

Natural Language Semantic to Formal Language Semantic

Om Prakash Singh¹, Dr. Santosh Pawar²

¹RS – Computer Science and Engineering, Dr. A P J Abdul Kalam University, Indore, MP

²Director Research (PhD Cell), Dr. A P J Abdul Kalam University, Indore, MP

Abstract: The Natural language has fundamental role in human communication and even in human way of thinking and behavior regarding the world. With the advancement in Natural Language Processing, it is very essential to study it in all the respects especially semantic at different level. The semantic is a basis of both tasks Natural Language Understanding and Natural Language Generation in Natural Language Processing. The ambiguity and vagueness increases the semantic complexity in semantic theory. These two ambiguities and vagueness are socially acceptable while it is highly overlooked academically during the human interactions. These two are consider as distinct and varieties of interpretive uncertainty that has significant meaning in philosophy of language. The Ambiguity deals with uncertainty about levels of representation with different structural characteristics and vagueness with uncertainty about true meanings of particular items. The formal semantics is the study of using the formal logic to analyze the meaning that human beings use in everyday communication.

This work presents the review of different semantics and their logics take place in natural language processing with computability of natural language semantics and reflexion on computability of natural language semantic within ambiguous and vagueness situation. It is a computational analysis of currently used logical models and algorithm in natural language semantics with different semantics that will make the foundation for new semantic models and logics.

Keywords: Semantic Theory, Natural Language Semantic, Formal Semantic, Distributional Semantic, Situation Semantic and Cognitive semantic, logical models

I. INTRODUCTION

It is hard to say that who human found before in either of these thoughts or languages. But the language may indeed influence the way that human think and the thought known as linguistic determinism. The natural language is one of the most complex structure human have met with. It plays a very basic role in human way of thinking, communication and regarding the world. With the advancement in Natural Language Processing, it is very essential to study it in all the respects especially semantic at different level. There must be much more works done in understanding its structure especially phonetic and syntactic but there is less efforts in understanding its semantics.

The goal of Natural Language Processing (NLP) is to be designing and developing algorithm to allow computer or other smart devices to understand natural language in order to perform the variety of tasks. The NLP has basically two tasks – first Natural Language Understanding (NLU) and second Natural Language Generation (NLG). But there are different levels of tasks from text or speech processing to semantic and pragmatics interpretation and discourse processing. In NLP, semantic plays an important role and one of the fundamental tasks for many NLP applications and its related areas.

The Turing Test for artificial intelligence [**Turing, 1950**], a human interrogator needs to decide which of his two interlocutors is a man and which is a machine with the help of question answering session with two terminals. The Turing Test presupposes automated natural language analysis and generation. These natural language processing components of the Turing Test are used in computer programs for many interested area like question-answering, translation and sentiment analysis etc. But the core is semantic in all.

If we define the semantics of a sentence or sequence of sentences (s) as the mapping to a representation called formal formula (f(s)) that is used for processing task in a machine than there are various ideas about semantics storming the mind?

First, the semantics describes what the (s) speaks about. This represents the meaning by using the word vectors. As we know word vectors deals with referential like nouns, adjectives, verbs, etc and not the grammatical words. Here the



semantics represent what (s) speaks about. This is mathematically computable. If we have n words, than we found the word meaning in the form of n-dimensional array.

Second, the semantics describes in logical nature. Here the computational semantics is the mapping of sentence to logical formula. This is carried out in the very compositional manner. **Frege's Principle** states that "the meaning of a compound expression is a function of the meaning of its components that is syntactic structure". This concept imposes some conditions on the syntax, semantics and interrelationship in both within a language.

This present work focuses on the compositional and logical form of semantics and analysis of their computability in with and without ambiguity and vagueness. Only computation analysis can make the foundation for new formal logic for natural language semantics in NLP.

II. NATURAL AND FORMAL LANGUAGE

The natural language can be thought as a mutually understandable communication system used by communicators. It may be highly ambiguity and vagueness. A formal language as a type of programming language is a set of string over an alphabet. The string or words are derived from production rules; the well known formal language is programming language that has been design to express the computations.

Here our purpose is derive firstly from use of formal languages in the study of the semantics of the natural languages where the formal languages are used to provide the natural language expressions with their assumed meanings and this meaning (formal language semantics) are than used as models of natural language semantics. Here, while the logical form and grammatical form of an expression are distinct but we use the formal language model for a natural language. In real world, a natural language differs from a formal language because formal language used in mathematics or logic or programming. Even more natural language is transparent in use but formal language is almost abstract. Therefore, we can say that natural language is natural and formal language is formal but in computing, formal language semantics is used for semantics of natural language.

