



ELS-JISH

ELS Journal on Interdisciplinary Studies on Humanities

Volume 2 Issue 1, 2019

ISSN (print) : 2621-0843

ISSN (online) : 2621-0835

Homepage : <http://journal.unhas.ac.id/index.php/jish>

Semantics in Natural Language Processing and Language Teaching

Ounu Sukaton¹

ounu.zakiy@machung.ac.id

Abstract

Semantics is one of the key elements in the study of language. This article's main objective is to describe and evaluate how semantic theories are implemented in the field of Natural Language Processing as well as Language Teaching. There are two branches of semantic theories that this article focuses on: Formal Semantics and Natural Semantic Metalanguage. The strengths and weaknesses of the theories mentioned are discussed. The most versatile theory is suggested along with future improvements.

Keywords: *Semantics, Language Teaching, Language Processing..*

How to cite: Sukaton O. (2019). Semantics in Natural Language Processing and Language Teaching. *ELS Journal on Interdisciplinary Studies in Humanities*, 2 (1), 58-65.

1. Introduction

In general linguistics, semantic analysis refers to analyzing the meanings of words, fixed expressions, whole sentences, and utterances in context” (Goddard & Schalley, 2010, p. 94). In order to achieve its goals, the analysis of semantics often relies on the use of another language or metalanguage (Goddard, 2011, p. 12). There are many approaches in the study of semantics and many of them offer their representational system or metalanguage in order to explain lexical meanings. Formal semantics is one of the earliest theory proposed by Richard Montague (1973; 1974) who used logical approaches to understand meanings. He applied mathematical based logic to analyse English utterances such as ‘John wishes to catch a fish and eat it’ and ‘the temperature is ninety but it is rising” (1973, p.222). Another example of a formal semantic approach is Discourse Representation Theory (DRT) which was developed in the early 1980’s by Kamp (1981). DRT aims “to capture the semantics of discourses or texts, that is, coherent sequences of sentences or utterances, as opposed to isolated sentences or utterances” (Goddard & Schalley, 2010, p. 97). DRT is also heavily based on mathematical based logic similar to what Montague (1973; 1974) proposed. In contrast, Anna Wierzbicka introduced semantic primitives (1992, 1997, 1999, 2014) as a metalanguage which are used to explicate semantic meanings. Wierzbicka emphasises cognitive based approaches in her work in order to account for the unquantifiable aspects, such as cultural concepts, embedded in lexical meanings.

¹Machung University, Indonesia

This essay will discuss the application of semantics theories in two fields which are Natural Language Processing (NLP) and language teaching. This essay aims to find out how semantic theories are implemented in both fields. Brief summary of semantics application in NLP and language teaching will be discussed in section 2. This essay will also assess the strengths and weaknesses of each theory and proposes the best theory for NLP and language teaching. Section 3 will provide detail arguments on how a certain theory is more practical than others. Finally, a conclusion will be provided in section 4.

2. The application of semantics

In this section, we will discuss the use of semantic theories in two fields which are NLP and language teaching. NLP is concerned with how humans communicate with one another through the use of language. This includes phonetics and phonology, morphosyntax, semantics, and pragmatics (see Figure 1). All of these information are then transferred to computers so that human and computers, also known as artificial intelligence (AI), can have proper interactions. Furthermore, Poesio (2000) said that “[t]he ultimate goal, for humans as well as natural language-processing (NLP) systems, is to understand the utterance—which, depending on the circumstances, may mean incorporating information provided by the utterance into one’s own knowledge base or, more in general performing some action in response to it. ‘Understanding’ an utterance is a complex process, that depends on the results of parsing, as well as on lexical information, context, and common sense reasoning. .” (p. 93). The theories of semantics are also prevalent in the field of language teaching. Lexical meaning is one of the most crucial foundation in learning a language. Therefore, language instructors will always benefit from having sufficient semantic knowledge.

