

# A DECADE OF USING HYBRID INFERENCE SYSTEMS IN NLP (2005 – 2015): A SURVEY

Ali Almiman

Department of Computer Science, University of Manchester, Manchester, UK

## **ABSTRACT**

*In today's world of digital media, connecting millions of users, large amounts of information is being generated. These are potential mines of knowledge and could give deep insights about the trends of both social and scientific value. However, owing to the fact that most of this is highly unstructured, we cannot make any sense of it. Natural language processing (NLP) is a serious attempt in this direction to organise the textual matter which is in a human understandable form (natural language) in a meaningful and insightful way. In this, text entailment can be considered a key component in verifying or proving the correctness or efficiency of this organisation. This paper tries to make a survey of various text entailment methods proposed giving a comparative picture based on certain criteria like robustness and semantic precision.*

## **KEYWORDS**

*Natural Language Processing, Textual Entailment, Semantic inference, Textual inference, Textual hypothesis,*

## **1. INTRODUCTION**

In today's world of textual information everywhere, and days of high amounts of information being generated every day, in the form of emails, chats, discussion forums and comments on articles, it has become a herculean task to make sense of the text and categorise them meaningfully. And the text generated here is generated by human interactions and is in natural language.

### **1.1. THE PROBLEM OF INFERENCE IN NLP**

There has always been a barrier in communications between human beings and devices, ever since the advent of computers. We solved this by creating syntactically robust languages. However, computer still fails to understand human beings, who beyond using multiple languages (problem of translation), make sense of a same sentence in different way in different contexts. There is a semantic element that difficult to extract from a given text. This precisely is the problem of Natural Language Processing. However as noted by Liddy [1], this area of research being active and young has no single agreed-upon definition. This has drawn interest of various researchers in the areas of computer science, artificial intelligence and linguistics equally in deciding how to make a machine make sense of human *natural language*. Early attempts to handle this problem involved direct hand-coding the rules, which in a way was very tedious. Most of the modern ways use methods similar to the machine learning, which are highly statistical in nature.[2]

## 1.2. TEXTUAL ENTAILMENT.

In the process of Natural Language Processing, when organising texts, we make various hypotheses which one could infer from the text. This inference has a directional relation to the given text. It is essentially to determine if meaning of one sentence T entails the hypothesis H. The RTE Challenge guidelines[3] defines the textual entailment in the following way - ‘our applied notion of textual entailment is defined as a directional relationship between pairs of text expressions, denoted by T - the entailing “Text”, and H - the entailed “Hypothesis”. We say that T entails H if, typically, a human reading T would infer that H is most likely true’.

## 2. THE PROBLEM

Looking at the brief history of NLP[4], where since the advent of computers, we have been trying ways to communicate with the machine effectively, we see an interesting trajectory of events. Earlier a lot importance was given in extracting meaning from each word that a human utters or seeks a computer as much information as possible from each of the words. This meaning making or inference involved processing information at different levels ranging from syntactic, lexical, semantic, discourse/pragmatic. But only later have we as a community started realising that this processing of each word often is not needed, instead one must focus on extracting as much sense from the sentence as possible as a whole. Thus the problem of NLP can be understood as extracting as much meaning as possible from the given sentence.

ID	TEXT	HYPOTHESIS	TASK	VALUE
1586	<i>The Republic of Yemen is an Arab, Islamic and independent sovereign state whose integrity is inviolable, and no part of which may be ceded.</i>	<i>The national language of Yemen is Arabic.</i>	QA	True
1076	<i>Most Americans are familiar with the Food Guide Pyramid- but a lot of people don't understand how to use it and the government claims that the proof is that two out of three Americans are fat.</i>	<i>Two out of three Americans are fat.</i>	RC	True
1667	<i>Regan attended a ceremony in Washington to commemorate the landings in Normandy.</i>	<i>Washington is located in Normandy.</i>	IE	False
2016	<i>Google files for its long awaited IPO.</i>	<i>Google goes public.</i>	IR	True
2097	<i>The economy created 228,000 new jobs after a disappointing 112,000 in June.</i>	<i>The economy created 228,000 jobs after disappointing the 112,000 of June.</i>	MT	False
893	<i>The first settlements on the site of Jakarta were established at the mouth of the Ciliwung, perhaps as early as the 5th century AD.</i>	<i>The first settlements on the site of Jakarta were established as early as the 5th century AD.</i>	CD	True
1960	<i>Bush returned to the White House late Saturday while his running mate was off campaigning in the West.</i>	<i>Bush left the White House.</i>	PP	False
586	<i>The two suspects belong to the 30th Street gang, which became embroiled in one of the most notorious recent crimes in Mexico: a shootout at the Guadalajara airport in May, 1993, that killed Cardinal Juan Jesus Posadas Ocampo and six others.</i>	<i>Cardinal Juan Jesus Posadas Ocampo died in 1993.</i>	QA	True

Figure.1: Table showing various NLP tasks as seen as TE problem [3]

However, broadly from the early 21st century, focus has been semantic inference and the area of Recognising Textual Entailment popularly known as RTE is a child of this pursuit. A typical system followed for any Natural Language processing is shown below[4].