[**Pinker, 1994**] claim that there is specific cognitive function for language known as language module in the mind. Thus cognitive science also considers the computability of languages that involves language analysis and production. [**Langacker, 2008**] also support this and claim for computational language.

In linguistics, many researchers think that sentences cannot have any meaning without a context. This context deals with both linguist and extra-linguistic information. Hence the computational function or algorithm has input with context. If we consider the linguistic context, than we have semantic and are able to produce discourse structure. But in the case of without linguistic context, it is not related to semantic but it is an aspect of pragmatics.

[**Corblin, 2013**] argued that if someone is given a few sentences on a sheet of paper without any further information, he starts imaging situations, may infer other statements from what he reads,.....and such thoughts are the semantics of the sentence.

[**Montague (1974)**], as the linguistic tradition initiator lacks some coherence regarding computability. He gives an algorithm for parsing of sentences and computing their meaning as a logical formula and defined the semantics as the set of possible words in which the semantic formula is true.

We argue here in support of computational philosophy that models for higher order are not as simple as assumed sometimes as in [**Montague 1974**] and they do not quite match the formulas. This is the situation of completeness fails.

Again we have support against alternative relevant interpretation of formulas. This involves dialogical interpretation, game theory semantics, and structure of normal proof and set of consequences of formula as in intuitionistic logic.

And finally we have, interpreting the formula is no longer related to linguistic. If there is a formula, than interpreting may carry out in any way that decide whether it is true or false in a model.

III. NATURAL AND FORMAL LANGUAGE SEMANTICS

For any natural language, the formal language semantics is the study that makes use of tools from formal logic to analyze the meaning of the language that human being use in communication.



Our study about computability of natural language semantics begins from very ancient. It was found as per “**the Complete Works of Aristotle**” [PUP, 1991] that meaning of natural language is clearly visible in ancient Greece. There are much more evidences that clarifying the meaning of many words like entity, truth, sine, knowledge, morality, beauty, possibility and many more.

But the first mathematical theory about meaning representation was formulated by [Gottob Frege, 1960]. This theory is based on logic base approach of meaning representation that includes logic theories of formal language, semantic theories of natural language and formal semantics. This theory was consider as incomplete because there is only formally explanation of reference component out of considered two components as reference and sense. All logic based subsequent theories had inherited this incompleteness.

[Gemma Boleda and Aureli Herbelot, 2016] describe that this incompleteness did not cause any problem in the study of formal language but root of many difficulties in interpreting the natural language. The non-consider component ‘sense’ in Frege’s theory make the abstract abject without any descriptive content that becomes domain of many semantic models. This makes the semantic model hard to recognized subtle difference of meaning and barrier about various reasoning methods like reasoning by analogy adapted by human practice.

[Distributional Semantics, 2016] is the next theory with overcome on all such lacks in formal semantics. It is a vector based approach for meaning representation. But it is difficult to capture the meaning of function words. This causes the two major considerable demerits, first as difficulties in characterization of compositionality of natural language and second one as a reasoning method like deduction. All after it, this semantic model uses statistics of surrounding words in corpora to represent the meaning where corpora are a collection of language instances produced by human. This corpora is updated constantly with occurrence of the instances time to time as human practices. This continuous change in corpora makes the continuous change of word meaning. This outcome like continuous change in meaning is contradicts the philosophy of language that meaning of many words are regarded to be invariant. It is another thing that human uses the words in what sense. For instant we consider an example to understand this phenomenon= the word “earth” is unique in real world, but as per this distributional semantic model the meaning of “earth” would be constantly changed as more words are added with various sentences in corpora. This makes it unsuitable for meaning theory of natural language.

The semantic named as Lexical semantics used frequently in NLP. This is used to clarifies the meaning of words and phrases from words and phrases and produces dictionaries like WordNet [Dan Jurafsky & James Martin, 2006]. As we know a dictionary has various kinds of useful information about the words and its language. But a dictionary could not provide an independent meaning representation for all those words or phrases it defines. Therefore, a computer hardly accepts a language by only looking through the dictionary of that language. This makes the Lexical semantic as supplementary resources for other semantic representation.

The cognitive semantics [Vyvyan Evans & Melanie Green, 2006] works on an assumption that language is a part of human cognitive ability. Thus it describes the world as people conceive it. It agreed with language as a tool for organizing, processing and conveying the information. This assumption makes a solid foundation for modern sciences for linguistic and NLP practices. All about it, the cognitive semantics is not a unified formal theory.

Situation Semantic becomes in existence with “Scenes and other semantics” [Jon Barwise, 1981] and “Situation and Attitudes” [Barwise and Perry, 1983]. It was an alternative to possible world semantics. Here the linguistic expression is evaluated partially never be complete world. The situations are structured entities that hold the relations and individuals standing in those relations [Ginzburg & Sag, 2000] while [Von Fintel & Gillies, 2011] describes the situation as particular.

