arrive_to(l)`, `open(o)`, `close(o)`, `move_to(l)`, `take(o)`, `leave(o)`, `enter(l)`, `exit(l)` where l represents locations, o represents objects and m are textual messages found in the robot's navigation path.

3.1. *The* narrate *dialogue model*

Any reader of almost any kind of tale (from Propp's folktales [24] to Barthes structuralist reading [1] of Balzac) would expect characters, actions and a plot of event turning situations. However, for the robot to generate an *homodiegetic* narration [6], that is, an inside-the-story first person tale of its experience, we need to translate from the robot's internal representation of a task to a robxp story model.

This translation process, where an algorithm scans the robot's knowledge representation of a recently realized task in order to identify the elements a robxp story structure N (see Section 3), is necessarily dependent on the robot's conceptual model and task structure. Actually, the generation process consist on transforming "*the spatial and temporal situation where the robot can be expected to achieve its goals*" [21] into "*the spatio-temporal universe designated by a narrative*" [6], or in other words a *robot experience diegesis*.

The theoretical and empirical framework for our robxp stories generation method is provided by the conceptual model for service robots introduced by Pineda et al. [21]. In this model, robot tasks are represented by a composition of behaviors defining the task structure in two layers: an upper application layer, where the task is specified as a set of situations the

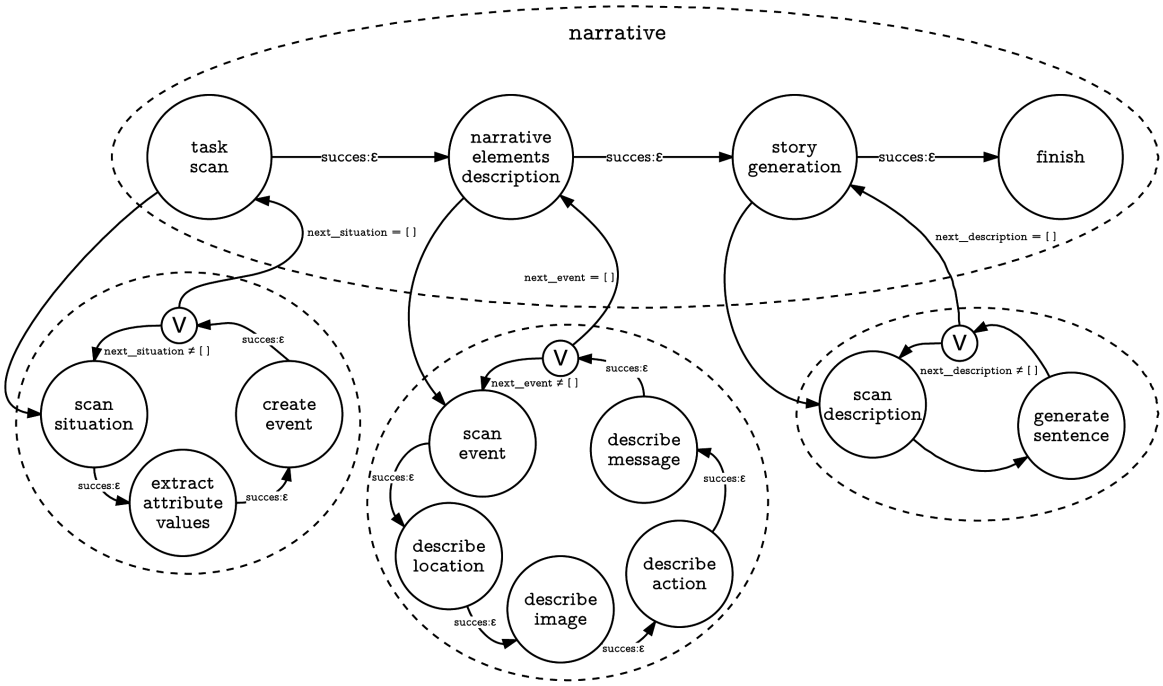

Fig. 1. *Narrate* task structure

robot needs to go through to fulfill a particular application, and a lower behaviors layer, where the robot's native capabilities are represented as a set of basic behaviors. The notion of situation is defined in terms of expectations, actions and expected results. Two kinds of task structures are considered by the conceptual model: static and dynamic, the distinction depending on whether it is possible to specify the task in advance or not. For the sake of simplicity, the scope of the robxp stories generator is limited to static task structures.

