

A Research on Discourse Access

Animesh R. Tayal, S.R. Tandan, M. A. Tayal

Abstract— We are living in the era of intelligent machines where everything is replacing by machine, interpretation of hidden meaning of text/speech becomes necessary to understand by machine. The last few years have seen an emergence of new perspectives in understanding the meaning of the given text. Researchers are working hard for this kind of complex problem. Discourse comes under this theme of understanding hidden information from the text. Syntactical and semantic analysis plays important role in discourse analysis. This paper focuses on the work done by various researchers in understanding discourse using various methods of machine learning and deep learning techniques.

Index terms: Discourse, Discourse Analysis

I. INTRODUCTION

Discourse is finding the hidden meaning from the piece of data we have. Discourse is defined by many researchers with different intent. Some more definitions are:

Discourse is “**Language above the Sentence or above the Clause**” (Stubbs, 1983).

Wikipedia depicts Discourse Analysis is a general term for a number of approaches to analyze semantically written, spoken, signed language as well as images also.

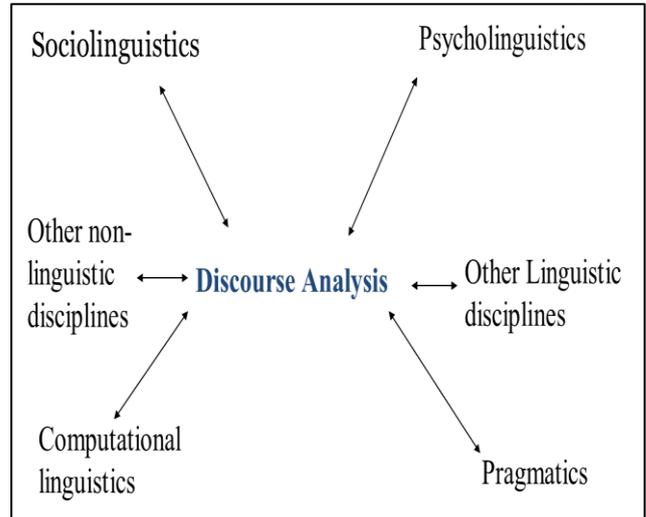
Discourse analysis is also concerned with language used in social platform, and in particular with interaction or communication between two persons. Discourse analysis is sometimes defined as the analysis of language 'Beyond The Sentence'. (Stubbs, 1983).

Discourse Analysis exists normally in the process of Natural Language Understanding. The words are analysed through morphological analysis and finds original word, then finds it appropriate tag using POS tagger. After having the syntax and semantic analysis discourse processing comes into the picture.

1.1 Influences on discourse analysis:

Following are the factors which influences Discourse analysis in broad sense. i.e. Sociolinguistics, psycholinguistics, computational linguistics, Pragmatics etc.

Figure 1. Factors which influences Discourse analysis



1.2 Discourse Analysis Levels:

We can analyse the different analysis levels of Discourse Analysis i.e. cohesion, coherence described below:

A. Cohesion Level (Textual Phenomenon)

- Reference (It, this, that, his, her, him...etc.)
- Ellipsis

Example:

- *The printer, I bought today doesn't work properly. It Surprised me.*
- *I know that person. Do you?*

It represent the grammatical relationship between sentences. It is essential to understand the exact meaning of sentences.

B. Coherence Level (Mental Phenomenon)

Example:

Her: It's the party time.

Him: Tomorrow having exam paper.

Her: Ok.

We can deduce that the above discussion with the assistance of straight action and by our contextual information that someone who is having exam paper cannot come for the party.

1.3 Approaches of Discourse Analysis:

Many attempts have been made to understand the hidden information from the text. Researchers have applied this concept on many fields/areas. So, broadly, Discourse Analysis can be applied on various areas/ field, represents in Fig. 1.2.

Revised Version Manuscript Received on 16 September, 2019.

