

# Leolani: A Robot That Communicates and Learns about the Shared World

Piek Vossen<sup>1</sup>, Selene Baez<sup>1</sup>, Lenka Bajčetić<sup>1</sup>, Suzana Bašić<sup>1</sup>, and Bram Kraaijeveld<sup>1</sup>

VU University Amsterdam, Computational Lexicology and Terminology Lab, De Boelelaan 1105, 1081HV Amsterdam, The Netherlands  
{p.t.j.m.vossen,s.baezsantamaria,l.bajcetic,s.basic,b.kraaijeveld}@vu.nl

**Abstract.** People and robots make mistakes and should therefore recognize and communicate about their “imperfectness” when they collaborate. In previous work [3, 2], we described a female robot model *Leolani(L)* that supports open-domain learning through natural language communication, having a drive to learn new information and build social relationships. The absorbed knowledge consists of everything people tell her and the situations and objects she perceives. For this demo, we focus on the symbolic representation of the resulting knowledge. We describe how *L* can query and reason over her knowledge and experiences as well as access the Semantic Web. As such, we envision *L* to become a semantic agent which people could naturally interact with.<sup>1</sup>

**Keywords:** robot, knowledge representation, ontology learning, communication

## 1 Introduction

In order to handle errors and possible conflicts, humans and robots need to be able to communicate about the surrounding world and each other. This requires the formal modeling of the perceived world within a specific context (i.e. the self, the surrounding objects, locations and people), as well as the accumulating knowledge that is the result of the perception (i.e. instances of concepts, properties of instances, relations between instances), and finally the provenance of the acquired knowledge (who stated what and when).

We developed a robot model that formally captures the complete process: from raw signals representing world experiences (both perceptions and natural language), to their interpretation resulting in symbolic knowledge that functions as the ‘brain’. The communication is used to learn about the world and to correct errors. Furthermore, the robot can reflect on the accumulated knowledge which results in states-of-the-brain that are formally modeled as *thoughts*. These thoughts trigger actions by the robot to acquire more knowledge, to resolve conflicts and uncertainties, and to build up trust on the sources.

---

<sup>1</sup> Copyright 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In this demo, we focus on the modeling of the different data layers in the process and the different initiatives or drives that are triggered by the thoughts. In Section 2, we describe a typical interaction with  $L$ . Section 3 contains the overview of the models and Section 4 describes the learning process as a result of the drives. Section 5 concludes and explores future work.

## 2 Interacting with the world

$L$  is a curious robot equipped with cognitive abilities and communication skills to support social behaviour. Initially,  $L$  scans the objects and people in her environment and relates them to a new instantiated context. Next, she tries to determine her location either by reasoning over previous contexts (i.e. the overlap with previous visits to a location) or asking an available source. Upon encountering people,  $L$  tries to discern whether the person is already known and should be greeted as such, or if she is meeting the person for the first time, in which case a get-to-know sequence is initialised.

$L$  then waits for the person to initiate conversation by asking a question or saying a statement. Questions trigger simple queries, while statements are processed to represent new knowledge along its provenance. As new information is added, this generates *thoughts*, which reflect on the current state of the brain and how it is affected by the input. Through these thoughts the robot raises questions or comments to the person, thus encouraging conversation.

## 3 Model

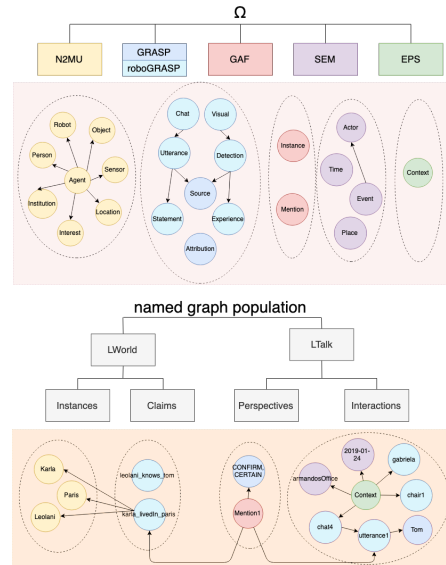
Firstly,  $L$  must represent knowledge about the **world**. “Nice to meet You” (N2MU) is a social ontology covering basic concepts for human-robot social interaction (e.g. a person’s name<sup>2</sup>, place of origin<sup>3</sup>, occupation, interest).

In order to model a theory of mind, we use GRaSP [1] to represent the notion of **mentions and perspectives**. *grasp:denotedIn* links an abstract *gaf:Instance* to their *gaf:Mention* in a specific signal (e.g. camera, human speech). Each of these mentions are related to a *grasp:Source*<sup>4</sup>, and a *grasp:Attribution* (i.e. denial/confirmation, sentiment/emotion, and certainty).

<sup>2</sup> Where possible, we follow the FOAF model: <http://foaf-project.org>

<sup>3</sup> Where possible, we follow the geonames model: <https://www.geonames.org/>

<sup>4</sup> Where possible, we follow the PROV-O model: <https://www.w3.org/TR/prov-o/>



**Fig. 1.** Linked ontologies and graph population: Tom says “Karla lives in Paris”.

The notion of **context** allows  $L$  to identify and be aware of different situations and the objects within. Thus we introduce the ontology for Episodic Awareness (EPS) that defines a *eps:Context* which *eps:hasDetection* (objects and people), and *eps:hasEvent* (chats). Additionally, we rely on the Simple Event Model (SEM)<sup>5</sup> for representing event properties like *sem:{Actor,Place,Time}*.

