Formalizing Semantic of Natural Language through Conceptualization from Existence

Duan Yucong, Christophe Cruz

To cite this version:
Duan Yucong, Christophe Cruz. Formalizing Semantic of Natural Language through Conceptualization from Existence. International Journal of Innovation, Management and Technology, 2011, 2 (1), p. 37-42, ISSN: 2010-0248. hal-00625002

HAL Id: hal-00625002
https://hal-univ-bourgogne.archives-ouvertes.fr/hal-00625002

Submitted on 20 Sep 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Abstract—We propose an outline of an approach to formalize semantic from conceptualization for both natural language (NL) and logic expression mechanisms. This goes beyond the level of discussions at conceptual level which has to either end in conscious/unconscious relativity of understanding or subjective enforcement in the form of definitions instead of expected objective semantic. This approach supports to view from a pure mathematical perspective, and explore and locate the fundamental problems. The semantic formalization mechanism realizes the integration of problem description and the solution expression at absolute semantic level. So a problem describing process is equivalent to the solution exploring process by integrating both in one. This essentially caters the ideology of proceeding with model refinement of model driven development. Other advantages include that it will reduce the need for validation for model migrations during a model driven development process, etc. Application is intended to cover specification refinement of both functional and quality requirement, and both static description and behavioral implementation, etc.

Index Terms—Semantic, knowledge, cognitive, formalization, conceptualization.

I. INTRODUCTION

For requirement phases of software development processes, it is important to get to know the information/intention of users to support modeling and implementation hence after. The key is to locate and determine the semantic [1], [2], [3] of the requirement specification expressions which is however not readily available. Our approach is proposed for this purpose. It is expected to cover all situations including incomplete, vague, inconsistent, and even not expressible situations of information. We would like to see scenery of whole picture of behavior [9] supported by integrated/whole picture of formalism.

A. Revelation on current situation

Why and how we can get to each other in spite of those argumentations of “expressible vs. not expressible” and “communicable vs. not communicable” [1], [2]?

Firstly, there are lots of unconscious not expressible and not communicable happen which just are not noticed usually. These situations could have been mistaken as expressible and communicable (notice: here the NL explanation actually needs to be transferred from implicit to explicit. However, at conceptual level, this is not feasible). And the even the rectifying approaches can be enforced and guided with feedbacks for some cases instead of demanding the explicit revelation in a theoretical manner.

Secondly, if the core of the semantic evolution mechanism is the same dualism then there will be same understanding/semantic as a result of the inspiration on the semantic evolution mechanism based on the dualism. But the same dualism is not necessary since that if some E =1 of semantic cannot be communicated, and then there will be no way to validate whether there are the “same”. What can be confirmed is that there exists consistency on both sides.

B. Proposition for formalization

An issue which deserves to be mentioned is that: sometimes customers who raise the requirement specifications might not really understand or keep conscious with what they have expressed. But if they really understand, it can be derived from this proposition that they will have the integration of the problem description and the corresponding solution expression. So the process involves enlightening the customer to attain a formalization of the semantic which they intend.

Here we use enlightening instead of transfer because that semantic cannot be transferred for cases that the existence which cannot be transferred composes the content to be transferred. The rules can be found in [1], [2]. An outside extension: we understand only when we understand. For this case, enlightening for the minds of individuals could be the only means for achieving consensus.

C. Motivation and goals

To our knowledge, no existing approach which deals with semantic formalization has touched so depth at the level of existence. Hence after, no approach for validation on static and behavior sides of a system can do more than what is expressed in the semantic of the modeling process. What is not touched is beyond conceptual level. This could be problem ranging from very small issues to the whole since that a CPT as a notation for semantic is actually a free variable which can represent anything from a pure sense.

It is expected to touch the fundamental issues about semantic which will be extended to the extent of both expressible and “expressible not expressible”. The background underlying this expression can be interpreted as extending from \{OWA\} \rightarrow \{CWA, OWA\} [10], [11],[12].
The ideal start point is to start from an objective semantic of the problem description. This requires a clarification [2], [8] mechanism for semantics of NL expressions which can be classified as natural language processing (NLP).