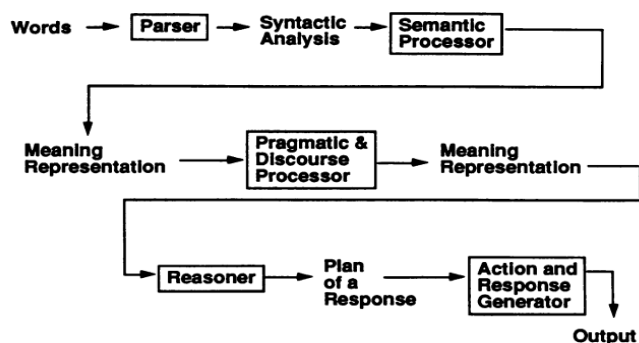


Figure.2: Pipeline view of a typical system of NLP [4]

Various tasks that are typically encountered include Question Answering (QA), Information Extraction (IE), (multi-document) summarization, and machine translation (MT) evaluation [3]. RTE is proposed as a generic task capturing various earlier existing tasks of QA, IE, MT, etc. RTE Challenge guidelines [3] specifically look for 7 such tasks namely Information Retrieval (IR), Comparable Documents (CD), Reading Comprehension (RC), Question Answering (QA), Information Extraction (IE), Machine Translation (MT), Paraphrase Acquisition (PP). These are proven with some strong empirical studies and arguments[5]–[7].

The efficacy of any given approach is understood by 2 key measures namely accuracy and a parameter defined as Confidence-Weighted score (*cws*). These are also used to compare various approaches. This is a measure to grade how well a system correctly entails the T-H pair. This is weighted based on the classification of the TE into positive (T entails H), negative (T does not entail H) and non-TE (T neither entails nor contradicts H) categories[3], [8]. A correct judgement gives higher confidence than the wrong judgements. It should work well in identifying and explaining both successful and failed T-H matches. *cws* is score between 0 and 1 and higher the score better is the system in meeting the requirement of RTE. A system with a good overall *cws*(on all the 7 tasks IR - - PP) is deemed better at TE.

### 3. SOME HYBRID METHODS TO SOLVE THE ENTAILMENT PROBLEM

Textual Entailment problem was solved collectively using platforms like RTE. Ever since, the launch of the RTE as a regular annual activity; it has gained momentum in both academic circles as well as corporate research houses, as this problem has a very high significance in linking human understanding with the high computational power of the machines (computers). Later, other data sets were provided to test the efficacy of various approaches. FraCas Suite has a new set of data based on the formal semantics.

We shall make a review of various studies in the context of the textual entailment. The key difference among various approaches that we discuss below lies in the structure and the steps that are followed to arrive at a conclusion between the dependency relations the text and the hypothesis pair. Broadly, it can be seen that they mainly employ a middle way between deep (semantic) analysis and shallow (word overlap like) analysis. Further, it can be seen that the semantic inferences can be of 3 kinds: propositional logic, first order predicate logic and others[4]. We will see in the following, some of these three approaches being used in isolation or in a mixture to accomplish the goal of RTE.

### 3.1. ATOMIC PROPOSITIONS [9]

One of the very early studies of the RTE age was this study, where the authors propose a textual entailment based on atomic propositions. The text and the hypothesis pairs are firstly parsed to extract atomic propositions. These atomic propositions are later compared using OTTER which uses the Knowledge rules. If for every atomic proposition in H has a counterpart in T, then the entailment holds, else the entailment breaks. This algorithm is shown in the picture below.

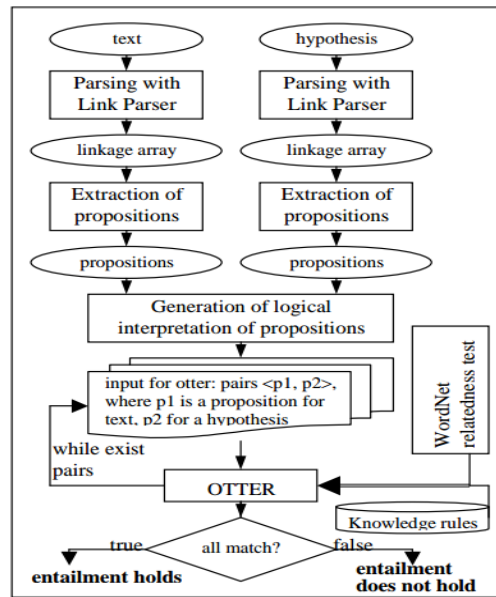


Figure.3: A system picture of the RTE using atomic propositions approach [9]

The performance is low with an accuracy of 0.5188 and overall *cwsof* 0.5067, which the authors attribute to lack of good knowledge rules database. Also, more work was needed to be done on proposition extraction and logical representation.

### 3.2. ABDUCTIVE THEOREM PROVING [10]

Here the text hypothesis pair is first converted to a logical representation using some already proven conventions [11]. This translates the textual inference problem to a logical inference problem. Then various abductive assumptions are analysed using a specially defined cost model. The assumption cost model needs to be minimised so that the entailment is more robust. Further the theorem prover (proving the Text Hypothesis relation) is allowed to learn good assumption costs iteratively. This among others compared gave better results. The Theorem prover which was trained separately for each class gave the best results with an accuracy of 0.57 and a *cws* of 0.651, significantly better than the simple atomic propositions approach.

