What Say You: An Ontological Representation of Imperative Meaning for Human-Robot Interaction

Robert PORZEL [a], and Vanja CANGALOVIC [b,1]
[a] *Digital Media Lab, Bremen University, Bremen, Germany*
[b] *Applied Linguistics, Bremen University, Bremen, Germany*

Abstract. There is a distinct difference between modeling linguistic knowledge and knowledge that is expressed linguistically. The research effort described herein concerns the latter, i.e. how to model the output of a natural language understanding system in a formal ontology in such a way that robotic agents can carry out the tasks given to them via natural language. For this, we build on a given foundational ontology that is suitable for our requirements and introduce the basic modeling principles and design patterns to model and represent the meaning of instructions. Linguistically, our model is informed by constructional approaches to language understanding, i.e. embodied- and fluid construction grammar. Ontologically, we base our model on the foundational framework provided by the Dolce Ultra Light + Descriptions and Situations ontology. The proposed model, called SOMA-SAY, is part of a larger system of ontologies that are employed in the Everyday Activity Science and Engineering collaborative research center.

Keywords. Natural Language Understanding, Everyday Activities, Robotics

1. Introduction

Numerous formal ontological models representing linguistic objects and their corresponding grammatical theories have been presented in the past. This is not another one of these models. While we will discuss some prior modeling efforts in Section 2, the scope and intent of our model differs from previous ones in several respects. Firstly, we do not focus on modeling linguistic objects themselves, such as morphological, lexical or syntactic constructions. We do, however, focus on representing the socially constructed meaning that is conveyed by instructions. Specifically, our work takes place in the challenging and practical application of giving robots underspecified and vague natural language instructions to perform certain everyday activities that are subsequently carried out by them. Representing an underspecified instruction also involves representing information that was left implicit by the textual or verbal command.

The ontology module presented in this work is part of the cognitive backbone of a greater robotic infrastructure [1]. It serves the function of representing the socially constructed semantic content of an instruction. More specifically, given that we are using a constructional approach to language understanding [2], we need to represent the pragmatic implications of the meaning pole of imperative constructions. This work, therefore, constitutes an attempt to employ state of the art knowledge engineering principles to connect formal ontologies to constructional cognitive linguistic theories [3,4].

2. State of the Art

In the past 20 years various approaches to model linguistic knowledge, i.e. the entities and features that make up human language, in formal ontologies have been proposed. These approaches differ in some respects such as alignment to upper layers, their modeling intent and their scope. One point of divergence lies in the specific alignment to a foundational layer. While, for example, the GOLD ontology [5] is aligned to the SUMO upper ontology [6], the OntoWordNet model [7], is aligned to the DOLCE foundational ontology [8]. The LingInfo model [9] can be used with any foundational framework as it relies on meta-classes to model information about the lexical entities. In contrast, the OntoWordNet aims at merging the linguistic information contained in WordNet with the respective classes employed in specific domain models, while both LingInfo and GOLD seek to incorporate more linguistic information, such as morphological and grammatical features of language. They all allow for a direct connection of the respective linguistic information for terms with corresponding classes and properties in a domain ontology.

These efforts are, in a sense, orthogonal to ours and each model could be integrated as an additional module to allow reasoning about linguistic information or as a link between lexical and ontological resources. For those purposes we employ a *Lower Semantic Model* that connects lexical and ontological information and is interchangeable with either of the models described above. More closely related to our approach is the so-called *General Upper Model* (GUM) [10]. GUM provides a detailed semantics for linguistic spatial expressions and is based on a principled ontological engineering approach. It covers language concerned with space, actions in space and spatial relationships for which an ontological organization is proposed that relates such expressions to general classes of fixed semantic import. However, as we seek to align our model with a specific foundational model and construct it as a module within the SOMA[2] ontological framework, we cannot employ the upper model GUM as is, but will re-use relevant details concerning schematic theories about functional relations, where applicable.

