NOTICE

Are there Semantic Primes in Formal Languages?

Johannes Fähndrich, Sebastian Ahrndt and Sahin Albayrak

DAI Lab, Berlin Institute of Technology, Berlin, Germany
{faehndrich,ahrndt,albayrak}@dai-labor.de

Abstract. This paper surveys languages used to enrich contextual information with semantic descriptions. Such descriptions can be e.g. applied to enable reasoning when collecting vast amounts of row data in domains like smart environments. In particular, we focus on the elements of the languages that make up their semantic. To do so, we compare the expressiveness of the well-known languages OWL, PDDL and MOF with a theory from linguistic called the Natural Semantic Metalanguage.

Keywords: Context Description Languages, Contextual Reasoning, OWL, PDDL, NSM

1 Introduction

Intelligent environments are made up of multiple pervasive or ubiquitous devices that provide a service to the user. One key indicator of such environments is the ability to adapt to changes. The changes are implied by external or internal influences like the introduction or removal of devices or changing application goals [20]. We expect that intelligent environments react to such changes and adapt themselves in a way that the service provided are still available for the users. More than a decade ago *R.J. Sternberg* [22] still emphasises this specifying that intelligence is the ability to adapt to changes in environments (to distinguish between the environment itself we refer to this as context). This point of view implies that an environment becomes more intelligent if it can cope with more or bigger changes in the context. To be able to adapt to contextual changes a cognition is needed to be aware of the actual context and appearing changes. We focus our analysis to environments where such cognition is available. That means, that there exist at least one entity able to perceive the context and able to communicate the actual perception to other entities in the environment.

One can distinguish two types of contextual information: The **defined context** and the **derived context** [13]. In both cases, the devices making up the intelligent environment have to agree on a language to interpret the data collected by the sensing devices. In a defined context (e.g., a specific application) this language can be given to the environment by a domain model. Another approach is to use semantic languages to annotate contextual information and use reasoner that derive knowledge or facts from this annotations. This approach is called derived context. Here every device has its own local model of the environment, without having to agree on a global context model providing information about all devices. Derived context is created by finding patterns in raw data form the sensing devices of an intelligent environment and annotate them with the given semantic language. A reasoner then reasons upon this annotated information to transform the information into a local domain model of the device. This emphasises the requirement to agree upon a semantic language used for the annotation. Furthermore, it underlines why no model of the whole context is needed.

Languages to describe semantic have been subject to research in many research areas. *Bikakis et al.* [1] surveys semantic based approaches and applicable reasoning methods in the domain of ambient intelligence. Two of them are the Semantic Web community and the Agent community. Both have developed a quasi standard language to describe semantics. In the semantic web community the Web Ontology Language[1] [17] (OWL) is been widely used. The agent community uses Planning Domain Definition Language[2] [7] (PDDL) to describe their planning problems. This paper will examine the fundamental concepts making up those two languages. Additionally the study includes the Meta Object Facility [6] (MOF) as a meta-language for artificial languages. We compare these approaches with a theory form linguistics named the Natural Semantic Metalanguage [8] (NSM). NSM states that every naturally developed language is based on 63 semantic concepts.

The paper is structured in the following way: In Section 2 introduces NSM and the basic concept of semantic primes in a nutshell. Furthermore, it describes the three semantic description languages and the difference between the languages. Section 3 takes such insights into account and compares the languages in a more detailed way. Afterwards, Section 4 wraps-up the paper with a discussion of the results.

2 Semantic Primes

The Natural Semantic Metalanguage (NSM) is a linguistic theory originated in the early 1970s [25]. It states that each meaning of an concepts created in a natural language can be represented using a set of atomic terms—so-called universal semantic primes. These primes have an indefinable word-meaning and can be identified in all natural languages [9]. In conjunction with associated grammatical properties NSM presents a decompositional system able to describe all concepts build in the appropriate language. Here, an expression is decomposed into less complex concepts, where the process ends if the expression is decomposed to the atomic level of semantic primes which can not be analyzed further. One can imagine that the decomposition builds a tree, where all leafs are semantic primes [27]. Consequently for each natural language a metalanguage exist that consist of the semantic primes in the specific syntax and their appropriated grammatical properties. About 63 semantic primes exist that can be divided into 16 categories [26].