### 3.3. COGEX [12]

COGEX was used earlier as a logic prover for a specific task of Question Answering (QA) [11]. However, here it is extended to the Textual entailment. Very much similar to the abductive theorem prover, the system here starts with converting text into logical form [10]–[12] along with the semantic relations. NLP axioms (used as linguistic rules) are then generated to connect the

concepts of T-H pairs. Some (310) knowledge axioms are also incorporated. Then WordNet is used to create lexical chains. The above processes are used to define the arguments of a predicate. This will make the logic prover, COGEX later not to miss any possible connections between the T-H. For, each case, a prover score is calculated iteratively till a refutation is reached, relaxing the arguments of the predicate step by step. This did very well with the CD with accuracy of 0.78 and *cws* of 0.822. But with other tasks the performance was not very satisfactory. The overall accuracy was 0.551 and *cws* was 0.56.

### 3.4. MODEL BUILDING [13]

As can be seen the above studies were not giving entirely satisfactory results. This approach gives a special focus on the semantic analysis. It performs deep semantic analysis using CGG Parser [14] to characterise the text and hypothesis pair individually. The figure.4 below shows an example. This is done along with several simple shallow semantic interpretations, or simple shallow word overlaps.

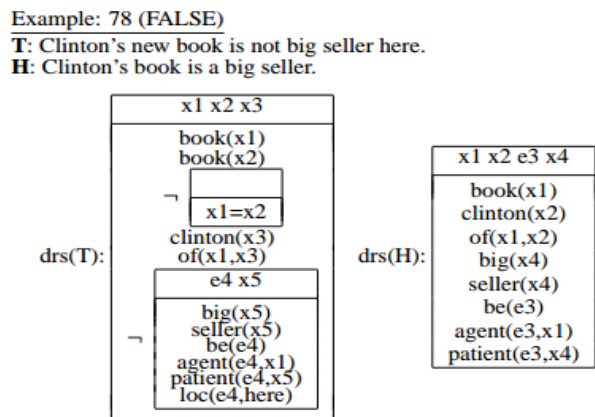


Figure.4: Semantic analysis of Text and Hypothesis pairs [13].

Further, model building approach if employed in contrast to the theorem proving ones, which will handle the negation well. Here theorem prover [10] and model builder are used in tandem to handle and account for both the negative and positive results effectively. This hybrid approach very clearly shows an advantage over the previous studies, with an accuracy of 0.61 and *cws* of 0.65. The same authors have later tried with logical inferences too, with no major improvements [15]

### 3.5. LCC'S GROUNDHOG SYSTEM[16]

The LCC's Groundhog system unlike the other systems is more robust and produced good results. It is based on the classifications and lexico-semantic information.

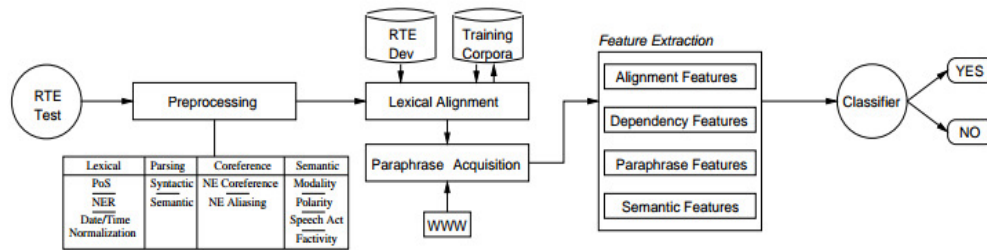


Figure.5: GroundHog System

In the GroundHog system, as shown in the above schematic, the T-H pair is first pre-processed and is annotated with four key sets of information (lexical, syntactic and semantic parse, co-reference, semantic). Then the system is trained with a large amount of test data acquired from World Wide Web (www). Later a lexical alignment probability is computed and paraphrase acquisition from top 500 documents of Google is done using a statistical approach similar to parallel corpora. Then the entailment is assessed by classifying using the four features as described in the figure.5 above (alignment, dependency, paraphrase and semantic). The entailment results were very positive with a good accuracy (>0.652) was reported, which increased with more training.

### 3.6. NATURAL LOGIC (LOGIC RUNNING OVER NATURAL LANGUAGE)[17]

As can be seen in the early examples studies running purely on lexical basis had little semantic precision [18], [19], but those based on the first order logic[9], [15] were less robust as we saw in the earlier studies. Here [17], we see a middle approach being proposed, where the logic is overlapped over the natural language, leading to something called Natural Logic. The system (known as *NatLog*) architecture is pretty similar to the groundhog [16] with three key stages namely pre-processing, alignment and entailment classification. This approach can be seen to have produced good precision of about 68.06%, which is a significant improvement. The hybrid, which the authors claim to be the middle path showed good accuracy and recall in both development (74.03%, 69.63%) and test stages (67.32%, 63.62%). These were tested on RTE3 test sets.[17] An improved study of Natural logic was later proposed [20] later, after the next approach bringing semantic and lexical-syntactical levels.