The ontology module described in this work is part of the EASE family of ontologies, collectively called SOMA. Like all of the SOMA ontologies, it is based on the DOLCE+DnS Ultralite (DUL) foundational framework [8], a decision that is greatly motivated by their underlying ontological commitments. Firstly, DUL is not a revisionary model, but seeks to express stands that shape human cognition. Furthermore, it assumes a reductionist approach – rather than capturing, for example, the flexibility of our usage of objects via multiple inheritance in a multiplicative manner, it commits to a reduced *ground* classification and use a *descriptive* approach for handling this flexibility. For this,

[2] SOMA stands for *Socio-physical Model of Activities (https://ease-crc.github.io/soma/).*

a primary branch of the ontology represents the ground physical model, e.g. objects and actions, while a secondary branch represents the social model, e.g. roles and tasks. All entities in the social branch would not exist without humans, i.e. they constitute social objects that represent concepts about or descriptions of ground elements. Every axiomatization in the physical branch can, therefore, be regarded as expressing some physical context whereas axiomatizations in the descriptive social branch are used to express social contexts. A set of dedicated relations is provided that connect both branches. For example the relation *classifies* connects ground objects, e.g. specific utterances, with the roles they can play, i.e. potential classifications. Thus, we can state that an interrogative can in some context be conceptualized as an assertion, a command or simply a query. Nevertheless, neither its ground ontological classification as an interrogative will change nor will interrogatives be subsumed as commands or assertions via multiple inheritance.

The semantics of the actions and entities executed by the robots performing everyday activities, such as setting a table or cooking a meal, are defined by this formal ontology. This ontology is designed to provide descriptions for everyday activities in terms of human physiology and human mental concepts, as well as enabling formal reasoning. The ontology supplying the labels for the objects has been designed using the principles proposed by Masolo et al. [8,11]. Specific branches of the former KnowRob knowledge model pertaining to everyday activities [1], such as those involved in table setting and cooking, have consequently been aligned to the DUL framework in the creation of the SOMA framework. Additional axiomatization that is beyond the scope of description logics is integrated by means of the Distributed Ontology Language [12].

3. The SOMA-SAY Module

Our application domain of everyday activites requires robotic agents to carry out natural language instructions, such as recipes or manuals. This involves flexibly parsing abstract and underspecified input into a meaning representation, which in turn enables further steps necessary for real grounding, such as simulations or additional reasoning. To this end, a deep semantic parser based on the Construction Grammar formalism is employed [13,14,15]. Both the constructions' meaning poles and the analysis itself make use of ontological knowledge, to disambiguate otherwise unclear instructions and to evoke unspecified parameters which need to be inferred by later processing steps. In this way, natural language commands are transformed into models of SOMA-SAY, which consist of series of state transitions specifying the evoked schemas. The employed parser's underlying mechanism is based on the unification and merging algorithms implemented in Fluid Construction Grammar[2], while its ontology integration and schema handling are inspired by the Embodied Construction Grammar framework [16]. The constructions are co-engineered with the concepts of SOMA-SAY and can be viewed as rules mapping between word-positional relationships and syntactic properties, and semantically rich ontological concepts. The parsing process integrates tightly into our knowledge base-focused understanding pipeline, with both the internal data structures as well as the semantic output being represented as semantic triples.

We can firstly define a LOCUTION as an ACTION[3] A LOCUTION represents any activity of expression using natural language that can be verbal, textual or else. The ac-

[3] An EVENT in DUL can be an ACTION, a PROCESS or a STATE.

tual linguistic form of such an expression, i.e. declaratives, imperatives or interrogatives, are modeled as a LINGUISTICOBJECT. A LINGUISTICOBJECT itself is an INFORMATIONOBJECT. Using the given *hasPart* relation we can express that only linguistic objects can be components of a locution, i.e. `Locution` $\sqsubseteq \forall$ *hasPart*.`LinguisticObject`.

DUL features the useful distinction between an EVENT that happens in the world and an EVENTTYPE, i.e. a classification or interpretation thereof.[4] Analogously, we can apply the same design pattern to the linguistic function of a LOCUTION, i.e. an ILLOCUTION. Additionally, we define three types of ILLOCUTION, i.e. ASSERTION, INSTRUCTIVE, and INTERROGATIVE. This solves certain aspects of semantic construal and pragmatic analysis [17]. We all know that utterances formalized as declaratives, e.g. *it is drafty in here*, are actually veiled instructions as in this example to close a window or door. Our models allows for all possible construals, i.e. questions that are meant as assertions, e.g. *how stupid is this* meaning *this is very stupid* as all other form-function combinations. This is realized via the *classifies* relation.

In our domain, instructions are important as they constitute the input for the robotic agents. So even if asked *could you put the plates on the table* the robot should not reply with a resounding *yes*, but rather execute the given command. In the DnS system [18] a DESCRIPTION is a SOCIALOBJECT that represents a conceptualization that can be understood as a 'descriptive context'. This context uses or defines concepts to create a view on a 'relational context'. This relational context is modeled as a SITUATION. A SITUATION is a view consistent with a DESCRIPTION, i.e. it *satisfies* it. It is created by an observer on the basis of a 'frame' provided by the DESCRIPTION. As such it used to represent reified n-ary relations, where *isSettingFor* is the top-level relation for all binary projections of the n-ary relation. For modeling the meaning of any locution that is construed as a command we employ the TRANSITION subclass of SITUATION. First of all a TRANSITION creates a context for two additional different Situation(s), i.e. the state before and the state after the transition takes place. For our scenario we create a subclass of TRANSITION called STATETRANSITION. Via the relation *includesEvent*, we can state that every STATETRANSITION includes exactly two EVENTS, e.g. two STATES.