Target: identify what is actually demanded for both customers and the implementation. The objective vs. subjective results should be firstly identified, and then dealt differently.

The rest of this paper is organized as follows. Section II presents related preparation for the subsequent discussion and understanding. Section III presents the strategy which would be used for the exploration. Section IV shows potential implementation with initial cases of applying the proposed approach. Section V discusses related methodology and ideas in comparison to our approach. The last section summarizes this work in progress.

II. PREPARATION

We see a question as a view to draw attentions to the target. We propose initial views on the topic through raising questions. We clear the standpoint towards attaining solutions here.

A. Questions which would be raised for enlightening before the beginning

Among numerous potential questions, we propose the follows:
1) Are there some technical difficulties which are still left unsolved? What are them?
2) Are there some theoretical challenges which define the scope of what can and cannot be accomplished? What are them?
3) For the confirmation of the existence of the above questions, are these positive answers expressible (being able to be modeled) or not in a formal semantic level with mathematical precision? If there are some of them not expressible, why and how?

B. Specifically catering refinement of requirement specifications

1) What will happen in customers’ minds?
2) What will not happen in customers’ minds?
3) What will happen in customers’ expressions?
4) What will not happen in customers’ expressions?

Think and express these answers from the view of a third party. Intuitively we cannot express in a positive and complete manner of what will or will not happen in customers’ mind and expression without reaching an abstractive level. We can be positive with that these answers at specific level will be not enumerable. Suppose the answers can be mapped to Y/N or T/F, and then they can be used as coordinates for scale down answers for less fundamental questions. We propose to reach some deterministic content related to these fundamental questions.

C. Intuitions on formal vs. not formal expression

Why start formalization from the beginning is necessary: Assuming there is counterpart for CPT of formal expressions, let’s say it as empirical. (* actually, from the view of our formalization, counterpart for CPT of formal expressions cannot be derived from the formalization mechanism. In another word (* let’s avoid such conjunctions of “because”, “then”[12]), it is not expressible at the formal level of expression even if its existence(E=1) is assumed in the background of CWA. Put it a little bit forward, even if it can be reached through the formalization of our approach, it is still not expressible/discussible more than the “E=1” which can be assumed. It can be revealed as that at the time of the CPT is reached with our formalization mechanism, its assumed existence changes simultaneously. It is not (negation) itself (which could be expected to be the target of previous discussion) anymore. anymore.

Positive description of the formalization: Semantic will be always traceable to the evolution of the formalization mechanism. The E=1 of the trace is bounded with the confirmation of the formalness of the semantic.

Reachability of formalness: No formal level semantic can be reached from empirical semantic and vice versa.

It can be used to deny the contribution of related not formal contents of previous works [1], [2], [3] at formal level. To strengthen the point, it can be put that:
1) <contribution>::= formal contribution.
2) To be complete: there are no other routines. Contributions have to contain formal contributions or at formal level to be reused by others as beneficial thoughts.
3) Considering expressible vs. not expressible: conclusion of above applies to both.

Attitude towards the enforcement of standards instead of exploring formal semantic: If the formal semantic is missing, unified semantic cannot be really enforced/attained with the absence of pure formal semantic.

D. From Expressible towards communicable:

A complete exploration of semantic involves explore semantic from its origin. We start the revelation from expressible vs. not expressible.

1) Expressible vs. not expressible:

If we view that the every conscious thought is bounded with an existing language. Or we thought with our language system. The coming into existence of any language system will be a result of conceptualization. Then we cannot reach out to anything which is related to the evolution mechanism of conceptualization in a consistent/continuous manner.

When the thought is bounded to the semantic evolving system, it is cannot reach anything which cannot be reached. Or there is nothing which is not expressible within the reachability. This situation maps to the background of OWA where there is no valid “negation”. In another word, even if a CWA is created to identify the E=1 of not expressible in comparison to the previous situation of no not expressible, the situation of E=1 of not expressible will automatically transform to E=1 of expressible synchronically at a semantic level.

