Grounding Language on Roboceptions

Agnese Augello[1] Ignazio Infantino[1] Umberto Maniscalco[1] Giovanni Pilato[1] Filippo Vella[1]

***Abstract*— The aim of this work is to investigate how a robot can learn the association between linguistic elements, such as words or sentences, to the sensations (roboceptions) it feels, coming from its body or the external environment. Through the interaction with human-beings a robot can learn to ascribe a meaning to its roboceptions, and express them in natural language. We discuss about the possibility to define such a process of roboceptions-words association that could be then used by the robot to detect in a verbal interaction some words recalling the previously experimented sensations.**

I. INTRODUCTION

Can a robot be aware of its bodily sensations and learn how to express his feelings? As a child learns from his mother to express and regulate his feelings, according to the cultural/social context in which it is situated, can a robot, by interacting with human beings, learn to conceptualize his sensations and learn the social norms to express and regulate them in society? In [1] a bio-inspired somatosensory system for a humanoid robot has been proposed to simulate the processes under which human beings perceive pleasant and unpleasant physical sensations from the stimulation of their sensory receptors. Employing the somatosensory system, embedded in a cognitive architecture, a robot can roughly classify its sensations, that we name here *roboceptions*. However, at present, the robot is not able to associate its roboceptions to a linguistic representation.

Starting from this first arrangement of the stimuli in roboceptions, we want to discuss the definition of a conceptualization process mingled with the grounding of roboceptions. The problem has been tackled by many authors [9][10]; we consider, analogously to what stated in [3] [4] [7], that language, and communication, in a socio-cultural context, play a central role in the process through which a robot can "understand" what happens in its body and the sensations it is feeling. In our vision, a robot can learn to ascribe a meaning to its roboceptions and express them through the interaction with human-beings.

In this position paper, we propose a *neural grounding process* that allows the robot to associate linguistic items, such as words or sentences, listened during an interaction with human beings, to the sensations (roboceptions) it feels at that moment.

We discuss the possibility of inducing such a process of roboceptions-words association that could be subsequently used by the robot to detect, during a verbal interaction some words recalling the previously experimented sensations, improving its capability of being empathetic, being able to "understand" the sensations that the human is describing.

II. THE ROBOT'S SOMATOSENSORY SYSTEM

The robotic somatosensory system consists of a layer of different soft sensors [2][6], defined on top of the robot's physical sensors. The somatosensory system has been defined to allow a humanoid to perceive and classify its physical sensations, that we called *roboceptions*.

We used soft sensors to coarsely emulate the human somatosensory capability. Starting form the built-in basic hardware sensors of a robot and by defining and implementing a set of soft sensors, we have designed a bio-inspired somatosensory system framework. For each kind of sense or stimulus, we tried as much as possible to emulate a biological model. We implemented the system on a Softbank NAO Humanoid Robot [1], however, the whole framework is completely general [1]. From the embedded sensors we can get basic information, e.g.: the current values and the temperatures of twenty-five actuators, the pressure of four switches located at the tip of each foot, nine (ON/OFF) touch information located on the head (*front, rear, middle*) and on the hands (*back, left, right*), three distances achieved by sonars located at the chest of the robot and some information about the energy system, etc. [1] [2]

We consider the transduction function and the data fusion of the basic sensorial data as fully-fledged soft sensors since they estimate quantities that can not be measured by physical instruments. They can estimate quantities like "muscular pain" caused by high current, or also the exertion caused by a high level of temperature in its actuators that affects the ability of a muscle to produce an effective force. They can also estimate pleasant sensations such as that generated by a caress.

The whole somatosensory state is coded as a vector which represents the set of sensations felt by the robot through the set of soft sensors.

III. GROUNDING LANGUAGE ON ROBOCEPTIONS THROUGH DUAL LEARNING FOR MACHINE TRANSLATION

The internal, primitive, representation of the robot roboceptions, coded as a vector, can be expressed in some sort of an elementary language. This language, which we name *"RoboLang"*, is homologous to that one used by babies to manifest their basic human needs.

*This work was not supported by any organization

[1] Institute for high performance computing and networking (ICAR), National Research Council (CNR), Via Ugo La Malfa 153, 90146 Palermo, Italy, name.surname@icar.cnr.it

[1]https://www.ald.softbankrobotics.com

Robolang is composed of a set of elementary statements that have this form *(RoboceptionType, Location, Value)* where:

- *RoboceptionType* $\in$ {*CurrentPain, TemperaturePain, CaressFeeling, BumperPain, Anxiety, Energy, Noise*}
- *Location* $\in$ {*0, 1, 2, $\cdots$, N*}, each number identifying one of the N physical position of the robot soft sensor. The value 0 indicates that the Roboception regards the whole robot and it has not a specific localization.
- *Value* $\in$ {*VeryLow,Low, Normal, High, VeryHigh*}, corresponding to the linguistic quantization of the robot soft sensors output value. If the sensor gives a binary output, the set of possible values are *VeryHigh* and *VeryLow*, associated to the values "True" and "False" respectively.