### 3.7. SEMANTIC INFERENCE AT LEXICAL-SYNTACTICAL LEVEL [21]

Semantic inference is something that is at the heart of the problem of textual entailment. However all the previous studies have had a mixed story in term of meeting semantic criteria and some nascent efforts can be seen to push this limit ahead slowly. Here, a new approach is proposed altogether, where the semantic inference is overlapped directly on the syntactical tree, so that the robustness of the shallow lexical-syntactical representations is taken advantage of. To do this, one transcends from text  $t$ , to the hypothesis  $h$ , in small incremental steps. That is the prover will try to generate  $h$  from  $t$  through a series of parse trees. This entire series of translations of  $t$  to  $h$  is guided by a set of entailment rules. The figure below give a picture of what a translation of the above kind would look like.

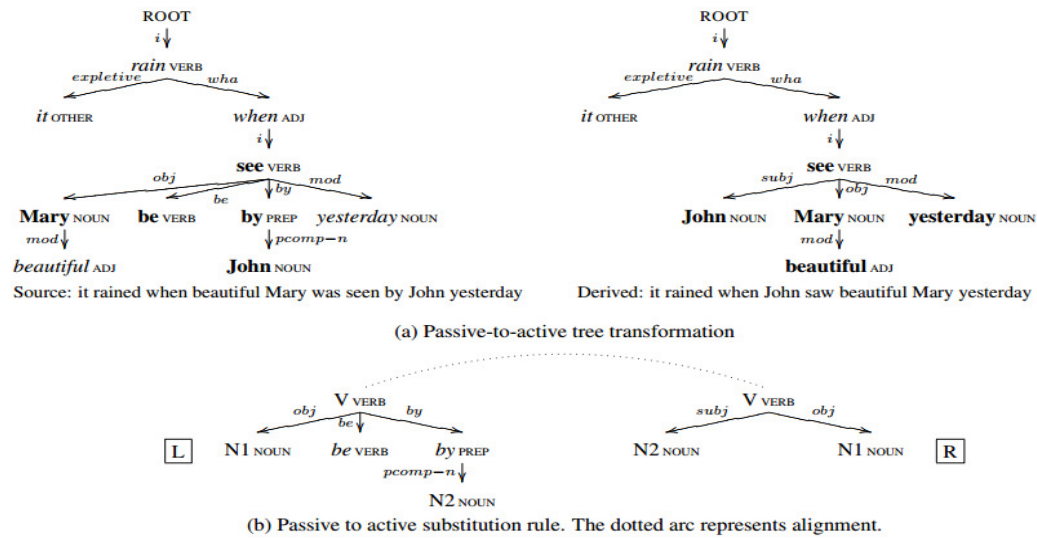


Figure.6: (a) *t* to *h* translation through intermediate parse trees  
(b) the entailment rule

A set of entailment rules like L-matching, R instantiation, Alignment copying. Other Annotation rules, generic-linguistic rules, lexical-syntactic rules are also manually created to guide the process. RTE data was not directly usable here due to the small sample size of the test set. However a different test with Base line and Proof shows a significant precision of 78.5%, which is quiet impressive. This also opens up a new framework for integrating semantic inferences along with lexical-syntactic level. This very clearly reveals the flaws in the approximate entailments allowed by RTE.

### 3.8. IMPROVED NATURAL LOGIC [20]

The previous study of Natural Logic [17] proposing a middle approach between the brittleness of semantic approaches and the robust approaches, which lack semantic precision. He the previous work is extended to incorporate semantic exclusion and implicativity. It is essentially similar to the previous study using intermediate trees [21] arriving at the hypothesis in small edits. Here, a statistical classifier is used at each edit to predict the lexical entailment and then feeds the relations upwards to higher level edits through the syntax trees as per the semantic properties of the intermediate nodes. In the process the entailment relation is obtained across the edits. Though this does not answer the entire problem of Natural language inference, it definitely works very well for the everyday patterns. The performance on RTE3 shows improvement when any existing RTE system is hybridized with the Natural logic system [20].

### 3.9. LOGICAL INFERENCE [22]

This is another approach that needs to be explored. Though has not proved its performance significantly better than others, this can be considered as an important approach in giving explanations to the entailment decisions. After the initial language processing using BLUE[23], [24], the entailment is verified using WordNet and DIRT paraphrases as knowledge sets. Following these entailment decisions, the explanations are provided. RTE4 is used as the test set of data. This opens up a new area of providing explanations to the entailment decisions in the TE problem.

### 3.10. Inference Rules[25]

Entailments rules were being used earlier studies to guide the incremental edits towards hypothesis[21]. Here we see how inference rules can be used to improve text entailment. This again falls into the group of TE approaches focussing on the semantic inferences. Using DIRT, the inference rules are acquired. DIRT algorithm includes determining the inference rules from Text. Then a few missing rules are identified, and some of them are acquired from the WordNet. A combination of DIRT and WordNet is applied on the RTE data. The figure.7 below shows a dependency tree skeleton is developed in the process.