A similar case of not expressible is for the CPT of “incomplete”/"absent": if the semantic of incompleteness or absent exists, it must be from the view of an indirect view of CWA other than the view of OWA where the E=1 of it is founded. While the semantic of it cannot be transferred as
with the $E=1$ of it in the form of $(E=1)_{OWA}$ since that the CWA and CWA is not compatible as an ultimate background for “negation”.

Ability to be expressed: just like the negation connective is valid as $E=1$ after the valid of the positive $E=1$. An interpretation, CWA comes after OWA in the evolution of the formalization mechanism: $\{OWA\} \rightarrow \{CWA\}$. This case is unnoticeable for pure descriptions where OWA and CWA can be put together as pure notations which do not really adhere to the formal semantic. An initial discussion on logic connectives can be found in [12].

2) Expressible vs. communicable:

For semantic which is identified as expressible, there will be another issue ahead: transferring from one mind to another mind.

The rule is that:

Only semantic/information which is explicitly complete can be transformed. (compete/independent::= “$E=1$”)

If there are multiple complete semantics, all of them will be equally transferred which is irrelevant to the preference/intention of the speaker. As an extension, this is a reason why misunderstanding exists even if the completeness is guaranteed. It is also the source of many language jokes originate in: the mismatch of the intended semantic.

3) More detail about what is transferable vs. not transferable

In case of $E=1$ which is not transferable by itself, the conceptualization of it can be transferred since that every conceptualization forms a CPT to represent the original intention with a notation. This process could be a replacement of the original intentions if not noticed or can be unconsciously.

For the case of $E=1$ which is not transferable directly, the $E=1$ of the negation of the transferability in the background of CWA can be transferred instead.

E. Relationship to reuse

From our analysis, reuse of thought::= reuse of semantic. Reuse of semantic::= is based on OBJ semantic instead of SUBJ semantic. Objective semantic::= formal semantic. Explanation for potential argumentations: the situation of reuse SUBJ semantic actually goes through an implicit transformation of SUBJ$\rightarrow$OBJ cognitively which can be revealed by explicitly denote the route of the target information transferring from SUBJ of specific individual to OBJ shared by more than one mind. It experiences the transformation of $(E=0)\rightarrow(E=1)$. To be complete: there are no other means of realizing semantic reuse than going through formalization which in our case through conceptualization.

F. The attitude towards the understanding of the draft

(i) You get what you see:= you get what you know.
(ii) Enhancement can be only expected through enlightening.
(iii) We can only focus on (ii).

III. STRATEGY

A. For formalizing from no concept

Our observations on the difference of the second language learning process by kids and adults from different mother languages matches to the explanation of theory of MDR(model dependent reality)[4]. It can be seen that human learn new things with their existing language system. Their senses of hearing get influenced. We expect to see similar influence for reasoning. It might be true, for example, if the existing reasoning system is the only choice for thought. What we would like to put it a little bit forward is that do we differ at the very beginning from nothing more than existence.

1) If from nothing to something, we experience the same conceptualization. It can be expected that we are destined to be the same in spite of all the difference at superficial level.

2) If we differ for this process, it will be another case, which we will investigate in the future.

An existing opinion from linguistic area is that human think with their language systems. We would like to extend it as that human formalize semantic with their existing language systems as well. The language systems could be different in some aspects. However if some seemingly different languages are evolved from the same conceptualization mechanism, their semantic will be the same essentially.

B. The attitude/strategy to unknown questions on existence

Here we explain our attitude/strategy to face questions on existence without existing determinate answers. Example of such questions:

1) Is there $E=1$ of a core of natural language which can reach the full expression power of the whole language?
   If it is true, what is it?
   It should be true as long as there are overlaps which can be reduced. Further criteria might be built on fundamental issues related to existence, e.g., the content of $E=1$ cannot be reduced if there is not an overlap and the content of $E=0$ can be, etc. Of course to understand the semantic does not apply for the idea of a core, since that every piece of the expression has to be present in the original expression.

