Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
VOLUME:9, ISSUE:1, JANUARY 2026, PAPER NUMBER:15

Discourse Analysis To Uncover Linguistic Structures From Texts

Amirti Shiva[1], Durga Prasad Rath[2], Malla Vinay[3] , Meela Mothilal[4]

CH Dhawaleswar Rao[5],

[1,2,3,4]Students,and [5] Faculty
Dept. of Computer Science Engineering,
Centurion University of Technology and Management
amirtishiva@gmail.com

Abstract*: Discourse analysis is a linguistic approach that looks behind the sentence-level structures of language to study meanings, their construction, and interpretation, in larger texts and social contexts. It emphasizes cohesion and coherence in the linguistic pattern as influenced by and constituting societal dynamics and power relations. Unlike general linguistics, it traces broader linguistic and social structures. Features such as pronoun use, deixis, and discourse markers are analyzed in methodologies like conversation analysis, critical discourse analysis, and corpus-based approaches to reveal deeper insights. This approach is instrumental in understanding human behavior, intent, and interaction and advancing fields like sociology, anthropology, and artificial intelligence. Discourse analysis has facilitated developments in natural language processing, enabling machines to interpret language meaningfully in context. This becomes a major instrument for analyzing the social functions of language and their consequences for communication.*

Keywords: *Natural Language Processing(NLP), SVO Structure, Dependency Parsing, Coreference Resolution, Flask API, Pronoun Resolution*

1. INTRODUCTION

The challenge is to identify underlying structures that lie at the root of computational and natural language processing, since human communication, is increasingly, going digital. It is exponentially calling for more refined methods of analysis of textual discourse.[1]This article presents an integrated methodology combining traditional linguistics with modern computing techniques for the examination of complex linguistic structures in texts. This approach extends beyond lexical and syntactic analysis to reveal complex relationships, hierarchical organization, and coherence in textual elements.[2]Discourse analysis is an important means of understanding how language functions beyond sentences. Deeper patterns and structures comprise meaning; thus, surface-level linguistic expressions correspond to deeper structures in modes of communication.[3]This approach consists of micro-linguistic elements such as lexical choices and syntactic patterns, as well as macro-linguistic structures such as thematic development and rhetorical organization.

It discusses how language plays with social, cultural, and contextual factors in constructing the more abstract systems of meaning.[4]There is an urgent need for sophisticated tools and methods in this digital world that can address complexities in communication with nuanced linguistic patterns.

This framework, therefore, offers a systematic approach for the identification, analysis, and understanding of complex structures that underpin textual communication. The framework links theoretical linguistics to its practical applications in computational analysis and natural language processing, hence combining robustness and adaptability to different contexts and disciplines.

2.RELATED WORK

A. Liu R., Mao,R.,Luu,A.T.,&Cambris, E.(2023)

Title: *A brief survey on recent progress in figuring out who's who in text*Journal: *Artificial Intelligence Review*Why this paper is a big deal:

- It rounds up all the cool new ways to tell who's who in texts, like those fancy transformer models – think BERT, ELECTRA – and even ways that work with different languages.
- It points out how this stuff helps with other computer language projects, like figuring out feelings in text or switching between languages.
- It spots some tricky bits, like helping out languages that don't have much tech support and making it work fast on loads of data.

B. Atkinson, J., & Escudero, A. (2022)

Title: *Evolutionary natural-language coreference resolution for sentiment analysis*Journal: *International Journal of Information Management Data Insights*Why it rocks:

- Makes a connection between coreference resolution and sentiment analysis tackling confusion in pronouns with feelings attached like "it" and "they".
- Evolves algorithms for the best linking of entities in messy texts like tweets or comments.
- useful for digging into opinions and getting insights for businesses.

C. Guarasci, R., Minutolo, A., Damiano, E., De Pietro, G., Fujita, H., & Esposito, M. (2021).

Title: *ELECTRA for figuring out who's referring to who in Italian*Journal: *IEEE Access*Why you should care:

- Shows how to sort out who's talking about who in multiple languages using ELECTRA for Italian, which isn't a language with lots of resources.
- Hits top scores by making pretrained transformers even better.
- Points out how neural tricks can be used for languages other than English.

D. Bohnet, B., Alberti, C., & Collins M. (2023)

Title: *Sorting out references in sentences with a seq2seq transition-based method*Journal: *Transactions of the Association for Computational Linguistics*Why you should care:

- This new sequence-to-sequence (seq2seq) model switches up the old clustering way of fixing coreference resolution.
- It's better at dealing with big long texts because of its step-by-step figuring out process.
- This bad boy shows up earlier models when you look at how well it does on tests like OntoNotes.

E. Evans R., & Orăsan, C. (2019)
Title: *Spotting clues of syntactic complexity in rule-based sentence streamlining*Journal: *Natural Language Engineering*Its significance:

- Ties "SVO analysis" with making syntax less complex, like when you need to make text easier to get.
- Figures out tricky structures with the help of dependency parsing and "SVO patterns", you know stuff like when sentences are packed inside other sentences.
- Super important for stuff like teaching and tools that help people out.