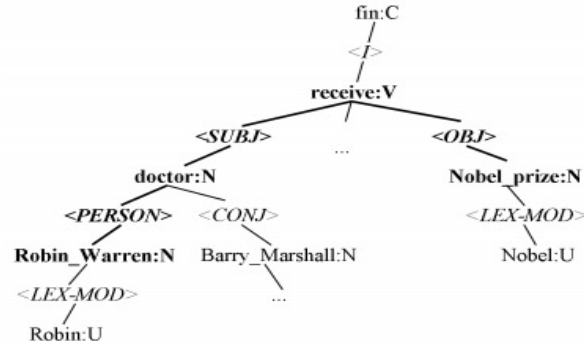


Figure.7: Dependency structure of text. **Tree Skeleton**

### 3.11. Distributional and Logical Semantics[17]

This involves mixing both the conventional semantics mapping to logical representations, but here the relational constants are given by distributional clustering at the level predicate-argument. This used FraCas and has shown impressive results beyond the existing studies. The figure.8 below shows the flow of steps in this approach. After the initial semantic analysis, the word typing (like identifying the types of the nouns) is performed. Certain logical forms are employed in this step. Then these are clustered using the distributional relations using statistics based on semantic relations. The last step is the semantic parsing using the relation clusters generated in the previous step. Experiments on FraCas suite shows a very impressive accuracy rate of 89% on single premise and 80% on multiple premises. Their study shows the first order logic approaches don't affect precision much. Also in terms of robustness of syntax, they admit their work being inferior to the work on Natural Logic [17].

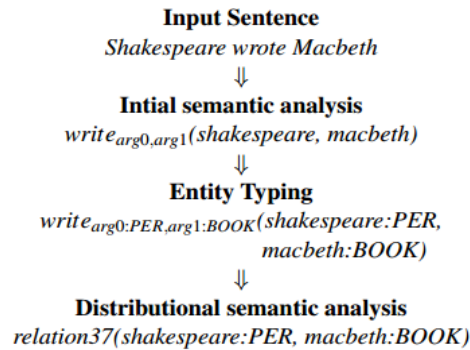


Figure.8: Flow of steps in Distributional Logical Semantics



### 3.12. Dependency based Compositional Semantics (DCS)[26]

This work is similar to the ones trying to transform the trees in the process of reaching hypothesis from text. DCS is a very good framework with simple structures as well as tight semantics. This approach merges the logical inference with the DCS giving promising results both on RTE and FraCas datasets.

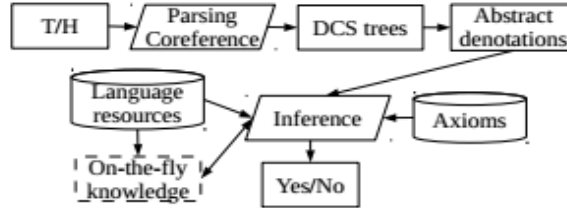


Figure.9: Figure showing the RTE implementation

Though the results are not good in comparison with the earlier studies on RTE2 and RTE3 datasets, the results are promising with RTE4 and RTE5 datasets. Further, on-the-fly implementations have also shown good accuracies.

### 3.13. Higher Order Logic [27]

This is the beginning of the higher order logic used in TE. [27]It is largely believed that higher order logic is not suitable for logical inferences; this study shows some promise in this direction. This opens up another approach to the text entailment bridging the parser and the semantic composition. The following picture below shows the CCG derivation of the semantic representation of a sample sentence.

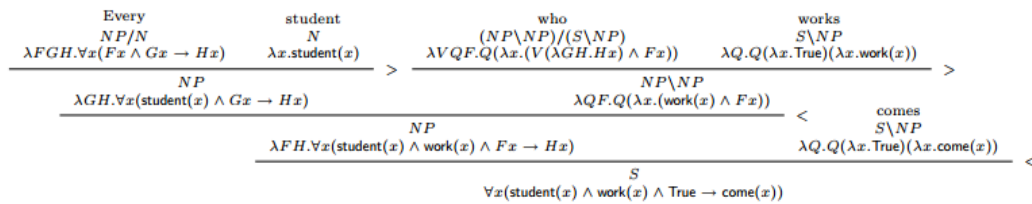


Figure.10 A CCG derivation of the semantic representation for a sentence[27]

Most of the results here very preliminary yet show promise in the fact that they out perform state-of-the-art first order logic system in the tests done on the FraCas datasets.

## 4. COMPARATIVE OBSERVATIONS

We can see that each of the above methods have different advantages and disadvantages from the logical point of view. Mixing up Semantic inferences with the lexical and syntactical inferences is something that the area is seems to struggle with. The Table in the appendix at the end shows a comparative overview of all the above mentioned approaches their performances quantitatively wherever is available. Also, most of these approaches still being tested using more and more data sets, as the knowledge base is also developing in parallel, as can be seen with the various versions of RTEs and the likes of FraCas.

## 5. FUTURE SCOPE

There has been an attempt to solve the Natural Language processing problem for many decades. However as can be seen in the above attempts to solve text entailment, a very concerted effort is being put a large community of researchers to make the machine process and understand natural language.

In the case of Text entailment, many algorithms have been proposed but, most of them are still not able to crack the problem or give a breakthrough. A revolution in the approach of solving this problem, (paradigm shift in the lines of historian and philosopher of Science Thomas Kuhn[28], is very badly needed. We can be more confident than ever, with the amount of computational prowess current technological age has.