The instantiation of Pineda et al. conceptual model for a service robot is possible using the SitLog programming language [22]. SitLog allows the specification of *dialogue models* as a set of situations defined in terms of expectations and actions. When actions are realized and expectations met, a new situation follows. Compositionality is effectively realized by embedding situations in a recursive way.

The motivation of choosing Pineda et al. model for robxp stories generation is twofold: on the one hand, we consider task structures defined in terms of compositional situations and behaviors as a *prediegetic* model of experience, which in a Genette's sense [6] could be defined as the experience preceding or inspiring the narrative. Therefore, the robxp stories generation process implies translating this *prediegetic* experience into a *diegesis*, meaning an narrative where a robot transforms significant elements of its spatial, temporal, performative, visual and cognitive context into an *homodiegetic* first person narration. On the other hand, the Sitlog task structure specification language is particularly well adapted to specify this *experience-into-narrative* transformation task as a dialogue model by itself. As a matter of fact, our goal is that first person situations and behaviors "experienced" by the robot could be "read" by the robot's interlocutor as a set of narrative events composing a robxp story, and furthermore, that a service robot could perform this task if its structure is defined as an algorithmic set of compositional behaviors specified in Sitlog.

Figure 1 shows the *narrate* dialog model in Pineda's et al. diagrammatic representation as a graph of situations [21]. In the first situation, the robot analyzes its own task history H in order to extract meaningful narrative elements: characters, objects, actions and messages organized as events. The second situation specifies a description cycle, where the robot analyzes the list of events in order to produce a descriptions. In the third situation, the robot scans each element of the de-

scription in order to produce sentences composing the robxp story.

4. Robxp stories generation with Sitlog

The SitLog robot programing language models a task as a sequence of situations which are experienced during the execution of the task. The situation represent a stage in the task in which the robot is expecting some events from the environment a in order to execute an specific action and to position itself into a new situation. Formally, a dialogue model is composed by a set of situations S which are defined by a set of triples composed by expectations e, actions a and next situation S_n. When the robot is in a situation s_i and there is a match with what the robot interprets from its environment as one of the expectations, the robot executes those actions associated to the expectation and reaches a new situation s_i. Thus the robot executes a task while unfolding situations and composite behaviors. SitLog composition mechanisms allows that a situation s_{rec}, before being executed, calls another dialogue model, executes it and returns the control s_{rec}.

A robot programmed in the SitLog has also access to the history H of the task, which contains situations, expectations and actions experienced by the robot through the task execution. The main input parameter of the *narrate* dialog model is H. However, the concrete implementation of the *narrate* dialogue model raises two questions: *how could we identify the start and the end of an arc?* and equally important: *how could we determine which events belong to the arc*. It would be easy to associate the initial situation s_i is actually the initial situation from the task's history H, but which of the situation from H should be the starting point of an arc? A possible approximation would be to consider sub-tasks as narrative arcs. However, we suppose that using the spatial information as a discursive axis leads to more natural robxp stories. In particular we refer to situations happening in different locations, for instance in a different room or a different hall. So we choose to split the history H of a task into meaningful locations (or landmarks) the robot passes through, where each split part represents a narrative arc. Such narrative arcs would be then linked by linguistic connectors providing coherence to the sequence of the arcs, like: *then* or *after that*.

Once the narration arcs are defined, we are expected to transform the sequence of situations from the history H into a sequence of events E. However, to translate every situation would render an excessive level of detail which might be received as too repetitive (or too *robotic*). In order to deal with this problem we choose to filter situations based on a list of actions which are interesting from a narrative point of view. This approach leads to a sequence of actions performed by the robot, which might again make it sound too robotic, so we decided to trigger extra behaviors to enrich the narration under certain situations. For instance, after navigating to a point, we launch a `see` behavior which takes a picture and analyses the scene by creating a description of it (i.e., `describe` action) and, in the case of a scene including a textual message, by reading the text (i.e., `read`).

This strategy of task perception enrichment is specially useful when an error occurs and a segment of the task can not been accomplished. This gives the robot the capacity of describing failed situation and explain the context in which the error happened, providing the robot programmer with the possibility of identifying scenarios that were not planned during the task programming phase.