As well as natural languages, formal defined artificial languages are based on a metalanguage like the Meta Object Facility. This leads to the implication that the concepts defined in artificial languages are semantic primes and that such primes can be compared among different languages. Since the bag of semantic primes presented by NSM

[1] For further information the interested reader is also refereed to: `http://lists.w3.org/Archives/Public/www-webont-wg/2001Dec/0169.html`

[2] For further information the interested reader is also refereed to: `http://www.cs.yale.edu/homes/dvm/`

is empirically well-researched, this work tries to compare three artificial languages utilising this bag of primes. For this comparison, we take the purpose and concepts of the languages into account and match the available primes with each other as foundation to discuss potentially missing primes in the languages.

Table 1. List of semantic primes with no equivalent found in the other languages.

Category	**Semantic prime**
Substantive	SOMEONE, PEOPLE
Determiners	OTHER/ELSE
Quantifiers	TWO, MUCH/MANY
Evaluators	GOOD, BAD
Descriptors	BIG, SMALL
Mental predicates	THINK, KNOW, WANT, FEEL, SEE, HEAR
Speech	SAY, WORDS, TRUE
Actions, events, movement, contact	DO, HAPPEN, MOVE, TOUCH
Space	WHERE/PLACE, HERE, ABOVE, BELOW, FAR, NEAR, SIDE, INSIDE
Intensifier, augmenter	VERY, MORE
Similarity	LIKE
Time	WHEN/TIME, NOW, FOR SOME TIME, MOMENT
Logical concepts	MAYBE

Table 1 lists all semantic primes presented by NSM not matchable with any concept in the examined artificial languages. In the following we will introduce the artificial languages and the list for each language, which semantic primes are used.

The Web Ontology Language (OWL) is a semantic markup language to create structured knowledge representations and enrich them with semantics. OWL is a W3C standard since 2004 and has been continuously developed since [11]. OWL is an extension of the Resource Description Framework [15] and has become one of the most used language to describe knowledge for AI. Since OWL is meant to describe structured knowledge the concepts used are abstract. Table 2 list all equivalents found in comparison with NSM primes.

Table 2. List of semantic primes with and equivalent found in OWL.

Category	**Semantic prime**	**OWL**
Substantive	I	self.entry
	SOMETHING/ THING	owl:thing
Relational	KIND	owl:SubClassOf
substantives	PART	owl:topObjectProperty
Determiners	THIS	owl:entityURI
	THE SAME	owl:equivalentClass

The Planning Domain Definition Language (PDDL) is a first-order logic based language defined as an extended BNF [7]. Commonly, it is used to provide a standardized way to describe planning problems and the associated domains. The syntax allows to define among others actions, effects, quantifications and constraints and was intended to enable developers to describe the ”physics” of a domain. Given such a description the reasoner uses a goal defined in PDDL to search for a plan that satisfies all constraints, requirements and preconditions. The concepts which are equivalent to semantic primes are listed in Table 3.

Table 3. List of semantic primes with and equivalent found in PDDL.

Category	Semantic prime	PDDL
Substantive	SOMETHING/ THING	:define
Determiners	THIS	::=
Existence	THERE IS	:exists
Time	BEFORE	:precondition
	AFTER	:effect
	A LONG TIME	:maintain
	A SHORT TIME	:wait-for
Logical concepts	NOT	:not
	CAN	:action
	BECAUSE	:imply
	IF	:when/constrained

The Meta Object Facility (MOF) has been introduced by the Object Management Group and is formally defined e.g. by *Smith et al.* [21]. MOF has been developed to model the structure and behaviour of entities in software development. For example, UML[3] implements MOF. Since MOF is quite abstract, the meta language like OWL mostly has structural semantic primes. Table 4 list all equivalents.

Table 4. List of semantic primes with and equivalent found in MOF.