2) Is there $E=1$ of not expressible by natural language?
   We propose to find out answers to these problems. Also before the reaching of the answers to these problems, if the answers are identified as necessary for proceeding to some solutions, our strategy is not leaving a gap/blank there which invalidate the continue of the transfer of consistency, but making hypotheses of $\{T/F\}$ on them and proceed from all the possible branches. With this manner, an answer cannot be missed from the limited amount of extended branches.

C. On the integration of “problem” and “answer”

At the level of $E=1$ or both real world and completeness based, related CPTs and expressions will be evolved out with consistency. Other CPTs and expressions if not related will be not confirmed with the consistency which we have identified as key factor of formal semantic. If viewed from
the view at conceptual level of CWA vs. OWA, the evolution is with OWA. There will be no “no” for E=1 for other CPTs and their expressions. These expressions include “problem expressions”. It can be explained that all what is meaningful is expressed out explicitly, and there are no place for the CPT of “problem/question” and the expressions intended by this CPT. Or we can say the problem and the answer are integrated if we has to use the CPT of “problem” and “answer” at conceptual level. A retrospect from the view of formal semantics, the existence of so called problems and answers are at conceptual level of semantic and suffers from the relativity of conceptual level when attempts were made towards achieving an ultimate answer. Formal semantic analysis: (ultimate/objective)\(\iff\) (E=1). A lot of so called problems are expressions which need to be answered due to the implicit semantic. The ideal answers actually supplies the missing semantic of the incomplete semantic and form complete semantic explicitly.

D. Argumentation for techniques
1) Why and how Y/N vs. T/F flow works?
   It is actually a conceptual extension of Dualism. The Y/N represents the necessary human side SUBJ decisions and T/F represents the OBJ semantic which can be related to E=1. In this manner, it however maintains a completeness from the integration of Dualism, and completeness and consistency can be maintained by following the guidance of semantic evolution and conceptualization. The formalness of the evolved semantic for CPTs can be testified by the criteria of what can be tracing back to along the semantic evolution trace. From an outside example view through OWA vs. CWA, the evolved semantic will be formal for logic connectives which have been identified as informal in the usual expression [ ].

   The evolution mechanism of semantic will fully implement both aspects of our hypothesis on the nature of computation (CP) as \(<\text{CLA},\ \text{ORD}>\)[]. On the other hand, when view CP as a CPT, what can be derived for a CP is no more than \(<\text{CLA, ORD}>\).

   By revealing semantic in a complete manner, the expected result will cover both explicitly what can be done/feasible and implicitly automatically deny what cannot be achieved (not what haven’t been touched.). It means that all that which can be processed without other individualized conditions will be processed and all that which cannot be processed will be excluded as well.

2) Steps for implementation
   (i) Applications will start from revealing the implicit side of existing semantic and construct corresponding complete semantic through explicitly supplementing the missing part. The transfers include: implicit \(\rightarrow\) explicit and incomplete \(\rightarrow\) complete. Here at conceptual level, explanations will suffer from the relativity of conceptual level, e.g., explicit could be identified in SUBJ manner as either complete or incomplete without notice if backgrounds are missing.

   (ii) Then the complete semantic will be sorted/developed as in the form of Y/N flows for representing necessary SUBJ interventions which can be reduced to one if connected and T/F flows for representing automatic processing/computing steps by a machine.

   The result will appear to be quite simple, but it might be the boundary of what can be progressed towards or achievable and cannot be surpassed at fundamental semantic level.

IV. TOWARDS INITIAL CASES

A. An example revelation with logic connectives
   Many previously intended OBJ expressions could end up as being revealed as SUBJ, the cases include the logic connectives of “T/F” which could be revealed as implicit semantic with the background of CWA vs. OWA [12]. Then any specific deterministic with either CWA or OWA implicitly could be identified as SUBJ instead of the claimed OBJ since that the deterministic is supported by not enough or incomplete OBJ evidence. Under these incomplete/implicit circumstances, any decision will be SUBJ instead of OBJ.

B. On employing T/F and Y/N flows
   1) “T/F vs. Y/N”
      T/F can be traced by formalization mechanism which is based on conceptualization from E=1/0. They will be complete/\(\iff\) automatically.