Algorithm 1 Generate events list

INPUTS: H, KB; **OUTPUT:** N

```
L set to main landmarks in the environment
A set to H/L
N set to empty
for each a in A do
    â set to filter_events(A)
    E set to empty
    for each h in â do
        e set to recover_event(h)
        push e to E
    end for
    push E to N
end for
```

4.1. Algorithms

There are three knowledge sources that the robot needs to include in order to be able to generate a robxp story:

Past events knowledge represented by the history H of the past perceptions, interpretations, and actions experienced by the robot as it performed the task [20].

Factual knowledge encoding information about concrete individuals, their properties, relations, states, events and processes. This represents what the robot knows about the world and the action it performs [23]. We will refer to this knowledge as KB

Linguistic knowledge encoding information on how the robot would express itself in natural language, how to name things, how to refer to persons and actions. In particular this type the knowledge is split on the one hand in the KB, through the specification of templates, and on the other hand in the algorithms producing the narrative.

Algorithm 2 Generate experience story

INPUTS: N, KB, **OUTPUT:** S

$S \leftarrow empty$
for all A in N **do**
  $s \leftarrow empty$
  for all e in A **do**
    if e_{action} **then** is dynamic or navigate
      b set to
      $generate\_template(e_a, e_{(ls,os)}, KB)$
    end if
    if e_{action} **then** is visual
      t set to $read(I)$
      b set to $describe(t)$
    end if
    push b to s
  end for
  $\hat{s}$ set to $rewrite(s)$
  push connector to $\hat{s}$
  push s to S
end for

Before generating the robxp story, we need to convert past situation knowledge into a list of events. This is done by Algorithm 1, the input for this process being the history (H) and the factual knowledge (KB). From the KB the landmark locations L and the history H is split into arcs. Next, each arc is run through and filtered, while the main events are collected. All the events e are collected in event lists E and these in the narration N.

Once the narration object has been created, Algorithm 2 is responsible of producing the robxp story. It receives the narration N and the factual knowledge KB as input parameters, and it runs through all the narrative arcs A and all the events e contained in each arc. Each event is rendered into the narrative bit b,

Fig. 2. Golem-II [21]

which, depending on the type of event, renders the content of the encoded KB template. In case of a navigation or a dynamic event, the template is rendered directly; in case of a visual event, the image scene or the textual message must be described using an image description service. Every time an arc is crossed, the arc story is rewritten in order to compact information, introduce some anaphora and create coherence connectors.

5. Experimental implementation

5.1. The robotic platform

We use the Golem II+ robotic platform for robxp, which is the second generation of a service robotic platform. Golem-II+ and its following version, Golem-III, have been extensively programmed to participate in the RoboCup competition [31]. Table 1 lists the main components of Golem-II+ and Figure 2 shows a picture of the robot.

Table 1
Golem-II+ hardware [21]

Mobile Robots Inc. platform	Hokuyo UTM-30LX Laser
On-board computer	One laptop
Microsoft Kinect	Point Grey Flea camera
Audio Fast Track Interface	RODE VideoMic
In house arms	In house neck

5.2. Experimental task

We established an 8 points trajectory at IIMAS lab (see Figure 3). At each point, the robot is expect to perform an action: move an object, read a sign, take a picture and describe the scene, proceed to the next location. The task trajectory includes an impossible action (point 2 in Table 2) with the intention of testing the narration of error states. This experimental trajectory was implemented both *in vivo* (in the Golem II+ robot) and *in vitro*, using a simulation of the robot's system outside the robot by means of a web interface developed for testing purposes.

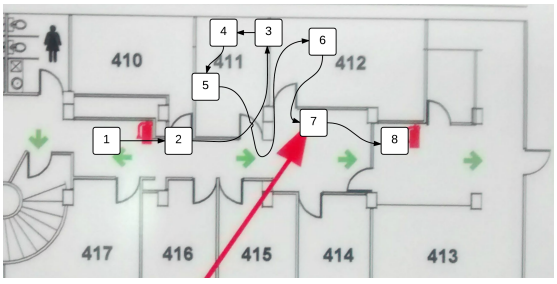

Fig. 3. Experimental task trajectory [16]

At points 1, 3, 5, 6 and 7 the robot should stop, take a picture and describe the scene (see Figure 4). Both task trajectory and the in vitro implementation is explained in detail in Meza et al. [16].