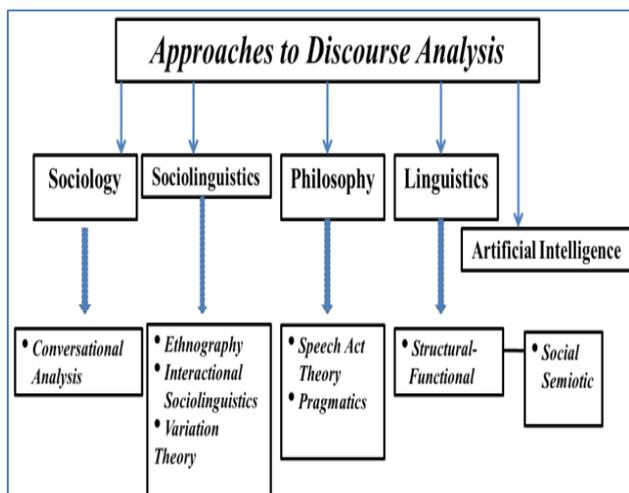
Animesh R. Tayal, Research Scholar, Department of computer science & Engineering, Dr. C.V.Raman University, Bilaspur (CG), India.
(Email: : jayamaniraja07@gmail.com)

S.R. Tandan, Associate professor, Department of computer science & Engineering, Dr. C.V.Raman University, Bilaspur (CG), India.

M. A. Tayal, Assistant Professor, Department of Information Technology, Shri Ramdeobaba College of Engineering and Management, Nagpur (MH), India.

Figure. 1.2: Approaches to Discourse Analysis

II. REVIEW OF LITERATURE & RESULTS



A. “Discourse Structure and Computation: Past, Present and Future”. (Bonnie Webber, 2012)

This work describes the discourse properties of text have extended been documented as serious to language technology, and over the past 40 years, our understanding of and capability to exploit the discourse possessions of text has grown up in several means. This paper demonstrates about many developments, technology, applications, support and the new challenges that each succeeding growth has elevated. (Bonnie Webber, 2012)

B. “CASCADE: Contextual Sarcasm Detection in Online Discussion Forums”, (et.al D. H., 2018)

This work focuses on automatic sarcasm recognition which is based on lexical, syntactic and semantic level investigation of text. Background, common sense knowledge are some of the ingredients of sarcastic sentence. This paper, author suggest a Contextual SarCAsm DETector (CASCADE), which accepts a mix method of content- and context-driven modelling. This was applied for online social media stuff. In the next phase this work extracts contextual information. CASCADE uses user embedding that encode stylometric and personality features of users as the sarcastic nature and form of expression can vary from person to person,. (et.al D. H., 2018)

C. “The Penn Discourse Treebank”, (et.al E. M., 2004)

This work focuses on discourse-level annotation project – the Penn Discourse Treebank (PDTB) – those goals to produce a large-scale corpus in which discourse connectives are glossed, along with their arguments. The PDTB is being constructed straight on top of the Penn Treebank and Propbank, thus associate the mining of valuable syntactic and semantic features and providing a richer substrate for the development and evaluation of practical algorithms. (et.al E. M., 2004)

D. “Lexical and Discourse Analysis of Online Chat Dialog”, (et.al E. N., 2007)

This work focuses on construction of a chat corpus, labelled with lexical (token part-of-speech labels), syntactic (post parse tree), and discourse (post classification)

information. This corpus can then be used to grow additional complex, statistical-based NLP presentations that accomplish tasks such as entity identification, author profiling and social network analysis. (et.al E. N., 2007)

E. “Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information”, (et.al F. B., 2018)

This work focuses on Discourse information which pays to a deeper thoughtful of the part of these communications to method social media data from a new viewpoint in discourse understanding. This overview primarily delivers the essential contextual to understand what context is from both the linguistic and computational linguistic perspectives, and then gives the most current context-based methods to NLP for social media. In this paper, author conclude with highlighting what they believe are the future directions in processing social media texts. (et.al F. B., 2018)

F. “Understanding spoken discourse”.(et al., Cornish 2014)

This work mainly emphases on spoken words, the current research on-going work is upon the Spoken discourse, located considerate of a speaker’s talkative senses, in which the another speaker’s expected and actual reactions are an integral part. The formation of discourse is thus a joint endeavour, joining the energetic association of all the members. The textual record on which this productive activity is based, in union with the invocation of a applicable context, is constituted not only by the verbal content of the utterances produced, but also by non-verbal signals.