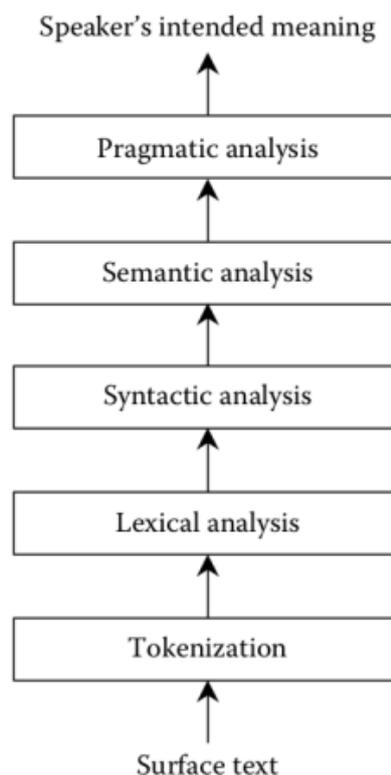


Figure 1. The stages of analysis in processing natural language (Dale, 2010)

2.1 Natural language processing (NLP)

Goddard and Schalley (2010) proposed some options of semantic approaches that will enable NLP to advance further. The use of semantic analysis in NLP might include machine translations, translation aids, data-mining, and text summarisation. Goddard and Schalley also addressed some problems in lexical semantics such as polysemy and homonymy. Lexical ambiguity will be a challenge for programmer to transfer such knowledge to AI because precise distinction is needed to tell one meaning to the others (e.g. polysemy in the word 'see' or 'hot'). The options that Goddard and Schalley proposed are logical approaches (e.g. Geurts & Beaver, 2008), DRT (e.g. Kamp, 1981), Pustejovsky's Generative Lexicon (e.g. Pustejovsky, 1995), Natural Semantic Metalanguage (NSM) (e.g. Goddard & Wierzbicka, 2004), and object-oriented semantics (e.g. Schalley, 2004).

Goddard and Schalley also pointed out two potential problems in semantics. The first one is the sense relations and ontologies. There are two sense relations which are horizontal sense relation and vertical sense relation. The horizontal sense relation includes synonymy and opposites (e.g. antonymy, complementary, conversity, and reversivity). The vertical sense relation includes hyponymy and meronymy. However, sense relations is not to be confused with ontological relation which has been used by NLP researchers in developing AI. "Ontological semantics is a theory of meaning in natural language and an approach to natural language processing (NLP) which uses a constructed world model, or ontology, as the central resource for extracting and representing meaning of natural language texts, reasoning about knowledge derived from texts as well as generating natural language texts based on representations of their meaning" (Nirenburg & Raskin, 2004, p. 10). The second one is semantic roles. Some examples of semantic roles are agent (the one that performs an action), patient (the one that is affected by the action), and instrument (the tool(s) that an agent use to perform an action). There are more examples of semantic roles (e.g. path, experiencer) which complicate the role of a lexicon in a given sentence. Thus, AI must be able to distinguish each lexicon precisely because one lexicon might be a patient or an experiencer depending on its relation with other lexicons in a sentence.

To address these problems, Goddard and Schalley argued that NSM is excellent at performing a fine-grained lexical-semantic analysis. Goddard and Schalley provided explications of 'sad' and 'unhappy' in English and contrasting them with their Chinese counterpart 'bei' and 'chou' (Ye, 2001). NSM seems to be successful in capturing subtle differences in these expressions. However, programming NSM into AI has never been done before and it will require a lot of work before NSM can be used in programming languages. Formal semantics is still the main approach of transferring human semantic knowledge to AI although formal semantics is incapable of performing fine-grained analysis as NSM does.