Category	Semantic prime	MOF
Substantive	YOU	uri
	SOMETHING/ THING	object
	BODY	instance
Relational	KIND	type, extent
substantives	PART	property
Determiners	THE SAME	element.equals
Quantifiers	ONE	multiplicityElement
Location	BE (SOMEWHERE)	link
Existence	THERE IS	element
	HAVE	classifier
	BE (SOMEONE/SOMETHING)	extend
Life and death	LIVE	create
	DIE	delete
Logical concepts	CAN	operation
	IF	event

[3] see: http://www.omg.org/spec/UML/

3 Comparison of Primes

The compared languages introduce additional concepts that are domain specific and which are not part of the semantic primes (e.g., 'OWL:VERSIONINFO'). Depending on the purpose of the language those additional concepts change. OWL for example was created to describe shared conceptualizations where versioning and backward compatibility is an issue. But from the theory of NSM those concepts could be described using the other semantic primes. Thus they are merely shortcuts. There are multiple extensions to those languages for special cases like the Semantic Web Rule Language (SWERL) [14], which introduces rules to OWL. We now discuss the 16 categories of semantic primes to analyse why such concepts do or do not exist.
Substantives are the first category. In natural language these semantic primes are used to distinguish actors and to separate humans from other things. To describe meaning, humans often reduce description of properties of things to the relation to humans or more precise them self [27]. For example, to describe the concept 'mouse' a semantic, context independent description most likely rely on a degree in Biology. Describing a mouse so that a naïve reader of the description might understand it, the description can refer to the potential context of the reader. In natural languages these readers are other humans, which implies that most description in NSM are in the context of humans and their relation to things. The distinction of 'YOU' and 'I' is needed if roles are described, i.e. in negotiation or contracting. Non of the reviewed language has a concept for 'PEOPLE'. On the one hand semantic description in these languages are thought for artificial reasoners and there a concept of 'PEOPLE' is not needed to describe most concepts. On the other hand, in the area of HCI the distinction of artificial agents and human agents can be of some concern and with that the concept of 'PEOPLE' might be required.
The category of *relational substantives* are well represented in two of the languages, except PDDL as it does not use type hierarchies to define domains. That means that PDDL does not semantically aggregate all instances of one 'KIND'. In semantic descriptions 'KIND' is for example used to describe water: 'something of one kind' [10].
Quantifiers are represented in all tree languages. The exception is the fuzzy representation of 'MUCH/MANY' and 'SOME'. There is a need to enable fuzziness in semantic description languages as motivated by *Stoilos et al.* [23].
The category of *Evaluators* is not represented in any language. PDDL from version 2.1 features numeric fluent to describe e.g., cost for actions, which could be interpreted depending the metric. Here an implicit metric is given to the reasoner, e.g. the *plan minimization metric* [7]. This metric states that the value to be optimized is better the smaller the value. Stating that less is better. We argue that such a metric can be explicitly formalized in the description itself and to define what is 'BAD' or 'GOOD' those concepts need to be part of the description language.
Descriptors are not represented in all three analyzed language. 'BIG' and 'SMALL' are fuzzy values and are defined in an description. For example, *Wierzbicka* [27] describes mice as small in the following way: 'They are very small. A person can hold one easily in one hand'. Giving a example on what small means and a relation to something every reader of the explanation knows: the size of a hand. In relation to some reference or as constant like the semantic primes 'A LONG TIME' and 'A SHORT TIME' these can

be used to describe relations in e.g. size, intensity, power or time. We can imagine that for example a timeout can be explained by defining ‘A LONG TIME’ as the maximum timeout. We argue that if the semantic description is used by a reasoner to create a heuristic, a metric needs to be defined and with that semantic descriptors are needed which classify the value which is subject to the metric. Further we argue that ‘LONG’ and ‘SHORT’ should be part of the descriptors since they are descriptors which could be used in addition to time with other concepts like spacial distances.