G. “Discourse Analysis of User Forums in an Online Weight Loss Application”, (et.al N. J., 1992)

This work focuses on initial investigation of language and discourse patterns in forum posts by users who lose weight and keep it off versus users with fluctuating weight dynamics. The result demonstrates polarity of sentiments; reveal differences the kinds of posts and semantic cohesion of posts made by users differ sideways with their weight loss design. This work is the first attempt of discourse-level analysis of linguistic and weight loss dynamics. Empirical Analysis done, which includes:

- Asking Questions
- Sentiment of Posts
- Cohesion with Previous Posts (Lydia Manikonda et.al, 2014)

H. “Textual Analysis of Print Media Text (et.al M. T., 2016)

In this paper the text is analysed for different types as well as linguistic formal structure in the newspapers language is also analysed. This work focuses on mainly political leaders, national politics, policy makers and the decision taken by leading politicians.

This paper analysed the newspapers “The Nation” for the month of May and September. Following points are analysed:



1. The linguistic formal structure of media discourse. (et.al M. T., 2016)

2. The text kinds used in media discourse (open and close text)

I. "A Discourse Analysis of Software Documentation: Implications for the Profession" (et.al N. J., 1992)

This paper compares two best-selling books for word processing and spread sheet programs to determine the resemblances and modifications among prime and subordinate computer handbooks, and to account for the popularity of the secondary texts, to documentation provided by the producer. They created the revision by emerging experiential for investigative software documentation based on reasoning and oratorical values. Then applied the heuristic to the corporate documentation. When examined conferring to emotional aspects for operational documentation, the four sets of texts were similar they differed when analysed according to linguistic aspects. To support learning all the texts participate causes calculated. They use methods that inspire short- and long-term memory, appeal present work schemas, build new schemas, and create text processing easier. (et.al N. J., 1992)

J. "Annotating the discourse and dialogue structure of SMS message conversations" (et.al N. J., 1992)

In this paper author presented an outline for interpreting the discourse and discussion arrangement of SMS message discussions. The annotation specifications assimilate essentials of coherence-based discourse relations and communicative acts in conversational speech. Also present annotation experimentations that show reliable annotation can be achieved with this annotation framework. (al N. X., 2016)

K. "Discourse Annotation in the PDTB: The Next Generation", (et.al R. P., 2018)

In this paper author highlights from their work on enriching the Penn Discourse Treebank (PDTB), which has to be released to the community in fall 2018 as the PDTB-3. They have increased its coverage of discourse relations (from ←-40K to ←-53k), the majority in intra-sentential contexts. This work on these new relations has led them to augment and/or modify aspects of the annotation guidelines, including the sense hierarchy, and all changes have been propagated through the rest of the corpus. (et.al R. P., 2018)

L. "Adopting New IT by Learning in Groups: Results of Discourse Analysis", (et.al T. V., 2002)

This work was proposed to take place evolutionarily, containing pilots in diverse departments and appropriate feedback. After some duration the structure was positively accepted by one group of users but unsuccessful with alternative. This idea was invented by Beaufort Method. (et.al T. V., 2002)

M. "News clustering approach based on discourse text structure", (al T. M., 2015)

This paper presents a method to post-retrieval snippet gathering founded on design structures building on improved syntactic parse trees. Authors suggest a reduction technique that permits to build a reduced pattern assembly and create it

accessible. In spite of how information is distributed between sentences, this algorithm takes into account discourse information to create clustering outcomes. (al T. M., 2015)

N. "Discourse Relations and Conjoined VPs: Automated Sense Recognition" (Valentina Pyatkin, 2017)

This work focuses on senses. The Logic identification of discourse relations is associated to shallow discourse parsing. Discourse relations can befall together crosswise sentences (inter-sentential) and inside sentences (intra-sentential), and more than one discourse relation can grip among the similar units. This paper demonstrated a consecutive classification system for their multi-label sense classification. It uses newly current corpus of discourse-annotated intra-sentential adjoined verb phrases. The importance of each feature is assessed in the classification. (Valentina Pyatkin, 2017)