With wider applications these days, with expanding reach of the technology, thanks to the smart phone revolution, the natural language processing is becoming a much sought after technology. Moreover, the latest studies have also started looking at languages other than English like this recent study in Arabic[29] Also, there are studies going on for speech recognition in the same flow of things. So, the approaches of higher order logical inference and semantic inferences which we ended with will be key thread to pursue in future directions.

There are two key aspects that could be realised to be hurdles to the TE problem. As shown in the above review, the first is the algorithm. The approach has slowly transformed from a mere syntactic one to more rigorous ones. We have stopped at the initial stages of higher order logic inference. The second key issue is the availability of sufficient data base of the resources like lexical resources and the more rigorous test data. Both these problems need to be handled together. The second issue however, could be easier, in the context of everyday interaction these days using digital media producing big data [30]–[32]. With more solid resources and potential test data, a crucial breakthrough is needed in the better algorithms, the second major problem. Attempts like Manning's Foundations of Statistical NLP [33] outline an endeavour in this very direction to chalk out the lookout for an effective algorithm. The latest handbook of semantic theory by Lappin[34] is another great guide book to start the pursuit of a breakthrough. Some latest approaches in this direction are the Bi-direction LSTM model and Inter-attention[35] and various other new models[36] addressing entailment using limited data[37] which is one of the key issues with the NLP problem.

## 6. CONCLUSIONS

This area of research as we can see is a very great meeting place for the people from Artificial intelligence, computer sciences, linguistics and philosophy. Also, practically, with the amount of data that is being generated and the information and insight these could carry in them, makes the problem even more relevant to today's requirements. Further, the communication with the machine is various interesting problem in the context of the entire human evolution and the scientific pursuit. The results of this research can be very productive in enabling intelligent machine and the robots to be more user-friendly and have a smooth transition with the natural world. Machines' understanding the human language as it is a going to break through in the entire history of science and technology.

## REFERENCES

- [1] E. D. Liddy, "Natural language processing," 2001.
- [2] Wikipedia contributors, "Natural language processing," Wikipedia, The Free Encyclopedia, 21-Jun-2016. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Natural\\_language\\_processing&oldid=726293558](https://en.wikipedia.org/w/index.php?title=Natural_language_processing&oldid=726293558). [Accessed: 24-Jun-2016].
- [3] I. Dagan, O. Glickman, and B. Magnini, "The PASCAL Recognising Textual Entailment Challenge," in *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment*, J. Quiñero-Candela, I. Dagan, B. Magnini, and F. d'Alché-Buc, Eds. Springer Berlin Heidelberg, 2006, pp. 177–190.
- [4] M. Bates, "Models of natural language understanding," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 92, no. 22, pp. 9977–9982, Oct. 1995.
- [5] C. Condoravdi, D. Crouch, V. de Paiva, R. Stolle, and D. G. Bobrow, "Entailment, Intensionality and Text Understanding," in *Proceedings of the HLT-NAACL 2003 Workshop on Text Meaning - Volume 9, 2003*, pp. 38–45.
- [6] L. Vanderwende and W. B. Dolan, "What Syntax Can Contribute in the Entailment Task," in *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Textual Entailment*, J. Quiñero-Candela, I. Dagan, B. Magnini, and F. d'Alché-Buc, Eds. Springer Berlin Heidelberg, 2006, pp. 205–216.
- [7] C. Monz and M. de Rijke, "Light-weight entailment checking for computational semantics," in *Proc. of the third workshop on inference in computational semantics (ICoS-3)*, 2001.
- [8] Wikipedia contributors, "Textual entailment," Wikipedia, The Free Encyclopedia, 26-Nov-2015. [Online]. Available: [https://en.wikipedia.org/w/index.php?title=Textual\\_entailment&oldid=692504148](https://en.wikipedia.org/w/index.php?title=Textual_entailment&oldid=692504148). [Accessed: 25-Jun-2016].
- [9] E. Akhmatova, "Textual entailment resolution via atomic propositions," in *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, 2005*, vol. 150.
- [10] R. Raina, A. Y. Ng, and C. D. Manning, "Robust textual inference via learning and abductive reasoning," in *AAAI, 2005*, pp. 1099–1105.
- [11] D. Moldovan, C. Clark, S. Harabagiu, and S. Maiorano, "COGEX: A Logic Prover for Question Answering," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1, Edmonton, Canada, 2003*, pp. 87–93.
- [12] A. Fowler, B. Hauser, D. Hodges, I. Niles, A. Novischi, and J. Stephan, "Applying COGEX to recognize textual entailment," in *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, 2005*, pp. 69–72.
- [13] J. Bos and K. Markert, "Recognising Textual Entailment with Logical Inference," in *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 2005*, pp. 628–635.
- [14] J. Bos, S. Clark, M. Steedman, J. R. Curran, and J. Hockenmaier, "Wide-coverage Semantic Representations from a CCG Parser," in *Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 2004*.
- [15] J. Bos and K. Markert, "When logical inference helps determining textual entailment (and when it doesn't)," in *Proceedings of the Second PASCAL RTE Challenge, 2006*, p. 26.
- [16] A. Hickl, J. Williams, J. Bensley, K. Roberts, B. Rink, and Y. Shi, "Recognizing textual entailment with LCC's GROUNDHOG system," in *Proceedings of the Second PASCAL Challenges Workshop, 2006*.
- [17] B. MacCartney and C. D. Manning, "Natural Logic for Textual Inference," in *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, Prague, Czech Republic, 2007*, pp. 193–200.
- [18] L. Romano, M. Kouylekov, I. Szpektor, I. Dagan, and A. Lavelli, "Investigating a Generic Paraphrase-Based Approach for Relation Extraction," in *EACL, 2006*.
- [19] V. Jijkoun, M. Rijke, and Others, "Recognizing textual entailment using lexical similarity," 2005.