2.2 Language teaching

In the field of language teaching, Goddard and Wierzbicka (2007) argued that 'semantic primes' should be taught as core vocabulary to second language learners. The reason is that semantic primes are simpler in meanings and can be combined to explain more difficult concepts. In this sense, semantic primes act as a bridge for learners to fully comprehend culture specific words such as 'privacy' or 'compassion' in English. Moreover, there are only 65 semantic primes (Goddard, 2011) which is a relatively small number of vocabulary. By introducing semantic primes to learners, Goddard and Wierzbicka hoped

that learners can use semantic primes to help them learn new languages. The syntactic frames governing NSM is also simple but powerful. Some of the examples provided in this study are:

Do: X does something

X does something to someone (patient)

Say: X says: “---“ (direct speech)

X says something to someone (addressee) (Goddard and Wierzbicka, 2007)

In contrast, some of the text books that Goddard and Wierzbicka used as samples in their study only introduced 30 to 40 primes in their first 500 vocabulary. Primes such as LIKE, MAYBE, and BECAUSE are not among the first 500 vocabulary although Goddard and Wierzbicka argued that semantic primes are highly frequent words. However, Goddard and Wierzbicka notion of semantic primes as high value and high frequency vocabulary needs to be studied further. In addition, language text books for beginners usually emphasise practicality of topics such as greetings, origins, and hobby rather than concepts such as causality and comparison. The latter concepts are seen to fit intermediate to advanced learners rather than beginners. One of the solution to this problem is to fit semantic primes within beginner topics and make it natural.

Another thing that should be noted is how useful semantic primes in daily conversations. There is no doubt that NSM is an excellent tool in explicating abstract concepts. However, in order to explain artefacts or concrete objects NSM is at a huge disadvantage compared to the use of visual teaching aids such as pictures or videos. The length of an artefact explication will make it difficult for students to follow let alone understanding it. Nevertheless, NSM is still a powerful tool to explain cultural keywords to culture insiders and outsiders alike. Goddard and Wierzbicka also pointed out that semantic primes are un-interesting from a cultural point of view. However, they argued that the used of semantic primes is essential to avoid the use of cultural-specific words to explain other cultural-specific words which may lead to poor precision.

3. Discussion

In the field of NLP, formal semantic approaches seem to be more appropriate. First, the language of programming such as python and JavaScript can accommodate the formulation of formal semantics. The use of mathematic based logic in this approach is in harmony with many if not all programming languages. Second, by using formal semantic approaches, programmers can manipulate their formula easily by changing small parts of their codes. In most cases, one formula can be used to process multiple instances. Therefore, programmers can create one core formula and manipulate it to fit in different calculations. Third, there are a lot of shared database that can be used to start a project. The extensive work of AI has generated a number of solid database such as WordNet (Fellbaum, 1998). However, formal semantic approaches also have some weaknesses. The formula used in formal semantics (e.g. DRT and logical approach) can only be used for 1 particular language. In order to use it on other languages which has different syntax rules, different formulas need to be implemented. Formal semantic approaches are also incapable of performing fine-grained analysis as NSM because they can't account for circularity or ambiguity for instance.

The use of NSM in NLP promised to overcome the obstacles that formal semantic approaches failed to address. From Goddard and Schalley (2010) study we can see that NSM can point out subtle differences between rivers and creeks (p. 110). NSM was also able to elegantly show how people from different cultures have different understanding of 'sadness' concept (Ye, 2001). However, it doesn't mean that NSM can be immediately used to replace formal semantics in NLP. The lack of mathematical representation in NSM has made it difficult for this approach to be implemented in NLP. NSM researchers need to work together with NLP researchers to enable the use of NSM concepts in programming languages.

According to Andrews (2016), NSM and formal semantics don't need to be seen as exclusive research areas which have different paths to take. NSM and formal semantics can benefit each other by filling the gap in each approach. Furthermore, NSM and formal semantics "both have a strong dependence on the same kind of data, intuitions of entailment (including anomaly of observed examples explained by conflicting entailments), although especially in NSM, these are supplemented by corpus data" (Andrews, 2016, p. 98). Although there are fundamental differences in how NSM and formal semantics construe their logic, Andrews (2016) argued that the algebraic principles in formal semantics and primes in NSM is somewhat similar in explaining relationships such as transitivity, symmetry, and reflexivity (see table 1).