O. "An Analysis of Spoken Discourse between Two Native Speakers",

This paper first registers records and then evaluates a small extract from an English chat between two natural utterers. It goals to disclose that rules and features undoubtedly happen in people's daily dialogues. Though they are quick and passing, dialogues are more or less prepared and organized. For this reason, both language learners and teachers should lay stress on rules and meanings of verbal discourse. (Zhang, 2010)

P. "Toward Fast and Accurate Neural Discourse Segmentation", (Yizhong Wang et.al, 2018)

This paper proposes an end to end neural segmenter based on BiLSTMCRF framework. To advance its accuracy, they address the problem of data insufficiency by moving a word representation model that is trained on a huge corpus. They also suggest a limited self-attention mechanism in order to capture useful information within a neighbourhood. Experimentations on the RST-DT corpus show that this model is significantly earlier than preceding methods, while attaining new state of-the-art performance. (Yizhong Wang et.al, 2018)

Q. "Improving STEM Classroom Culture: Discourse Analysis" (et.al Y. V., 2013)

This paper recommends that discourse analysis may assist as a device to improve trainers understanding of their classrooms and to assist as an involvement mainly useful for junior faculty as they are beginning their teaching career. In addition to this, it defines the theoretic basis of discourse analysis, determines its application, effectiveness applicability in STEM classrooms, Qualitative Methodology applied in three dissimilar courses. (et.al Y. V., 2013)

IV. CONCLUSION

According to the reviewed work of various authors, it is detected that discourse investigation is important for human machine interaction, machine intelligence, question answering system, etc. this may further lead to content



generation from images. Proposed model will help to donate in examining text for its hidden sense.

REFERENCES

1. N. X. (2016). "Annotating the discourse and dialogue structure of SMS message conversations". *Proceedings of Association for Computational Linguistics*. ACL.
2. T. M. (2015). "News clustering approach based on discourse text structure". *Proceedings of Association for Computational Linguistics and The Asian Federation of Natural Language Processing*. ACL.
3. Ashraf Ullah, K. B. (September 2014). Mining opinion components from unstructured reviews: A review. *Journal of King Saud University - Computer and Information Sciences*, 26, Pages 258-275.
4. Bonnie Webber, A. J. (2012). Discourse Structure and Computation: Past, Present and Future. *ACL*.
5. C. (2014). "Understanding spoken discourse". *Elsevier Ltd. Encyclopedia of Language and Linguistics(2nd edition)*.
6. D. H. (2018). "CASCADE: Contextual Sarcasm Detection in Online Discussion Forums". *27th International Conference on Computational Linguistics*.
7. E. M. (2004). "The Penn Discourse Treebank". *Proceedings of the ACL-2004*.
8. E. N. (2007). "Lexical and Discourse Analysis of Online Chat Dialog". *International Conference on Semantic Computing*.
9. F. B. (2018). "Introduction to the Special Issue on Language in Social Media: Exploiting Discourse and Other Contextual Information". *Special Issue on Language in Social Media*.
10. M. T. (2016). "Textual Analysis of Print Media Text Discourse". *New Media and Mass Communication, IISTE*.
11. N. J. (1992). "A Discourse Analysis of Software Documentation: Implications for the Profession". *IEEE transactions on professional communication, vol. 35, no. 3*.
12. R. P. (2018). "Discourse Annotation in the PDTB: The Next Generation", *Proceedings of ACL. Proceedings of ACL*. ACL.
13. T. V. (2002). "Adopting New IT by Learning in Groups: Results of Discourse Analysis". *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03), IEEE*. IEEE.
14. Y. V. (2013). "Improving STEM Classroom Culture: Discourse Analysis". *IEEE transaction*.
15. Lydia Manikonda et.al. (2014). "Discourse Analysis of User Forums in an Online Weight Loss Application".
16. Stubbs. (1983).
17. Valentina Pyatkin, B. W. (2017). "Discourse Relations and Conjoined VPs: Automated Sense Recognition". *Proceedings of Association for Computational Linguistics*. ACL.
18. Yizhong Wang et.al. (2018). "Toward Fast and Accurate Neural Discourse Segmentation". *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics*.
19. Zhang, Y. (2010). "An Analysis of Spoken Discourse between Two Native Speakers", . *Review of European Studies*.