3.METHODOLOGY

Input Processing Phase handles raw text that's organized or not organized such as documents, articles, or conversations. This step makes sure to preserve the original features of the text. This helps to maintain the data's integrity for future tasks. The input that's now in a standard form becomes the base for the language changes that follow..

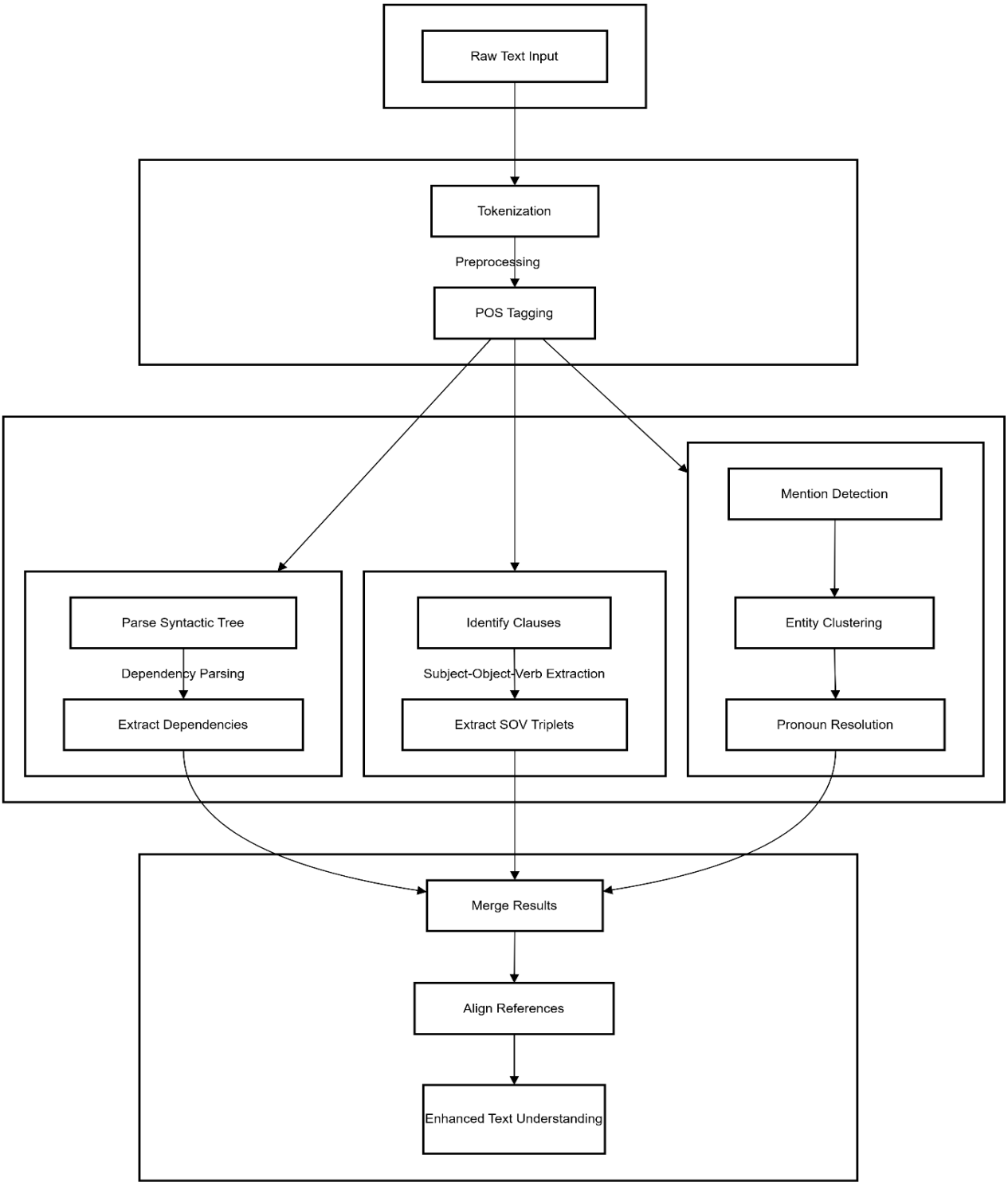
The preprocessing Phase changes raw text into organized language units. It breaks down the text into separate tokens (words punctuation) using methods based on rules or libraries. This process deals with shortened forms and words that are joined together to keep the text making sense. Then, it labels each token with its grammar role (like nouns and verbs) using models based on statistics or neural networks. This uses information from the surrounding text to get the data ready for deeper analysis of sentence structure and meaning.