Table 2
Experimental task actions [16]

Point	Instruction
1	Start and read the sign
2	Get into Mr. Bribiesca's office and move the water bottle
3	Get into Mr. Meza's office and read the poster
4	Take the water bottle from the desk under the poster
5	Leave the water on the desk behind and describe what is on it
6	Go to the meeting room and describe the scene
7	If the meeting room is busy, look for a sign on the door and read what is going on
8	Go to the lab and stop

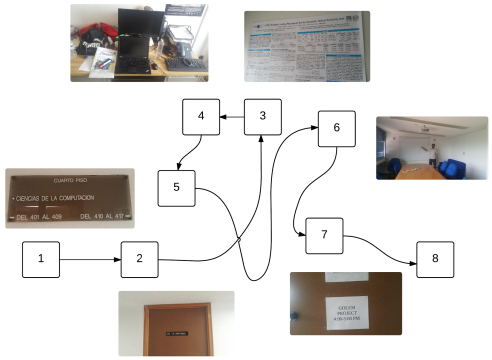

Fig. 4. Photo sequence of points 1, 3, 5, 6 and 7 [16]

5.3. In Vitro implementation

A web application was developed in order to simulate the robxp generation process outside the robot. Its main function is to link the embedded robot's task execution system in SitLog with out-of-the-robot services, like the image description web service. In the future, this architecture would allow us to include a text generator [4,8,5] and a machine reading system [3] to be able to explain text messages read by the robot during the task execution. However, the current implementation scope is limited to a SitLog task specification and an external web service where the pictures of the task trajectory are sent to an image description service in order to return a natural language description. For the in vitro implementation, the trajectory pictures (see Figure 4) are uploaded manually. For the in vivo implementation, they are sent by the robot to the web service during the task execution. The image description service used is Microsoft Cognitive Services computer vision API[2].

The following robxp story was generated by the in vitro implementation at the end of the simulated execution of the experimental task specified in Table 2

> *"I woke up at the hall there was a sign on the door there it said COMPUTER SCIENCE DEPARTMENT. After that I wanted to walk towards the office of Ernesto but couldn't, after that I walked towards the office of Ivan to a desk there it said SOFA: Random Forests Regression for the Semantic Textual Similarity task I took a bottle then I moved to a desk I leaved the bottle I noticed a cluttered desk with a laptop there it said "sausenc &aceta para consejoros - ASO". Afterwards I moved to the meeting room to a table I saw a man standing in a room then I went to a door there it said GOLEM. After that I moved to the laboratory to a*

[2]https://www.microsoft.com/cognitive-services/en-us/computer-vision-api

Fig. 5. In vivo task implementation

desk I noticed a desk with a computer and a chair afterwards I finished."

A human evaluation of this text made by Meza et al. [16] showed that, while the human judges considered this text very complete, it lacks of fluency and readability and correct sentence separation.

5.4. In Vivo implementation in the Golem-II+ robot

Figure 5 shows the Golem-II+ robot executing the experimental task for robxp stories generation. An online video of the execution is available[3]. The in vivo implementation was the first prototype of the algorithm, and the generated robxp story lacks the style improvements in punctuation and connectors reported in Meza et al. [16]. Here's the generated robxp story:

Before the task execution:
Welcome to this demo. I am Golem. I will perform some actions, at the end I will narrate my experience.

After the task execution:
I started at the hall. I begun at the hall. Where it said COMPUTER SCIENCE DEPARTMENT. Afterwards, I tried to the desk. Afterwards, I went to the office of Ivan... to a desk. I grabbed the bottle, then, I walked towards the desk. I noticed the desk with a laptop in a room. I dropped the bottle. After that, I walked towards the meeting room to a table. I saw a person standing in front of a mirror. Then, I went to the door. There it said GOLEM. I noticed a woman standing in front of a mirror. Afterwards I wanted the desk. After that I stopped.

