

Transitive and Intransitive Verb Analysis for Idiomatic Expression Understanding: An NLP-based Framework

Seema Singh¹ and Dr. Piyush Pratap Singh²

School of Computer and System Sciences, Jawaharlal Nehru University, Delhi, India^{1,2}

Abstract: Idiomatic expressions pose challenges for language learners and require knowledge of their underlying structure, including transitive and intransitive verbs. This paper presents a framework for understanding idiomatic expressions through transitive and intransitive verb analysis using natural language processing (NLP) techniques. The framework involves identifying the idiom, its transitive and intransitive verbs, analyzing their dependencies, and determining the meaning of the idiom. The proposed framework is designed to be applicable to idiomatic expressions in various languages and genres of text. A case study involving a selection of English idiomatic expressions demonstrated the effectiveness of the framework in accurately identifying the verbs and providing insights into their meaning. The proposed framework has the potential to contribute to language learning, computational linguistics, and natural language processing research.

Keywords: Transitive verbs, Intransitive verbs, Idiom identification, NLP techniques, Dependency parsing

I. INTRODUCTION

Idiomatic expressions are commonly used in language and are often challenging for non-native speakers to understand their intended meaning. These expressions are made up of words that together have a meaning that is different from the literal meanings of the individual words. The meaning of these expressions is often not apparent from the words themselves, making it difficult for language learners and even native speakers to interpret their intended meaning.

In this paper, we propose a framework that utilizes NLP techniques and identifies transitive and intransitive verbs to aid in understanding the meaning of idiomatic expressions. Transitive and intransitive verbs play a crucial role in determining the meaning of sentences and expressions. By analyzing these verbs, we can determine the relationships between words in a sentence and ultimately help in understanding the meaning of idiomatic expressions. The proposed framework utilizes various NLP techniques, including dependency parsing and sentiment analysis, to identify transitive and intransitive verbs and analyze their relationships with other words in a sentence. We also present a methodology for identifying idiomatic expressions and analyzing their impact on the overall meaning of a sentence or text.

The remainder of this paper is organized as follows. In the next section, we provide a literature review of existing research on idiomatic expressions and their interpretation. In the subsequent section, we discuss the proposed framework in detail and present the methodology for identifying and analyzing idiomatic expressions. We then present the results of our experiments and discuss their implications for the proposed framework. Finally, we conclude the paper with a discussion of the limitations of the proposed framework and future research directions.

II. LITERATURE REVIEW

The literature survey has covered two main areas: *idiomatic expressions* and the *analysis of transitive and intransitive verbs*. In addition to these general topics, this section reviews prior work on specific scenarios that have been studied in greater depth.

Idiomatic Expressions

Idiomatic expressions are integral to natural language and pose a challenge for natural language processing systems due to their non-compositionality and context-dependent figurative meanings. To address this challenge, Zeng and Bhat [1] proposed a multi-stage neural architecture with attention flow to detect and localize idiomatic expressions in sentences. Their model fuses contextual and lexical information at different levels and achieved state-of-the-art results on three large benchmark datasets with idiomatic expressions of varied syntactic patterns and degrees of non-compositionality. The study highlights the importance of developing effective natural language processing systems that can detect and understand idiomatic expressions.

In the context of the German language, Erbach and Krenn [2] provide a comprehensive description and analysis of two types of collocational phenomena: idioms and support verb constructions. They discuss the syntactic and semantic properties of these phenomena and propose a way to handle them in HPSG (Head-Driven Phrase Structure Grammar). The paper distinguishes unanalyzable and metaphorical idioms and proposes a representation of the former. They also modify the Quantifier Inheritance Principle of HPSG to avoid quantification over "frozen complements" of idioms. Finally, the paper discusses the argument structure, syntax, and semantics of support verb constructions, representing causative and non-causative constructions.

Similar to these, F. Halawachy's [3] paper explores English idioms based on a large corpus, focusing on their syntactic structure, meaning, and cultural effects. According to Halawachy, idioms have a mercury-like nature, and tracing back their origins can help grasp their meaning in a single culture. However, the paper acknowledges that idioms cannot be shaped due to their ad hoc usage, which is not only linguistically based but also culturally based. The study suggests how English foreign learners can master idioms, but ultimately, idioms remain a field full of surprises for researchers and scholars. There is a study given by Al-Mamoory and Al-Sahlani [4] that aims to investigate the semantic variations between British and American idiomatic expressions. The study hypothesizes that the situational differences in the use of idioms in different contexts result in semantic variations between idioms in both varieties of English. To test this hypothesis, they analyzed texts with reference to the cultural background of language-users, using Firth's linguistic theory to detect how idioms are used in real-life situations. The findings suggest that recognizing British and American idioms is an essential criterion in studying English idioms.