Problem with ambiguity and vagueness

The ambiguity and vagueness are two different essence of interpretive uncertainty while both are discussed together. There two varieties of uncertainty have their separate essential features and significance for semantic theory and philosophy of the language. Ambiguity deals with uncertainty about mapping between levels of representation with different structural characteristics. While vagueness is uncertainty about actual meaning of particular word or sentence. [Gillon, 2004] urged that it is more important to identify type of interpretive uncertainty.

In NLP from phonemes to discourses, ambiguity is appearing with utterance chunks. It is characterized by a string with more than one meaning. It is an intrinsic characteristic in human communication. It is particularly challenging in natural language understanding. Actually ambiguity is a major part of cognitive science that does not have a well defined solution. [Bolinger (1967, Siegel (1976), Cinque (1993))] all are agree with importance of ambiguity in the



development of semantic theory. Here, the ambiguity provides the crucial data for building and evaluating theories of lexical representation and semantic composition.

The semantic associated with ambiguity is typically related to the interpretation of the sentences. The ambiguity is evaluated from structural, lexical or compositional aspect throughout the architecture of a theory of the syntax-semantics interface.

The vagueness appears in all grammatical categories and also across lexical structure. It is pervasive in natural language therefore it is fundamentally important for semantic theory. The vagueness is multifaceted phenomenon like degree and combinatory vagueness, semantic and pragmatic vagueness.

Computational Analysis

[Dirk Geeraerts & Hubert Cuyckens, 2007] describe the primary function of natural language in communication as exchanging the information between the users of this language. It stands that the information conveyed by a natural language expression is the major part of meaning of this expression. The situation, distributive and cognitive semantics are also the part of this information. Therefore, the first and most important thing is to understand the information before understanding the meaning of natural language.

[R J McElice, 2004] is using the entropies of probability distributions to provide the quantitative characterization of the information in his information theory. Therefore our general aim is to find out how semantics construct its objects. The object construction is a complex process that involves empirical elements. There are no any brute semantic facts for semantics that come, describe and explain them. Truly there are some empirical elements. The semantic is no exception. This is special in this sense that different semantic theories conceptualize their objects and meaning in different way. Therefore, the meaning concept of one theory cannot reconstruct within another theory and vice versa.

The computational analysis is carried out in three steps:

- a. Mapping of sentences to logical formula
- b. Compute the syntactic analysis
- c. Incorporation of lexical terms

Mapping of sentences to logical formula

The mapping of considered sentences to logical formula may carry by different methods. We shall like to perform it with categorical grammar like Lambek-grammar [Moote & Retore, 2012]. The Lambek-grammar is obtained from a lexicalized formalism- Lambek calculus and a lexicon. This lexicon assigns a fine set of types to each word used in lexical structure.

If X and Y are two categories than $X \setminus Y$ and Y / X both are categories where “\” and “/” are two binary operators. Here, $X \setminus Y$ is a category X to its left that form a category Y and similarly Here Y / X is a category Y to its right that form a category X. If the sequences of words during natural language processing are $w = w_1, w_2 \dots w_n$ of types $C = c_1, c_2 \dots c_n$, than lexicon $\text{lex}(w)$ provides a finite set of categories whenever –

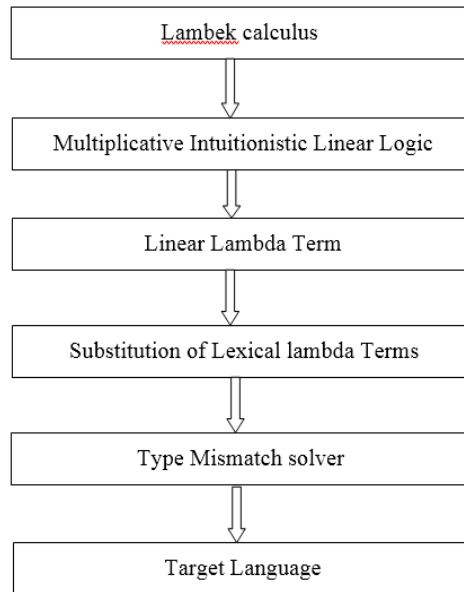
$$\forall i \exists c_i \in \text{lex}(w_i) c_1, c_2 \dots c_n \vdash C$$

Compute the syntactic analysis

Now time to perform categorical derivations from derivations in Multiplicative Intuitionistic Linear Logic to obtained the Linear Lambda Terms. This categorical derivation is proper subset of derivations in Multiplicative Intuitionistic Linear Logic. It is actually mapping from syntactic categories to semantic or logic type. Actually the typed lambda calculus is enough to express higher order logic.