[3]https://www.youtube.com/watch?v=JOu2eMvfrZ0

6. Perspectives

While the robxp stories generated with this approach take into account actions, perceptions and the navigation experience of a robot while performing an experimental task, there is still plenty of room for improvement. In vitro robxp story (see section 5.3) needs improvement in discourse connectors and in text recognition from images (the computer vision API read *"SOFA"* instead of *"SOPA"*. Furthermore, arcs splitting is still confusing, as we can see in the lack of punctuation separating different phrases in *"there it said SOFA: Random Forests Regression for the Semantic Textual Similarity task I took a bottle then I moved to a desk I leaved the bottle I noticed a cluttered desk"*. Main future improvements of our approach will include the integration of the Geni text generator, which relies on a surface realizer based on tree adjoining grammars [4,8,5]. Furthermore, the reading processing of text message found in the robot's visual perception would be extended with explanations found by the machine reading tool FRED [3]. A human evaluation like the one reported in Meza et al. [16], but extending the number of experimental tasks, would be necessary once the full fledged pipeline implementation is done. A promising perspective for this approach would be to apply the in vitro implementation in non robotic tasks, like chatbots or autonomous car navigation.

7. Conclusion

How was your day, robot? The present work aims at providing robots with narrative capabilities, which

would allow to tell in a human friendly way (meaning: a story) a summary of the robot's activity. This work contributes with the Robot Experience Story concept: an experience narration based not exclusively on the spatial experience of the robot, but in a holistic approach of the robot actions, movements and visual perceptions. The *robxp stories* aim is to extract pertinent information from the robot's task history in order to generate a narrative of the recently performed task. Therefore, the second contribution of our work is a *narrate* dialogue model, specified as a graph of situations in the framework of Pineda et al. [21] concept model for a service robot. This dialogue model, which can be instantiated as a standard behavior of the Golem robot for any abstract task, is detailed in two algorithms for the SitLog robot programming language, where the robot's task history is analyzed in order to extract significant elements (objects, characters, locations and messages) to fill a narratology inspired robxp story model, which will be used to produce the sentence of the story and linguistic connectors. Finally, two implementations of this approach are provided: an in vitro one, where the taks is simulated outside the robot, and an in vivo one, where the task is effectively performed by the Golem-II+ robot. The first two robxp stories generated with this approach shows that this narrative still lack of fluency and clarity, so further work will include syntactic and stylistic improvements by means of a text generator and a machine reading tool.

References

[1] R. Barthes. *S/Z*. Seuil, Paris, 1970.

[2] C. Brémond. *Logique du récit*. Seuil, Paris, 1973.

[3] A. Gangemi, V. Presutti, D. Reforgiato Recupero, A. G. Nuzzolese, F. Draicchio, and M. Mongiovì. Semantic Web Machine Reading with FRED. *Semantic Web*, Preprint(Preprint): to appear, 2016.

[4] C. Gardent and E. Kow. GenI, un réalisateur basé sur une grammaire réversible. In *14e conférence pour le Traitement Automatique des Langues Naturelles-TALN 2007*, pages 10–p, 2007.

[5] C. Gardent and L. Perez-Beltrachini. A statistical, grammar-based approach to microplanning. *Computational Linguistics*, 43(1):1–30, 2017. doi: 10.1162/COLI_a_00273. URL http://dx.doi.org/10.1162/COLI_a_00273.

[6] G. Genette. *Figures III*. Seuil, Paris, 1972.

[7] N. Gwo-Dong Chen and C.-Y. Wang. A survey on storytelling with robots. *Edutainment Technologies*, page 450, 2011.

[8] B. Gyawali and C. Gardent. Surface realisation from knowledge-bases. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 424–434, Baltimore, Maryland, June 2014. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/P14-1040.

[9] J. Ham, R. Bokhorst, R. Cuijpers, D. van der Pol, and J.-J. Cabibihan. Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *Social robotics*, pages 71–83. Springer, 2011.

[10] B. Jensen, R. Philippsen, and R. Siegwart. Narrative situation assessment for human-robot interaction. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 1, pages 1503–1508. IEEE, 2003.

[11] J. Kory and C. Breazeal. Storytelling with robots: Learning companions for preschool children's language development. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Washington, DC., 2014. IEEE.

[12] J. J. M. Kory. *Storytelling with robots: Effects of robot language level on children's language learning*. PhD thesis, Massachusetts Institute of Technology, 2014.

[13] A. H. Leversund, A. Krzywinski, and W. Chen. Children's collaborative storytelling on a tangible multitouch tabletop. In *Distributed, Ambient, and Pervasive Interactions*, pages 142–153. Springer, 2014.