## 4 Learning and Drives

After knowledge has been acquired, **thoughts** on conflicts, trust, gaps, novelty, and overlaps are derived through reasoning over the graph. Conflicts link back to theory of mind and represent information that one or more actors have claimed but which logically cannot coexist because of a) a previous statement directly negates this one (Following example in Figure 1: "Karla told me she does not live in Paris") or b) only one object is allowed for this predicate ("I heard Karla lives in Amsterdam, not in Paris").  $L$  builds on this information to calculate the *trust* she assigns to the source of the information, based on how many times she has interacted with the source, how much she has learned from them, and how many conflicts their claims produce ("I trust you, Tom. You teach me new things"). In parallel,  $L$  focuses on identifying sparse areas in her knowledge graph, as these represent learning opportunities around people as subjects of knowledge ("Where is Karla from?"), or objects as elements that either people talked about or  $L$  perceived ("Where is Paris located?"). Finally,  $L$  is able to show awareness when an actor claims knowledge that was acquired before ("Gabriela told me last week that Karla lives in Paris"), convey excitement if it represents *novel* information ("I did not know where Karla lives!"), or guide the conversation by emphasizing on *overlapping* information across entities, for example, to discover groups of people with similar properties ("My friend Armando also lives in Paris"). These help in making the conversation smooth and engaging the actor with  $L$ .

**Open-domain learning** from conversation entails that extracted entities may be stored as instances of unknown class (*owl:Thing*), labeled only by the output text from speech recognition without further interpretation. However, typing entities empowers  $L$  to generate more interesting and meaningful thoughts. Therefore, a specific typing pipeline first employs Wordnet, followed by an attempt that exploits  $L$ 's access and interoperability with linked open data (LOD) by querying over DBpedia and WikiData, using predefined heuristics for the accepted types. Once the entity's class has been established, the learned information is used to represent properties at an instance level, and to expand and enrich N2MU by creating new classes and object properties, as well as *owl:sameAs* mappings to the original LOD resources. This is a promising avenue to explore ontology learning through communication and human-robot interaction, using  $L$  as a method for data collection and possibly LOD enrichment.

**Episodic memory** is modeled to create **spatial awareness** and identify instances of objects within. At first, the location is determined by judging the

<sup>5</sup> <https://semanticweb.cs.vu.nl/2009/11/sem/>

overlap of the physical objects observed in the current context with those from all previously modeled contexts. If there is sufficient overlap with a previous context,  $L$  hypothesizes that she is now in the same location. Otherwise, she assumes a new location and will temporarily model this as 'Unknown'. Once given the chance,  $L$  will ask for the unknown location name and perform on-the-fly ontology modification to model this new location. Location information is in turn used alongside object properties for **disambiguating object instances** within and across contexts.

## 5 Conclusion and Future Work

$L$  is a semantic agent that absorbs knowledge and reasons over everything she has heard or experienced. As such, it is easily imaginable that malicious users may feed  $L$  with false information, thus misguiding its learning. However, one of the main strengths of this model is that, after sufficient interactions and using prior knowledge,  $L$  would be able to identify suspicious information, judge its veracity, and gather evidence for its reasoning.

Ongoing work on this project focuses on representing temporal information, specifically with regards to object permanence and event duration. Furthermore, the current NLP pipeline consists of rule-based syntactic and constituency parsing components developed specifically for parsing English spoken natural language and extracting RDF. Thus, one utterance is exploded into several triples, depending on the complexity of the information and limited by the coverage of the rule-based system. Experimenting with different SOTA systems for Named Entity Recognition and Relation Extraction may bring improved results and add flexibility to the NLP pipeline, previous to representing knowledge.

Our code is available on Github<sup>6</sup> and the project progress is reported on our website<sup>7</sup>. Links to videos of the demo set up can be found as well<sup>8</sup>

## References

1. Fokkens, A., Vossen, P., Rospocher, M., Hoekstra, R., van Hage, W.: Grasp: Grounded representation and source perspective. In: Proceedings of KnowRSH, RANLP-2017 workshop, Varna, Bulgaria (2017)
2. Vossen, P., Baez, S., Bajčetić, L., Basvić, S., Kraaijeveld, B.: A communicative robot to learn about us and the world. In: Proceedings of Dialogue-2019, Moscow (2019)
3. Vossen, P., Baez, S., Bajčetić, L., Kraaijeveld, B.: Leolani: a reference machine with a theory of mind for social communication. In: Proceedings of TSD-2018, Brno, <https://www.tsdconference.org/tsd2018> (2018)

<sup>6</sup> <https://github.com/cltl/pepper>

<sup>7</sup> <http://makerobotstalk.nl/>

<sup>8</sup> <https://drive.google.com/drive/folders/1RZcIM8JeIFxYw1tly5dDglQYPZH2WJkZ?usp=sharing>