- [20] B. MacCartney and C. D. Manning, "Modeling Semantic Containment and Exclusion in Natural Language Inference," in Proceedings of the 22Nd International Conference on Computational Linguistics - Volume 1, Manchester, United Kingdom, 2008, pp. 521–528.
- [21] R. Bar-Haim, I. Dagan, I. Grental, and E. Shnarch, "Semantic inference at the lexical-syntactic level," in Proceedings of the National Conference on Artificial Intelligence, 2007, vol. 22, p. 871.
- [22] P. Clark and P. Harrison, "Recognizing textual entailment with logical inference," in Text Analysis Conference (TAC 2008) Workshop-RTE-4 Track. National Institute of Standards and Technology (NIST), 2008.
- [23] P. Clark and P. Harrison, "Boeing's NLP System and the Challenges of Semantic Representation," in Proceedings of the 2008 Conference on Semantics in Text Processing, Venice, Italy, 2008, pp. 263–276.
- [24] P. Clark, C. Fellbaum, J. R. Hobbs, P. Harrison, W. R. Murray, and J. Thompson, "Augmenting WordNet for Deep Understanding of Text," in Proceedings of the 2008 Conference on Semantics in Text Processing, Venice, Italy, 2008, pp. 45–57.
- [25] G. Dinu and R. Wang, "Inference Rules and Their Application to Recognizing Textual Entailment," in Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Athens, Greece, 2009, pp. 211–219.
- [26] R. Tian, Y. Miyao, and T. Matsuzaki, "Logical Inference on Dependency-based Compositional Semantics," in ACL (1), 2014, pp. 79–89.
- [27] K. Mineshima, P. Martinez-Gómez, Y. Miyao, and D. Bekki, "Higher-order logical inference with compositional semantics," in Proceedings of EMNLP, 2015, pp. 2055–2061.
- [28] T. S. Kuhn, *The Structure of Scientific Revolutions*. Chicago: Third Edition, 1996.
- [29] F. AL-Khawaldeh and V. Samawi, "Lexical cohesion and entailment based segmentation for arabic text summarization (Iceas)," *World ComputSciInfTechnol J*, 2015.
- [30] M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile NetwAppl*, vol. 19, no. 2, pp. 171–209, Jan. 2014.
- [31] L. Manovich, *The Language of New Media*. MIT Press, 2001.
- [32] D. V. Shah, J. N. Cappella, and W. R. Neuman, "Big data, digital media, and computational social science possibilities and perils," *Ann. Am. Acad. Pol. Soc. Sci.*, vol. 659, no. 1, pp. 6–13, 2015.
- [33] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [34] S. Lappin and C. Fox, *The Handbook of Contemporary Semantic Theory*. Wiley, 2015.
- [35] Y. Liu, C. Sun, L. Lin, and X. Wang, "Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention," *arXiv [cs.CL]*, 30-May-2016.
- [36] B. Magnini, "Recognizing Textual Entailment: Models and Applications Ido Dagan<sup>1</sup>, Dan Roth<sup>2</sup>, Mark Sammons<sup>2</sup>, and Fabio Massimo Zanzotto<sup>3</sup> (1Bar-Ilan University, Israel, 2University of Illinois, Urbana, IL, and 3University of Rome 'Tor Vergata,' Italy) Morgan & Claypool (Synthesis Lectures on Human Language Technologies, edited by Graeme Hirst, volume 23), 2013, xx+ 200 pp; paperbound, ISBN 978-1-59829-834-5; e-book, ISBN 978-1-59829-835-2," *Comput. Linguist.*, 2015.
- [37] C. Shivade, P. Raghavan, and S. Patwardhan, "Addressing Limited Data for Textual Entailment Across Domains," *arXiv [cs.CL]*, 08-Jun-2016