[14] M. Marge, C. Bonial, A. Foots, C. Hayes, C. Henry, K. A. Pollard, R. Arstein, C. R. Voss, and D. Traum. Exploring variation of natural human commands to a robot in a collaborative navigation task. In *Proceedings of the The First Workshop on Language Grounding for Robotics*, ACL'2017, pages 58–68. The Association for Computational Linguistics, 2017. ISBN 978-1-945626-64-7.

[15] N. Mavridis. A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35, 2015.

[16] I. Meza, J. G. Flores, A. Gangemi, and L. A. Pineda. Towards narrative generation of spatial experiences in service robots. In *IJCAI 2016 Workshop on Autonomous Mobile Service Robots.*, IJCAI'16, 2016.

[17] I. Meza, C. Rascon, G. Fuentes, and L. A. Pineda. On indexicality, direction of arrival of sound sources and human-robot interaction. *Journal of robotics*, 2016. To appear.

[18] B. Mutlu, J. Forlizzi, and J. Hodgins. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Humanoid robots, 2006 6th IEEE-RAS international conference on*, pages 518–523. IEEE, 2006.

[19] R. Paul, A. Barbu, S. Felshin, B. Katz, and N. Roy. Temporal grounding graphs for language understanding with accrued visual-linguistic context. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, IJCAI'17. AAAI Press, 2017.

[20] L. Pineda, I. Meza, H. Aviles, C. Gershenson, C. Rascon, M. Alvarado, and L. Salinas. Ioca: interactionoriented cognitive architecture. *Research in Computer Science*, 54:273–284, 2011.

[21] L. Pineda, A. Rodríguez, G. Fuentes, C. Rascon, and I. V. Meza. Concept and functional structure of a service robot. *International Journal of Advanced Robotic Systems*, 2015.

[22] L. A. Pineda, L. Salinas, I. Meza, C. Rascón, and G. Fuentes. SitLog: A Programming Language for Service Robots' Tasks. *International Journal of Advanced Robotic Systems*, 2013.

[23] L. A. Pineda, A. Rodríguez, G. Fuentes, C. Rascón, and I. Meza. A light non-monotonic knowledge-base for service robots. *Intelligent Service Robotics*, 10(3):159–171, 2017.

[24] V. Y. Propp. *Morphology of the Folktale*. University of Texas Press, Austin, 2 edition, 1968.

[25] C. Rascon, I. Meza, G. Fuentes, L. Salinas, and L. A. Pineda. Integration of the multi-doa estimation functionality to human-robot interaction. *International Journal of Advanced Robotic Systems*, 12(2):8, 2015.

[26] A. Rohrbach, M. Rohrbach, N. Tandon, and B. Schiele. A dataset for movie description. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[27] S. Rosenthal, S. P. Selvaraj, and M. Veloso. Verbalization: Narration of autonomous robot experience. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, pages 862–868. AAAI Press, 2016. ISBN 978-1-57735-770-4.

[28] Ting-Hao, Huang, F. Ferraro, N. Mostafazadeh, I. Misra, A. Agrawal, J. Devlin, R. Girshick, X. He, P. Kohli, D. Batra, C. L. Zitnick, D. Parikh, L. Vanderwende, M. Galley, and M. Mitchell. Visual Storytelling. In *In Proceedings of NAACL'16*, 2016. to appear.

[29] A. Torabi, P. Chris, L. Hugo, and C. Aaron. Using descriptive video services to create a large data source for video annotation research. *arXiv preprint*, 2015. URL http://arxiv.org/pdf/1503.01070v1.pdf.

[30] M. R. Walter, S. Hemachandra, B. Homberg, S. Tellex, and S. Teller. A framework for learning semantic maps from grounded natural language descriptions. *Int. J. Rob. Res.*, 33(9):1167–1190, Aug. 2014. ISSN 0278-3649. doi: 10.1177/0278364914537359. URL http://dx.doi.org/10.1177/0278364914537359.

[31] T. Wisspeintner, T. van der Zant, L. Iocchi, and S. Schiffer. RoboCup@Home: Scientific Competition and Benchmarking for Domestic Service Robots. *Interaction Studies*, 10(3):392–426, 2009-12-01T00:00:00. doi: 10.1075/is.10.3.06wis.

[32] L. Yao, A. Torabi, K. Cho, N. Ballas, C. Pal, H. Larochelle, and A. Courville. Describing videos by exploiting temporal structure. *ArXiv e-prints*, abs/1502.08029, Feb. 2015. URL http://arxiv.org/abs/1502.08029.