Additionally, we create a new type of situation, called a SCENE. Using the given DUL relations of *hasInitialState* and *hasTerminalState*, we can connect situations to situations, i.e. two scenes to one transition. Based on the constructional approach to language understanding, representing the meaning of utterances is impossible without recourse to schematic knowledge in the form of image schema, x-schema or a FrameNet frame [3,19,20,4,21]. We subsume these schematic theories under the class THEORY making them a type of DESCRIPTION. We can now start to populate our model with the schemas that have been proposed in the literature [3,19] with the addition of an ESTSCHEMA that is evoked in expressions featuring the existence of an entity, as in "there is a cup".

We now define a new *evokes* relation that can hold between two social objects. In our case we employ it to state that schemas can evoke other schemas, thereby making the evoked schemas' constituent roles accessible to itself [16], e.g. `CausedMotionSchema` $\sqsubseteq \exists$ *evokes*.`SourcePathGoal`. Following the computational practice [22,16], we want to assign constituents to the individual schemas, whereby only entities of type ROLE can be a constituent of a schema, i.e. `SchematicTheory` $\sqsubseteq \forall$ *hasConstituent*.`Role`

[4] DUL defines an EVENTTYPE as a concept that *classifies* an EVENT. It should, therefore, describe how an EVENT should be interpreted, executed, expected, seen, etc.

As provided by the DUL foundations, the classification of objects is realized via the Role pattern. Roles are Concepts and, as such, reside in the SocialObject branch of DUL. The *classifies* relation is used to constrain what can be classified as a source or goal or what can be classified as a trajector. This role pattern is if paramount importance, especially in restricting the type of entity that become the filler of a schematic constituent. The SOMA framework imports the roles that have been established in the field of frame semantics [21]. The selectional restrictions imposed by the *classifies* relation are used in a number of reasoning processes ranging from natural language understanding to tool selection. As certain roles can only classify physical agents or specific types of designed artifacts, these axiomatizations provide substantial information about context dependent *meaning* of objects, e.g. `Destination` $\sqsubseteq \forall$ *classifies*.`Location`. Lastly, we can now link situations to schematic theories, as every scene and action in a given situation evokes specific schematic theories that consequently need to be satisfied as well, i.e. `StateTransition` $\sqsubseteq \forall$ *satisfies*.`SchematicTheory`.

Please note that the ensuing structures are still not sufficiently specific to drive a robotic control system. A detailed overview of further parametrizations and explication of these plans via physical simulation and human computation is given in Pfau et al. [23].

4. Conclusion

We introduce a modeling approach *cum* model that seeks to represent the social interpretation of imperatives, regardless of their linguistic form. The ensuing model is part of the SOMA framework that constitutes a socio-physical model of activities for autonomous robotic agents. As SOMA itself is aligned to the DUL foundational ontology, we showcase where and how our model connects to this axiomatic theory about the high-level domain-independent categories in the real world. There are obvious limitations to our model, as numerous constraints are outside of the scope of description logics. More expressive models of image schemas and their implications, in first order logic, have been proposed and discussed recently [24]. In other on-going work, we define theories for image schemas in a more tractable formalism of propositional defeasible logic and use them for counterfactual simulation. As future work we are considering a complete reformulation of constructional and schematic knowledge using some form of propositional defeasible logic. Additionally, further reasoning and *mental simulation* is needed to explicate the given instructions completely. This concerns both the grounding of referents as well as the parametrization of the ensuing actions.

As with the work presented herein, we aim to put the insights provided by cognitive linguistic theories to use in making robotic agents more flexible and robust in carrying out the vague and underspecified instructions that are given to them by their human interlocutors. This increase in flexibility is a result of combining physical models that represent actual events, such as a Locution with social descriptions thereof, e.g. an Illocution. This two-pronged approach is employed throughout the proposed SOMA-SAY module, whether it is in the Role - Object or the Task - Action pattern. Employing this throughout the cognitive infrastructure of robotic agents enables them to reason about their actions and react felicitously to a given linguistic input. This work, therefore, represents another step towards moving from robotic agents that can perform a single task to ones that actually master an everyday activity, such as preparing a meal and setting a table.