      Y/N is not related to the formalization. It is not expressible when it is assumed to be E=1 with CWA at the level of formalness, and not discussible at formal level as there is a transformation for its existence from E=1 to E=0 at the moment formalization reaches it.

   2) “T/F vs. Y/N” vs. consistency
      Consistency[2] is the first quality which is more fundamental than “T/F vs. Y/N”. It is “related” by itself. Even “not related” has to rely on the presence of it. This discussion implies the fundamental order for OWA vs. CWA as OWA\(\rightarrow\)CWA at this stage. Extended implication for consistency:

      It is all what can be really gained from formalization of semantic;

      From a view of CWA: it can be repeated or emphasized as no more fundamental can be expected/achieved.

      However the meaning of that consistency is the only fundamental semantic at a pure reasoning level could be easily misused for engineering applications where “T/F” represents the OBJ of the E=1 of an engineering target which goes through the explicit or implicit transformation from SUBJ to OBJ along with the transformation from E=0 to E=1.

C. An example on a case of fallacy
   We are going to reveal the multiple semantics as the origin of misconception [5] fallacy for the following case.
   The first example [5]:
   1) Nothing is better than eternal happiness.
   2) A ham sandwich is better than nothing.
   3) Therefore, a ham sandwich is better than eternal happiness.

   Firstly this discussion is at conceptual level, so a semantic of E=1/0 cannot be revealed inside as a coordinate for avoiding the absurd which originates from inconsistency. But we can reveal that the inconsistency as a result of mistaking
unrelated semantic flows along with the notations of CPTs. The CPT of “nothing” has at least two semantics: (s1) nothing is a thing or something, and (s2) nothing is not anything which can be described or maybe exist. If they are viewed from conceptual level from introduced backgrounds of CWA vs. OWA, it can be explicitly distinguished as that (s1) is from OWA where everything which is evolved out is assigned synchronously with “it is a thing in OWA”. And (s2) is from the CWA. From our semantic evolution approach, this revelation can end here as that CWA and OWA are not supposed to be semantically valid [12],[13],[14] synchronically when they represent their semantics of the controversial semantics. Any subsequent inconsistencies or fallacies can be initiated from the source which is revealed here, as long as it is not explicitly solved.

V. RELATED WORK

A. Compared to Model-Dependent Realism

Hawking & Mlodinow [4] propose Model-Dependent Realism (MDR) “on the idea that our brains interpret the input from our sensory organs by making a model of the world. When such a model is successful at explaining events, we tend to attribute to it, and to the elements and concepts that constitute it, the quality of reality or absolute truth.” We would like to see this expression as utilizing “negation” for excluding the possibility of constructing a model which has the same E=1 as the reality does. If we interpret reality as semantic of our topic, we would agree with the coincidence of that at conceptual level, every definition or explanation referring to E=1 will be destined as relative. But this will not apply to situations where extended discussions are based on CPTs of E=1 which is attained at pure formal semantic level such as integers, etc.

However instead of stop argumentation at that individuals will have individualized realities which rely on individualized views. We contend that these views could be consistency as long as they might originate in the same base: dualism. In another word, while MDR proposes that we will have different interpretations of modern science, we propose that these interpretations if all of them are right, they will be consistent. We still see that there are possibilities of one reality.

Similar to what is expressed on that there is pointless [4] to argue whether a model is real or more real, we propose that the only target achievable is no more than consistency [4].

In our approach, reality can also be mapped at conceptual level to CPTs of T/F vs. Y/N, or SUBJ vs. OBJ. While we would like to go even deeper by arguing in the manner of avoiding the usage of any CPTs which include “reality”, “model”, “independent”, etc. From our approach, we do not assume that there is a shared semantic for any CPTs/words which are not evolved from our conceptualization mechanism. Instead, we assume no condition, or it can be interpreted that we assume that we share nothing/negative (vs. something/positive) beforehand except the perception/observation for existence/E. Our hypotheses which back our approach:

1) It is more efficient to start from nothing or no concept than start from something unclear explicitly. A controversial thought is why we should introduce CPTs which are not bounded with formal semantics in a pure sense. If not formalized semantic could contribute nothing for strengthening of the formalness positively if not in the controversial direction by increasing the amount of chaos.