Mental predicates are not represented in all three languages doe to the fact that these predicates are based on the senses of human being. We separate mental predicates in two groups: The first group is based on human cognition, ‘FEEL’, ‘SEE’ and ‘HEAR’. First of all there are two of the senses missing: ‘SMELL’ and ‘TASTE’. Additionally for the domain of intelligent environments and if sensors needs to be described, such predicates can be used. Of cause this is specific for humans, since not every agent involved in an intelligent environment has such cognitive functions. Additionally there could be sensors which extend the human cognitional like a barometer, altimeter or a localization like GPS, which should then be incorporated in the semantic primes as well. Even thou these concept are not used in the analysed modelling languages, they are part of the semantic description of many fundamental concepts like ’HANDS’ [27]. The second group of mental predicates is the mental state of mind: ‘THINK’, ‘KNOW’, ‘WANT’. These are philosophical terms and rarely used in artificial languages. Braubach et al. [2] e.g. describe a Believe, Desire, Intent (BDI) paradigm for agents. Here ‘I believe’ is considered a subset of ‘I think’ [26]. In the BDI paradigm ‘believe’ can be mapped to the semantic prime ‘KNOW’, since it represents the knowledge of the agent and ‘desire’ can be mapped to ‘WANT’ since it describes the internal goals of the agent. But in our analysed languages, all of these concepts are missing.

Speech is - at first - the category which hold one of the basic logical operators ‘TRUE’. All three languages use an implicit representation of the concept ‘TRUE’ since they assume that a reasoner interprets an axiom as fundamentally true. PDDL e.g. ‘define a model to be an interpretation of a domain’s language that makes all its axioms true’ [16]. Thus here again it can be argued to explicitly describe such truth values and with that add a semantic prime to the metalanguage. But we think that ‘TRUE’ should be caped in the category ‘Logical concepts’. Further we use ‘WORDS’ as basic building blocks for our description, and thus need a semantic prime for them. ‘SAY’ has been represented in a formal way as agent communication speech acts [4] and could be directly part of the metalanguage.

Semantic primes in the category *Action* are often defined in an context dependent manner, where the semantic is given by the reasoner of the evaluating the axioms. In PDDL for example the blocks world defines‘MOVE(A,B,T)’ [12]. NSM proposes to add such primes to the metalanguage, to be able to describe events, movement and actions.

The semantic primes in the category *possession* e.g., ‘HAVE’ can be seen as the specialization of composition and aggregation of the semantic prime ‘PART’. The *specification* ‘BE’ denotes a location where something is located and at the same time to be of one type.

Life and death is a category which is not subject to research in formal languages because computer systems rarely need a concept of death or living. Semantically there are

many things that can not be described without the concept of ‘LIVE’ and ‘DEATH’. In agent communication e.g.m agents send ‘alive messages’ to other agents, where the interpretation is left to the programmer of the agent.
Time has found its way into almost every formal language. Even a own logic—the Temporal logic—which is a kind of modal logic has been created to model something like: ‘I am hungry until I eat something’. In formal languages time has often been included into the language e.g., PDDL from version 2.1.
The category *space* is subject to research and is formulated in contextual models like the CORBA-ONT [3]. Nevertheless non of the surveyed languages presents primitive elements to describe special properties. The fact that a OWL ontology is required, shows that such semantic primes are necessary for the modelling of contexts. The same can be argued for the semantic primes ‘BE (SOMEWHERE)’ of the category location. Other fuzzy primes like ‘NEAR’ or ‘FAR’ are again hard to grasp in a formal language.
Logical concepts are of cause part of most formal languages. The hurdle primes are again the fuzzy ones: ‘MAYBE’ and ‘CAN’. To describe the meaning of probability, those primes could be part of a language like they are in epistemic logic [24].
Intensifier can be modeled as lexical functions [18] and fuzzy decision systems have been subject to research [19] and thus to make their semantic explicit intensifier should be part of the metalanguage.
Similarities are a huge research area and measures have been studied in depth [5]. The here developed methods like recommender systems try to find entities which are alike. Those methods try do define the prime ’SAME’ for different domains.

4 Conclusion

We have analysed three common semantic description languages and compared their meta languages with the set of semantic primes taken from NSM. We have found that already many of the semantic primes are part of the three formal description languages depending on their focus. The semantic primes that are not yet part of the description languages has been collected in Table 1. Future work will include an in-depth analysis of those primes. Here we want to examine which primes are useful for formal languages and define a set-theoretic semantic for each of them.
Acknowledgement: Research in this paper has been financed by the Schaufenster Elektormobilität http://www.schaufenster-elektromobilitaet.org in the project 16SBB007A.