Incorporation of lexical terms

There is a semantic lexicon also that maps any word of type in syntactic formula to a typed lambda term. Thus we have corresponding semantic type for each and every syntactic type. The miracle takes place, due to semantic lexicon with syntactic lexicon, the lambda terms are same on both output one from syntactic lexicon and another from semantic lexicon. These three steps are making the ground for an algorithm that show that computability of natural language semantic within ambiguous and vagueness situation. This is represented in a work flow diagram as follows:



IV. CONCLUSION

The computational semantics is fundamental of natural language semantic in NLP. It proves true percentage of natural language semantic and help in development of appropriate logic. This is because the only the formal logic capable to perform natural language processing in a very accurate way. This supports the philosophy that the grammar size is at least important as sentence length for the empirical complexity of parsing algorithm [Gomez-Rodriguez, 2006]. We have a ground believe that both the complexity of computing the semantics and the semantic representation are most important factors that the Turing test is out of approach presently.

REFERENCES

- [1]. P. W. Foltz, W. Kintsch, and T. K. Landauer, "The measurement of textual coherence with latent semantic analysis," *Discourse Processes*, vol. 25, no. 2-3, pp. 285–307, 1998.
- [2]. T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of Proceedings of the International ACM SIGIR Conference (SIGIR '99)*, pp. 50–57, 1999.
- [3]. T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, vol. 284, no. 5, pp. 34–43, 2001.
- [4]. Dekker, P., 2004, "Cases, Adverbs, Situations, and Events," *Context Dependence in the Analysis of Linguistic Meaning*, Hans Kamp and Barbara Partee (eds.), Amsterdam: Elsevier, 383–404.
- [5]. Raffman, Diana 2005. *How to understand contextualism about vagueness*. Reply to Stanley. *Analysis* 65, 244–248.
- [6]. Y. Li, D. McLean, Z. A. Bandar, J. D. O'Shea, and K. Crockett, "Sentence similarity based on semantic nets and corpus statistics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 8, pp. 1138–1150, 2006.
- [7]. A. Formica, "Ontology-based concept similarity in Formal Concept Analysis," *Information Sciences*, vol. 176, no. 18, pp. 2624–2641, 2006.
- [8]. Kennedy, Christopher 2007. Vagueness & grammar. The semantic of relative & absolute gradable adjective. *Linguistics & Philosophy* 30, 1–45.
- [9]. J. C. Valle-Lisboa and E. Mizraji, "The uncovering of hidden structures by Latent Semantic Analysis," *Information Sciences*, vol. 177, no. 19, pp. 4122–4147, 2007.
- [10]. M. Ruiz-Casado, E. Alfonseca, and P. Castells, "Automatising the learning of lexical patterns: an application to the enrichment of WordNet by extracting semantic relationships from Wikipedia," *Data and Knowledge Engineering*, vol. 61, no. 3, pp. 484–499, 2007.
- [11]. Dan Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing*. Prentice Hall, 2008.
- [12]. Barbara H. Partee. *Compositionality in Formal Semantics: Selected Papers by Barbara H. Partee*. John Wiley & Sons, 2008.
- [13]. A. Islam and D. Inkpen, "Semantic text similarity using corpus-based word similarity and string similarity," *ACM Transactions on Knowledge Discovery from Data*, vol. 2, no. 2, article 10, 2008.
- [14]. Roberto Navigli. Word sense disambiguation: a survey. *ACM Computing Surveys*, Vol. 41, No. 2, Article 10, 2009.
- [15]. K. Zhou, X. Gui-Rong, Q. Yang, and Y. Yu, "Learning with positive and unlabeled examples using topic-sensitive PLSA," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 1, pp. 46–58, 2010.
- [16]. P. Moreda, H. Llorens, E. Saquete, and M. Palomar, "Combining semantic information in question answering systems," *Information Processing and Management*, vol. 47, no. 6, pp. 870–885, 2011.
- [17]. V. Janev and S. Vraneš, "Applicability assessment of Semantic Web technologies," *Information Processing and Management*, vol. 47, no. 4, pp. 507–517, 2011.
- [18]. Schaffer, J. & Szabo, Z., 2014, "Epistemic Comparativism: a Contextualist Semantics for Knowledge Ascriptions," *Philosophical Studies*, 168: 491–543.
- [19]. I. Beltagy, Stephen Roller, Pengxiang Cheng, Katrin Erk, and Raymond J. Mooney. Representing meaning with a combination of logical and distributional models. *Computational Linguistics* December 2016, Vol. 42, No. 4: 763–808., 2016.
- [20]. Gemma Boleda and Aurelie Herbelot. Formal distributional semantics: Introduction to the special issue. *Computational Linguistics* December 2016, Vol. 42, No. 4: 619–635., 2016.