## APPENDIX: TABLE

S. No	Approach	Key Claim/ task	overall cws (only for RTE)	accuracy/ precision (when mentioned)	FraCas score	Advantages/ Disadvantages	Authors	Year
1	Textual Entailment Resolution via Atomic Propositions	Describing a computer tool which can extract text entailments by comparing atomic propositions.	0.5067	0.5188	-	Simple. Inadequate Knowledge rule database for a robust Entailment.	Elena Akhmatova	2005
2	Robust Textual Inference Via Learning and Abductive Reasoning	Parsing sentences into logical-formula like representations. Then a minimum cost set of assumptions are realised using an abductive theorem prover. If one sentence follows the other low cost set of assumptions. This marries precision of logical reasoning with the robustness of machine learning.	0.651	0.57	-	More adaptive and flexible. Highly syntactical, Semantic aspects still weak.	RajatRaina, Andrew Y. Ng and D. Manning	2005
3	Applying COGEX to Recognize Textual Entailment	Transforming T-H pair into logic form representation with semantic relations. The system then generates axioms as linguistic rewriting rules and lexical chain axioms to connect T-H.	0.56	0.551	-	Has better semantic connectivity. Creating knowledge base could be tedious.	Abraham Fowler et al.	2005
4	Recognising Textual Entailment with Logical	Using Model building and machine learning	0.65	0.61	-	Better semantic inference and hybrid model	Johan Bos and Katja Markert	2005

	Inference					leading to more robust entailment.		
5	Recognising Textual Entailment using LCC's Ground Hog	Introducing a new system for recognizing textual entailment (known as GROUNDHOG), which utilizes a classification-based approach to combine lexico-semantic information derived from text processing applications with a large collection of paraphrases acquired automatically from the WWW.	-	>0.652	-	Machine learning based approach and high accuracies possible. Requires large amount of training examples.	Hicks, et al.	2006
6	When logical inference helps determining textual entailment (and when it doesn't)	Comparison of logical inference (shallow method) in its efficacy in text entailment.	-	0.616 (shallow) and 0.606 (both)	-	No significant improvement in results using the logical inference (except for a few tasks)	Johan Bos and Katja Markert	2006
7	Natural Logic for Textual Inference	Most approaches sacrifice semantic precision for robustness. But those based on first order logic and theorem-proving are highly brittle. This is a middle way	-	0.673	(accuracy) 59.56%	Tries to use both deep and shallow using Natural logic ( <i>NatLog</i> ), among the first to use FraCas	Bill MacCartney and Christopher Manning	2007
8	Semantic Inference at the Lexical-Syntactic Level	Classical approaches to semantic inference rely on complex logical representations. However,	-	78.5 (precision) RE (Relation Extraction) setting	-	Does not use standard RTE or FraCas setting. More precise and	Bar-Haim, Dagon, Greental, and Shnarch	2007

		<p>practical applications usually adopt shallower lexical or lexical-syntactic representations, but lack a principled inference framework. We propose a generic semantic inference framework that operates directly on syntactic trees. New trees are inferred by applying entailment rules, which provide a unified representation for varying types of inferences.</p>				<p>less approximate than the other studies. A new framework to incorporate semantic as well as lexical-syntactic levels</p>		
9	<p>Modeling Semantic Containment and Exclusion in Natural Language Inference</p>	<p>Using a new method to reduce the error incorporating semantic inference, by using a sequence of atomic edits between T-H</p>	-	<p>64.5 (accuracy RTE3)</p>	<p>70.49 (accuracy-FraCas)</p>	<p>Hybrid method. More reliable results than earlier NatLog. Using semantic containment, exclusion and implicativity explains many everyday patterns.</p>	<p>Bill MacCartney and Christopher Manning</p>	<p>2008</p>
10	<p>Recognizing Textual Entailment with Logical Inference</p>	<p>First the semantic interpretation of the sentence is performed and then check if the logic for the H is implied by some inference elaborated</p>	-	<p>~ 65% correctness</p>	-	<p>Provides explanations. Produces preliminary positive results to start with. Deductive style of reasoning.</p>	<p>Peter Clark and Phil Harrison</p>	<p>2008</p>

		version of T. The system also tries to produce explanations for the entailments, sometimes also erroneous.				Use of noisy knowledge	
11	Inference Rules and their Application to Recognizing Textual Entailment	Starting with automatically acquired collection and refine it and obtain more rules using hand crafted lexical resource. Then produce a dependency based structure representation from texts, with an aim to provide proper base for the inference rule application.	-	high precision (>55) across tasks	-	Flexible combinatorial approach. Tedious inference rules writing.	Georgia Dinu and Rui Wang 2009
12	Combined Distributional and Logical Semantics	Following formal semantics in mapping language to logical representations but differ in that the relational constants used are induced by offline distributional clustering at the level of predicate-argument structure.	-	-	89% (accuracy single premise) 80% (accuracy multiple premises)	Distributional logical Semantics. Advanced clustering techniques deployed for textual entailment	M Lewis and Mark Steedman 2013
13	Logical Inference on Dependency-based Compositional Semantics	Equipping the DCS framework with logical inference, by defining abstract denotations as an abstraction of the computing process of	-	accuracies of above 59%	79.5% (accuracy single premise) 80% (accuracy multiple premises)	DCS and logical inferences. On the fly implementations	Tian, Miyao, and Matsuza ki 2014



		denotations in original DCS. An inference engine is built to achieve inference on abstract denotations. Furthermore, we propose a way to generate on-the-fly knowledge in logical inference, by combining our framework with the idea of tree transformation.						
14	Higher-order logical inference with compositional semantics	Developing a bridge between parser and semantic composition. higher order logic used for logical inference	-	-	69% (accuracy)	Opens gates for higher order logic inferences.	Koji Mineshima, et al.	2015