References

1. Antonis Bikakis, Theodore Patkos, Grigoris Antoniou, and Dimitris Plexousakis. A survey of semantics-based approaches for context reasoning in ambient intelligence. In *Constructing Ambient Intelligence*, pages 14–23. Springer, 2008.
2. Lars Braubach, Alexander Pokahr, and Daniel Moldt. Goal representation for bdi agent systems. *multi-agent systems*, pages 44–65, 2005.
3. Harry Chen, Tim Finin, and Anupam Joshi. An ontology for context-aware pervasive computing environments. *The Knowledge Engineering Review*, 18(03):197–207, 2003.

4. Marco Colombetti and Mario Verdicchio. An analysis of agent speech acts as institutional actions. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3*, pages 1157–1164. ACM, 2002.
5. Richard O. Duda, David G. Stork, and Peter E. Hart. *Pattern classification and scene analysis. Part 1, Pattern classification*. Wiley, 2 edition, November 2000.
6. Jean-Marie Favre. Foundations of meta-pyramids: languages vs. metamodels. In *Episode II. Story of Thotus the Baboon, Procs. Dagstuhl Seminar*, volume 4101, 2004.
7. Maria Fox and Derek Long. PDDL2.1: An extension to PDDL for expressing temporal planning domains. *Journal of Artifical Intelligence Research*, 20:61–124, Decemeber 2003.
8. C. Goddard and A. Wierzbicka. *Semantic and lexical universals: Theory and empirical findings*, volume 25. John Benjamins Publishing Company, 1994.
9. Cliff Goddard. *Cross-Linguistic Semantics*, chapter Natural Semantic Metalanguage: The state of the art, pages 1–34. John Benjamins Publishing Company, 2008.
10. Cliff Goddard. Semantic molecules and semantic complexity:(with special reference to" environmental" molecules). *Review of Cognitive Linguistics*, 8(1):123–155, 2010.
11. Bernardo Cuenca Grau, Ian Horrocks, Boris Motik, Bijan Parsia, Peter Patel-Schneider, and Ulrike Sattler. Owl 2: The next step for owl. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):309–322, 2008.
12. Naresh Gupta and Dana S Nau. On the complexity of blocks-world planning. *Artificial Intelligence*, 56(2):223–254, 1992.
13. Karen Henricksen and Jadwiga Indulska. Modelling and using imperfect context information. In *Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on*, pages 33–37. IEEE, 2004.
14. Ian Horrocks, Peter F Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, Mike Dean, et al. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member submission*, 21:79, 2004.
15. Ora Lassila and Ralph R Swick. Resource description framework (rdf) model and syntax specification, 1999.
16. Drew McDermott. The formal semantics of processes in pddl. In *Proc. ICAPS Workshop on PDDL*, pages 101–155. Citeseer, 2003.
17. Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(2004-03):10, 2004.
18. Igor Melčuk and Leo Wanner. Lexical functions and lexical inheritance for emotion lexemes in german. *Lexical functions in lexicography and natural language processing. Amsterdam/Philadelphia. John Benjamin*, pages 209–278, 1996.
19. Grundlagen der Theorie Unscharfer Mengen. Fuzzy decision support-systeme. 1994.
20. Mazeiar Salehie and Ladan Tahvildari. Self-adaptive software: Landscape and research challenges. *ACM Transactions on Autonomous and Adaptive Systems*, 4(2):1–42, May 2009.
21. Jeffrey Smith, Scott DeLoach, Mieczyslaw Kokar, and Ken Baclawski. Category theoretic approaches of representing precise uml semantics. In *of the ECOOP Workshop on Defining Precise Semantics for UML*, 2000.
22. Robert J Sternberg. *Beyond IQ: A triarchic theory of human intelligence*. 1985.
23. Giorgos Stoilos, Nikos Simou, Giorgos Stamou, and Stefanos Kollias. Uncertainty and the semantic web. *Intelligent Systems, IEEE*, 21(5):84–87, 2006.
24. Wiebe van der Hoek. *Epistemic logic for AI and computer science*, volume 41. Cambridge University Press, 2004.
25. Anna Wierzbicka. *Semantics: Primes and Universals*. Oxford University Press, 1996.
26. Anna Wierzbicka. *English: meaning and culture*. 2006.
27. Anna Wierzbicka. Mental lexicon. In *The Slavic Languages: An international handbook of their history, their structure and their investigation*. Mouton de Gruyter, Berlin, 2009.