References

[1] Beetz M, Bessler D, Haidu A, Pomarlan M, Bozcuoglu AK, Bartels G. Know Rob 2.0 - A 2nd Generation Knowledge Processing Framework for Cognition-Enabled Robotic Agents. Proceedings - IEEE International Conference on Robotics and Automation. 2018:512–519.

[2] Steels L. Basics of Fluid Construction Grammar [Journal Article]. Constructions and Frames. 2017;9(2):178–225.

[3] Lakoff G. Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. University of Chicago Press; 1987.

[4] Feldman J, Lakoff G, Bailey D, Narayanan S, Regier T, Stolcke A. L_0—The First Five Years of an Automated Language Acquisition Project. AI Review. 1996;10:103–129.

[5] Farrar S, Langendoen T. A Linguistic Ontology for the Semantic Web. GLOT. 2004 06;7:97–100.

[6] Niles I, Pease A. Towards a Standard Upper Ontology. In: Proceedings of the International Conference on Formal Ontology in Information Systems - Volume 2001. FOIS '01. New York, NY, USA: Association for Computing Machinery; 2001. p. 2–9.

[7] Gangemi A, Navigli R, Velardi P. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. In: Meersman R, Tari Z, Schmidt DC, editors. On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. Berlin, Heidelberg: Springer Berlin Heidelberg; 2003. p. 820–838.

[8] Masolo C, Borgo S, Gangemi A, Guarino N, Oltramari A. Wonderweb deliverable d18, ontology library (final). ICT project. 2003;33052:31.

[9] Cimiano P, Buitelaar P, Frank A, Racioppa S, Sintek M, Kiesel M, et al. LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies. In: Proceedings of the OntoLex Workshop at LREC. ELRA; 2006. p. 28–32.

[10] Bateman JA, Hois J, Ross R, Tenbrink T. A Linguistic Ontology of Space for Natural Language Processing. Artif Intell. 2010 Sep;174(14):1027–1071.

[11] Mascardi V, Cordì V, Rosso P. A comparison of upper ontologies (disi-tr-06-21). Dipartimento di Informatica e Scienze dell'Informazione, Universitŕ degli Studi di Genova. 2008;35:16146.

[12] Mossakowski T. The Distributed Ontology, Model and Specification Language–DOL. In: International Workshop on Algebraic Development Techniques. Springer; 2016. p. 5–10.

[13] Fillmore C. The mechanisms of construction grammar. In: Berkeley Linguistics Society. vol. 14; 1988. p. 35–55.

[14] Goldberg AE. Constructions: A Construction Grammar Approach to Argument Structure. University of Chicago Press; 1995.

[15] Fillmore C, Kay P. Construction grammar. Stanford, CA: CSLI; 1999.

[16] Chang N, Feldman J, Porzel R, Sanders K. Scaling Cognitive Linguistics: Formalisms for Language Understanding. In: Proceedings of the 1st Workshop On Scalable Natural Language Understanding; 2002. .

[17] Katz J. Propositional Structure and Illocutionary Force. Mass.: Harvard University Press; 1980.

[18] Gangemi A, Mika P. Understanding the Semantic Web through Descriptions and Situations. In: Proceedings of the ODBASE Conference. Springer; 2003. .

[19] Johnson M. The Body in the Mind Metaphors. University of Chicago Press; 1987.

[20] Langacker RW. Foundations of Cognitive Grammar, Vol. 1. Stanford University Press; 1987.

[21] Baker CF, Fillmore CJ, Lowe JB. The Berkeley FrameNet Project. In: Proceedings of the 17th International Conference on Computational Linguistics - Volume 1. ACL '98/COLING '98. USA: Association for Computational Linguistics; 1998. p. 86–90.

[22] Narayanan S. Moving Right Along: A Computational Model of Metaphoric Reasoning about Events. In: Proceedings of the Sixteenth National Conference of Artificial Intelligence. Menlo Park; 1999. .

[23] Pfau J, Porzel R, Pomarlan M, Cangalovic VS, Grudpan S, Höffner S, et al. Give MEANinGS to Robots with Kitchen Clash: A VR Human Computation Serious Game for World Knowledge Accumulation. In: Entertainment Computing and Serious Games. vol. 11863 of Lecture Notes in Computer Science. Springer; 2019. p. 85–96.

[24] Hedblom MM, Kutz O, Mossakowski T, Neuhaus F. Between Contact and Support: Introducing a logic for image schemas and directed movement. In: 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017); 2017. p. 256–268.