2) This is the only choice towards formalization by starting from the level before existing CPTs. Otherwise either relativity will be achieved explicitly, or the exploration for the solutions will get lost unconsciously.

* Here we would like to add to our previous discussion on definitions [12]. There is no explicit/determinate unconsciousness. Or the state of unconsciousness cannot be expressed with “precise/formal” semantics. Or it cannot be able to be correctly defined since that every explicit expression will construct either formal semantic or meaningless compositions of notations. Another case is that chaos cannot be explicitly and formally introduced as pertaining to the semantic of “chaos” since that the semantic of “chaos” refers to negation of clear which is not expressive directly without a background of CWA instead of OWA.

We would also see Niels Bohr’s quote [6] of “It is wrong to think that the task of physics is to find out how Nature is. Physics concerns what we say about Nature.” as a shared thought. We would like to say that most semantics experts has thrown lots of effort to achieve not OBJ semantic but instead repeatedly replacing SUBJ semantic with their version of SUBJ semantic.

Compared to modeling with UML, this approach reaches more fundamental level for semantic than the meta-meta level modeling. Then there will be no problem for lacking of formal semantics for those modeling elements of UML which previously suffer from lacking formal semantic.

Similarly we can find some similar fractions of wisdom which is mentioned in expressions of extremeness of epistemological ideas empirically and intuitively in Tao Te Ching [7]. For example, the paragraph of “Ineffability or Genesis” can be viewed as metaphors of discussion on conceptual level and conceptualization.

By introducing our approach, we would like to expect extending the scope of current NLP content, e.g., distinguish the contents which is not able to be processed currently.

VI. SUMMARY

We propose to retrospect on existing and old problems from a fundamental semantic view backed by conceptualizations [15], [16], [17] which are beyond conceptual level. This approach actually provides a coordinate which is needed by missing previously for measuring/locating semantic for CPTs. This coordinate can be used as not only a start point for formalizing semantic with consistency, but also as basic criteria to evaluate semantic as formal or not formal, etc.

A. Techniques

1) Existence vs. identification for the start of conceptualization.
2) Y/N and T/F for transfer of “implicit $\rightarrow$ explicit” for identifying SUBJ vs. OBJ of expressions.

3) Y/N and T/F flows for gaps, overlaps and inconsistency identification.

4) Retrospect: the logic connectives.

5) Application: Proof and application of the conjunctures of mathematics.

B. Expected merits

1) Target problems can be investigated/clarified and distinguished as feasible or not.

2) For the feasible problems: the essence of the problems will be extracted towards mathematical expressions.

   The two characteristics include:
   
   (i) Completeness/$<>$; the formal semantic which is derived/revealed through our approach maintains a level of completeness of explicit either OWA or CWA.

   (ii) Consistency: a consistency will be attained as a consistent flow in the form of explicit Y/N and T/F flows. On the other side, the gaps and inconsistency will be identified through the transformation of semantic from implicit and incomplete toward explicit and complete: (implicit, incomplete) $\rightarrow$ (explicit, complete)[3], [12].

   We expect to see this mechanism works also for answering related linguistic problems.

REFERENCE


Yuncong Duan is a senior member of IACSIT. He received a PhD in software engineering from Institute of Software, Chinese Academy of Sciences, China in 2006. He was Post-doc fellow in School of Software, Tsinghua University, China from 2006-2007. He was a Post-doc fellow in the Software Engineering Laboratory at Pohang University of Science and Technology (POSTECH), South Korea, from 2007-2008. He was a lecturer of Biomedical Engineering Institute, Capital University of Medical Sciences, Beijing, China, from 2008 to 2009. He was a Post-doc fellow in Le2i, CNRS, University of Bourgogne, France, from 2009-2010. He has been a member of IEEE since 2005. His research interests include: theoretical and empirical software engineering, model driven software development, knowledge engineering, ontology modeling, Semantic web, etc.