Table 1. Relationships concepts comparison (adapted from Andrews, 2016, p. 108)

Formal semantics

Transitivity: For all x, y, z , if xRy and yRz , then xRz (e.g. richer than)

Reflexivity: For all x , xRx (e.g. identical to)

Irreflexivity: For all x , not xRx (e.g. richer than)

Symmetry: For all x, y , if xRy , then yRx (e.g. similar to)

Asymmetry: For all x, y , if xRy , then not yRx (e.g. richer than)

NSM

OTHER: Irreflexive, Symmetric, (not Transitive)

MORE: Transitive, Asymmetric (and therefore Irreflexive)

BEFORE: Transitive, Asymmetric (and therefore Irreflexive)

AFTER: Transitive, Asymmetric, (and therefore Irreflexive)

ABOVE: Transitive, Asymmetric, (and therefore Irreflexive)

BELOW: Transitive, Asymmetric, (and therefore Irreflexive)

In the field of language teaching, formal semantics will have a significant disadvantage simply because it is not user friendly. The notation used in formal semantics require a certain degree of mathematic skill for both teachers and students. The fact that mathematical notations are intimidating to students might not be favourable for teachers to implement the use of formal semantics in their classrooms. NSM, however, is more practical compared to formal semantics in this field. The semantic primes are composed of 65 words which is relatively small in number and do not need a lot of effort to memorise. The 65 words are even common among what beginners usually get in their classroom.

However, there are also some issues in using NSM in language teaching. First, NSM is not an effective tool to explain concrete objects. The time it takes for students to understand a 'kangaroo' by looking at a picture compared to reading an NSM explication of a 'kangaroo' is significantly different. In a way, NSM explication of a 'kangaroo' might be as intimidating as formal semantics notations. Second, although NSM grammar is claimed to be universal, beginners will have difficulties combining semantic primes to form a sentence in NSM. If the target language and NSM have significantly different syntax (e.g. NSM and Japanese), students should study both system which means they have to make extra efforts. This case is of course not applicable if the students' first language has similar syntax to NSM, but even university students have difficulties when they try to make explications without proper instructions and hints. Even some published articles used NSM in an incorrect manner (e.g. Murtisari, 2013).

Goddard and Wierzbicka (2014) proposed the use of minimal English as global lingua franca. In my opinion, minimal English can also be implemented in English language teaching settings. This is especially useful for beginners with limited vocabulary. The use of semantic primes and molecules can replace complicated words such as evidence, fact, or proof (p.10). Moreover, semantic primes have been translated to multiple languages such as Chinese, Russian, and Japanese to name a few. This means that minimal Chinese, minimal Russian, and minimal Japanese are also available for teachers. The universality of semantic primes is the best selling point of using NSM in language teaching.

4. Conclusion

The proposed use of NSM in both NLP and language teaching demonstrate that NSM is a versatile and useful tool in both fields. Compared to other formal semantic approaches in NLP, NSM promises better accuracy and precision in semantic analysis. In language teaching, NSM can be used to equip learners with powerful semantic primes to ease the process of learning languages. The universality of semantic primes and their syntax can be useful for students in learning new languages.

There are some improvements which need to be done regarding the use of NSM in NLP and language teaching. NSM researchers need to figure out how to program NSM into AI by either integrating its concept in existing programming languages or developing a new programming language specifically for NSM. Finding common grounds in NSM and formal semantics is a good starting point and further research will be needed to apply NSM in NLP. In language teaching, teachers and researchers need to work together to integrate NSM in language teaching. This can be done by including semantic primes in the early vocabulary introduced to students. NSM workshops and trainings for teachers can also be used as a solution to equip teachers with NSM knowledge. NSM as a cultural-free metalanguage is useful to explain cultural-specific words to culture outsiders. In English learning settings, teachers can use semantic primes to teach students minimal English which is a powerful application of semantic primes in day to day conversations.