Parallel Processing Streams deal with different parts of language at once. The first stream, **Dependency Parsing**, creates trees to show how words connect in a sentence. It does this by spotting grammar links using set rules or patterns from data. This creates structured graphs showing these dependencies. At the same time, the **Subject-Object-Verb (SOV) Extraction** stream finds clauses and pulls out groups of three related words. This captures the main meaning of sentences working with both simple and complex sentence structures. The third stream, **Coreference Resolution**, spots mentions of entities and groups them together using metrics based on context and similarity. It clears up unclear references by connecting pronouns to the words they stand for.

In the **Integration Phase**, we bring together the outcomes from these different processes. We combine syntactic dependencies semantic triplets, and coreference chains to build a single unified representation. This step has an impact on settling conflicts like overlapping entities and makes sure all streams stay consistent. To keep entity tracking coherent across sentences and documents, we align references, which helps to clear up any lingering uncertainties. The final result is a complex linguistic network that ties together syntactic structure semantic relationships, and entity groups. This network enables a deeper grasp of the text.

In the **Integration Phase**, we bring together the outcomes from these different processes. We combine syntactic dependencies semantic triplets, and coreference chains to build a single unified representation. This step has an impact on settling conflicts like overlapping entities and makes sure all streams stay consistent. To keep entity tracking coherent across sentences and documents, we align references, which helps to clear up any lingering uncertainties. The final result is a complex linguistic network that ties together syntactic structure semantic relationships, and entity groups. This network enables a deeper grasp of the text.

The results—syntactic trees semantic triplets, and resolved entity clusters—help with tasks like summing up text, pulling out information, and answering questions. This approach strikes a balance between being modular (through parallel streams) and working as a whole using tools such as spaCy or Stanza to put it into action. It focuses on being able to scale up and analyze with awareness of context, checked against language benchmarks to make sure it's strong enough for real-world NLP jobs.



4. Conclusion

A conclusion section is Discourse analysis is the best method that uncovers some inherent linguistic structure within the text, which brings ideas of syntactic, semantic, and referential relationships. More advanced NLP techniques may include dependency parsing, SVO extraction, and coreference resolution, which can allow structured representations of the language to help computers understand meaning much like the way humans do.

Dependency Parsing produces a dependency representation of sentences, but now words are presented as they are dependently connected via grammatical dependencies. The ability to identify dependencies leads to the breakdown of complex sentence structures into comprehensible syntax trees which in turn is the ground for extracting meaningful grammatical roles such as subjects and objects and modifying elements and, in turn, forms crucial support for any further process of NLP analysis done on the text by structuring it in a way that is both readable and effective for further application.

Thus, SVO Extraction continues this pattern by focusing on basic unit meanings: subject, verb, and object, in every clause. This limits the meaning of a sentence to triplets of semantic value, so the intent of a sentence becomes easy and straightforward to express, thereby allowing applications such as information retrieval and summarization by simplifying the intent of the statement.

Coreference Resolution identifies the cohesiveness of meaning within sentences by relating mentions of the same entity. Mention detection, entity clustering, and pronoun resolution ensure that the references are tracked coherently in the text, which is fundamental for tasks like text summarization and dialogue systems. This lets discourse analysis move beyond just the understanding of one sentence and helps in having a global interpretation of text where references are consistent and meaning is preserved across a wider context.

Collectively, all these techniques enable discourse analysis to delve into the linguistic structures of a text transforming apparently unsystematic language into systematic knowledge. Discourse analysis captures intricate patterns of human communication by linking syntactic, semantic, and referential layers, providing wide applications for AI and language processing.

5. References

[1] Liu, R., Mao, R., Luu, A. T., & Cambria, E. (2023). A brief survey on recent advances in coreference resolution. Artificial Intelligence Review, 56(12), 14439-14481.

[2] Šteflovič, K., & Kapusta, J. (2023). Coreference Resolution for Improving Performance Measures of Classification Tasks. Applied Sciences, 13(16), 9272.

[3] Scheffczyk, J. (2023). Introducing the component coreference resolution task for requirement specification (Doctoral dissertation, Universität Bonn Germany).

[4] Liu, R., Mao, R., Luu, A. T., & Cambria, E. (2023). A brief survey on recent advances in coreference resolution. Artificial Intelligence Review, 56(12), 14439-14481

[5] Luu, A. T., & Cambria, E. (2023). A brief survey on recent advances in coreference resolution. Artificial Intelligence Review, 56(12), 14439-14481

[6] Chong, M. Y. M. (2013). A study on plagiarism detection and plagiarism direction identification using natural language processing techniques.

[7] Lapponi, E. (2012). Why Not!: Sequence Labeling the Scope of Negation Using Dependency Features (Master's thesis).

[8] Toh, Z., & Wang, W. (2014, August). Dlirec: Aspect term extraction and term polarity classification system. In Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (pp. 235-240).

[9] Meng, Y. (2012). Sentiment analysis: A study on product features.

[10] Eskander, R., Muresan, S., & Collins, M. (2020, November). Unsupervised cross-lingual part-of-speech tagging for truly low-resource scenarios. In Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP) (pp. 4820-4831).