The statement is expressed only if a specific condition of interest is satisfied. Each specific condition, at present, is hardwired in the somatosensory system (i.e. a set of thresholds have been set). In the futue these conditions will be learned by the robot in an autonomous manner.

The statements can be concatenated, indicating the co-occurrence, during a given time interval, of multiple values of different relevant roboceptions at a given time.

As an example, these simple "sentences" are possible:

- *(CurrentPain 2 High), (Energy 1 VeryLow)*
- *(Energy 1 High), (CaressFeeling 3 VeryHigh), (TemperaturePain 3 High)*

These *RoboLang* sentences represent the basic expression of the robot state or its primitive needs.
The goal of the idea illustrated in this paper is to link these simple expressions in *Robolang* with more sophisticate sentences expressed in natural language. The consequence of this approach is twofold: to let the robot learn how to better express naturally its feelings, and, vice-versa, to let the robot being capable to "understand" and "internalize" the sensations that a human being communicates through the use of natural language. Since, basically, we have two languages to bridge, the grounding of roboceptions can be treated as a machine translation problem. One of the most popular approaches to perform automatic translation between languages is the Neural Machine Translation. However, one of the big drawbacks of this kind of techniques is that they require a huge amount of bilingual sentences to be trained, while, in our case, the interaction between the human and the robot is very limited. This leads to a lack of training data. In [5] it has been presented a dual-learning mechanism, named dual-NMT, which can be used on a Neural Machine Translation (NMT) system to automatically learn from unlabeled data by exploiting a dual-learning game. The advantage of this system is that it uses only the 10% of bilingual data for warm start. The dual-NMT approach learns mainly from monolingual data, and it is capable to reach an accuracy which is comparable to traditional NMT that have been trained from full bilingual data. The approach illustrated in [5] is particularly suitable for our problem since the dual learning technique is generally applicable to any kind of problem where tasks are in dual form: it is possible to learn the two tasks at the same time by using unlabeled data by exploiting reinforcement learning techniques. According to this approach, we only need two well trained, or defined, language models $LM_A(.)$ and $LM_B(.)$ which, given a sentence as input, outputs a real value which estimates how the sentence is a natural sentence in its own language.
Furthermore, we need a limited number of sentence pairs used to make a warm start of the learning phase of the system, e.g., *("(CurrentPain 3 VeryHigh), (TemperaturePain 3 High),"You have a headache")* or *("(CurrentPain 2 VeryHigh), (BumperPain 2 VeryHigh),"You slammed your right foot")*. These pairs can be obtained, at present, by interacting with the robot though a terminal: the robot writes on a screen in *"Robolang"*, and the human being tries to describe the sensation in natural language.

By using, therefore, two language models and an always growing set of such pairs, by exploiting this technique, we are confident that the robot can eventually learn how to describe its own sensations directly from the interaction with a human being, making the robot more convincing and more "human". Furthermore, when the robot hears someone saying "I have a headache", it can "feel" the sensation, "computationally understanding" how does it feel to have a headache, making it more empathetic, mimicking somehow the mirror-neuron theory [8].

IV. Conclusions

In this position paper, we have presented the general idea of a system that could be capable to roughly emulate the learning process of babies when they try to express their sensations and needs in natural language. We are implementing the system, and we will make the first experiments to assess the effectiveness of the proposed approach.

References

[1] Augello, A., Infantino, I., Maniscalco, U., Pilato, G., Vella, F. (2018) "Robot Inner Perception Capability Through a Soft Somatosensory System" International Journal of Semantic Computing, 12(01), 59-87.

[2] Augello, A., Infantino, I., Maniscalco, U., Pilato, G., Vella, F. (2017). The effects of soft somatosensory system on the execution of robotic tasks. In Robotic Computing (IRC), IEEE International Conference on (pp. 14-21). IEEE.

[3] Burr, Vivien. (2006) "An introduction to social constructionism" Routledge

[4] Boiger, Michael, and Batja Mesquita. (2012) "The construction of emotion in interactions, relationships, and cultures" Emotion Review 4.3 pp.221-229

[5] He, D., Xia, Y., Qin, T., Wang, L., Yu, N., Liu, T., Ma, W. Y. (2016). Dual learning for machine translation. In Advances in Neural Information Processing Systems, pp. 820-828.

[6] Maniscalco, U., Pilato G., Vassallo G. (2010) "Soft Sensor based on E-αNETs." Frontiers in Artificial Intelligence and Applications, Volume 226: Neural Nets WIRN10, pp. 172-179

[7] Mesulam, M-Marsel. (1998) "From sensation to cognition." Brain: a journal of neurology 121, no. 6 pp.1013-1052.

[8] Rizzolatti, Giacomo; Craighero, Laila (2004). "The mirror-neuron system". Annual Review of Neuroscience. Vol.27 No.1 pp.169192.

[9] Tani, Jun. (2016) Exploring robotic minds: actions, symbols, and consciousness as self-organizing dynamic phenomena. Oxford University Press.

[10] Taniguchi, T., Nagai, T., Nakamura, T., Iwahashi, N., Ogata, T., Asoh, H. (2016). Symbol emergence in robotics: a survey. Advanced Robotics, 30(11-12), 706-728.