Consequently, future studies of NSM in NLP should focus on how to apply NSM in AI programming. In language teaching, there are two steps that need to be taken for NSM researchers. First, NSM researchers need to study how to equip teachers with NSM knowledge as efficient as possible. This will affect teachers preference in using NSM as one of the main theoretical frames in their teaching routines. Second, teachers and NSM

researches need to work together to include semantic primes as core vocabulary in text books and make it interesting for students.

References

- Andrews, A. D. (2016). Reconciling NSM and formal semantics. *Australian Journal of Linguistics*, 36(1), 79-111.
- Dale, R. (2010). Classical approaches to natural language processing. In N. Indurkha & F. J. Damerau (Eds.), *Handbook of Natural Language Processing, Second Edition*: CRC Press.
- Fellbaum, C. (Ed.) (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
- Geurts, B., & Beaver, D. I. (2008). Discourse representation theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2008 ed.). Stanford, CA: Stanford University.
- Goddard, C. (2011). *Semantic analysis: A practical introduction* (2nd ed.). Oxford: Oxford University Press.
- Goddard, C., & Schalley, A. (2010). Semantic analysis. In N. Indurkha & F. J. Damerau (Eds.), *Handbook Of Natural Language Processing* (2nd ed.): CRC Press.
- Goddard, C., & Wierzbicka, A. (2004). Cultural scripts: What are they and what are they good for. *Intercultural pragmatics*, 1(2), 153-166.
- Goddard, C., & Wierzbicka, A. (2007). Semantic primes and cultural scripts in language teaching and intercultural communication. In F. Sharifian & G. Palmer (Eds.), *Applied cultural linguistics* (pp. 105-124). Amsterdam: John Benjamins.
- Goddard, C., & Wierzbicka, A. (2014). *Global English, Minimal English: Towards better intercultural communication*, Australian National University.
- Kamp, H. (1981). A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal Methods in the Study of Language* (pp. 277-322). Amsterdam, The Netherlands: Mathematical Centre.
- Montague, R. (1973). The Proper Treatment of Quantification in Ordinary English. In K. J. J. Hintikka, J.M.E.Moravcsik, & P. Suppes (Eds.), *Approaches to Natural Language* (Vol. 49, pp. 221-242): Springer, Dordrecht.
- Montague, R., & Thomason, R. H. (1974). *Formal Philosophy: Selected Papers of Richard Montague*: Yale University Press.
- Murtisari, E. T. (2013). Some traditional Javanese values in NSM: From God to social interaction. *International Journal of Indonesian Studies*, 1, 110125.
- Nirenburg, S., & Raskin, V. (2004). *Ontological Semantics*. Cambridge, MA: MIT Press.
- Poesio, M. (2000). Semantic analysis. In R. Dale, H. Moisl, & H. Somers (Eds.), *Handbook of Natural Language Processing* (Vol. 1, pp. 93-122). New York, NY: Marcel Dekker.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

- Schalley, A. (2004). *Cognitive Modeling and Verbal Smeantics. A Representational Framework Based on UML*. Berlin, Germany: Mouton de Gruyter.
- Wierzbicka, A. (1992). *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*: Oxford University Press.
- Wierzbicka, A. (1997). *Understanding Cultures through Their Key Words: English, Russian, Polish, German, and Japanese*: Oxford University Press.
- Wierzbicka, A. (1999). *Emotions across languages and cultures: Diversity and universals*: Cambridge University Press.
- Wierzbicka, A. (2014). *Imprisoned in English*. New York: Oxford University Press.
- Ye, Z. (2001). An inquiry into "sadness" in Chinese. In J. Harkins & A. Wierzbicka (Eds.), *Emotions in Crosslinguistic Perspective* (pp. 359-404). Berlin, Germany: Mouton de Gruyter.