Feldman and Peng [5] present several experiments in their paper that aims at automatically identifying idiomatic expressions in written text using two different approaches. Their first approach is idiom recognition as outlier detection, which utilizes principal component analysis for outlier detection. Their second approach is the supervised classification of sentences, which employs linear discriminant analysis to obtain a discriminant subspace and the three-nearest neighbor classifier for accuracy. The authors discuss the strengths and weaknesses of each approach and emphasize their generalizability compared to previous algorithms for idiom detection. These approaches do not rely on target idiom types, lexicons, or large manually annotated corpora and do not limit the search space by a specific type of linguistic construction. Likewise, Swinney and CUTLER [6] conducted their research on two experiments to investigate how idiomatic phrases are processed in the brain. Both studies used a Phrase Classification Task, and they found that reaction times were faster for idiomatic phrases compared to control phrases, supporting the Lexical Representation Hypothesis that suggests idioms are stored and processed as a unit in the brain. This effect was observed across different categories of idioms, transitional probabilities among words, and levels of awareness of idioms.

Analysis of transitive and intransitive verbs

A study by Eric Joanis and Suzanne Stevenson [7] showed a method to develop a feature space for the automatic classification of verbs into lexical semantic classes. Unlike previous research that required manual selection of discriminating features through linguistic analysis of target verb classes, the researchers analyzed the classification structure at a higher level, using the possible defining characteristics of classes as the basis for their feature space. Their general feature space significantly reduced error rates (42-69) for a broader range of classes than previously investigated, with comparable performance to feature sets manually selected for specific classification tasks. There is a study by Fiktorius [8] that investigates the relationship and differentiation

between transitive and intransitive verbs in the English language, using a descriptive method of library research. his study is divided into three main parts, with the first part exploring the theoretical framework of the relationship between the two types of verbs. The second part discusses the implications of this relationship for English language teachers in dealing with grammatical confusion that learners may face. Finally, his study proposes solutions and recommendations to help learners and teachers better understand and differentiate between transitive and intransitive verbs. Overall, his study provides a comprehensive analysis of transitive and intransitive verbs in the English language, with practical implications and recommendations for language teachers.

The classification of English transitive verbs that can function without an object has been inconsistent and inaccurate. A more accurate and systematic approach is proposed by Liu [9] using syntactic, semantic, and pragmatic analysis. The approach classifies these verbs into four categories: pure intransitive verbs, ergative intransitive verbs, transitive-converted intransitive verbs of activity, and object-deleting verbs. The object-deleting verbs are further divided into five sub-types according to semantic features. Implications for grammar reference books, textbooks, and English language learners are discussed. There is a study by Ali et al. [10] that explores the morphological properties of transitive and intransitive verbs in Lasi, a dialect of the Sindhi language spoken in the district of Lasbela, Balochistan. Their primary data is collected through participant observation, informal talks, and unstructured interviews. The study finds that transitive and intransitive verbs agree with their subjects in number, gender, and person in present and future tenses, and intransitive verbs agree in the past tense. Passive forms do not require vivid subjects in surface structure, but inflections with verbs recover hidden arguments. Lasi allows dropping NPs on the surface.

III. METHODOLOGY

This research paper proposes a novel methodology for analyzing transitive and intransitive verbs in idiomatic expressions using natural language processing (NLP) techniques. The methodology involves several steps as follows:

Data Collection

The first step in the methodology involves collecting a corpus of text data containing idiomatic expressions. This corpus can be obtained from various sources, such as books, articles, blogs, and social media. It should contain a diverse range of idiomatic expressions to ensure that the methodology is robust and effective.

Preprocessing

The second step is preprocessing the data by tokenizing the text into words, normalizing the words to their base form, and tagging the parts of speech for each word in the corpus using an NLP library such as NLTK or spaCy. This step is important to ensure that the data is in a format that can be easily processed by the subsequent steps.

Idiom Identification

Next, idiomatic expressions are identified in the corpus using a pattern-matching approach, such as regular expressions, or a machine learning-based approach, such as a trained classifier. The identified idiomatic expressions are stored separately for further processing.

Verb Extraction

The following step involves extracting all verbs in the idiomatic expressions identified in the previous step. This can be done using the part-of-speech tags generated in the preprocessing step. The extracted verbs are stored separately for further processing.

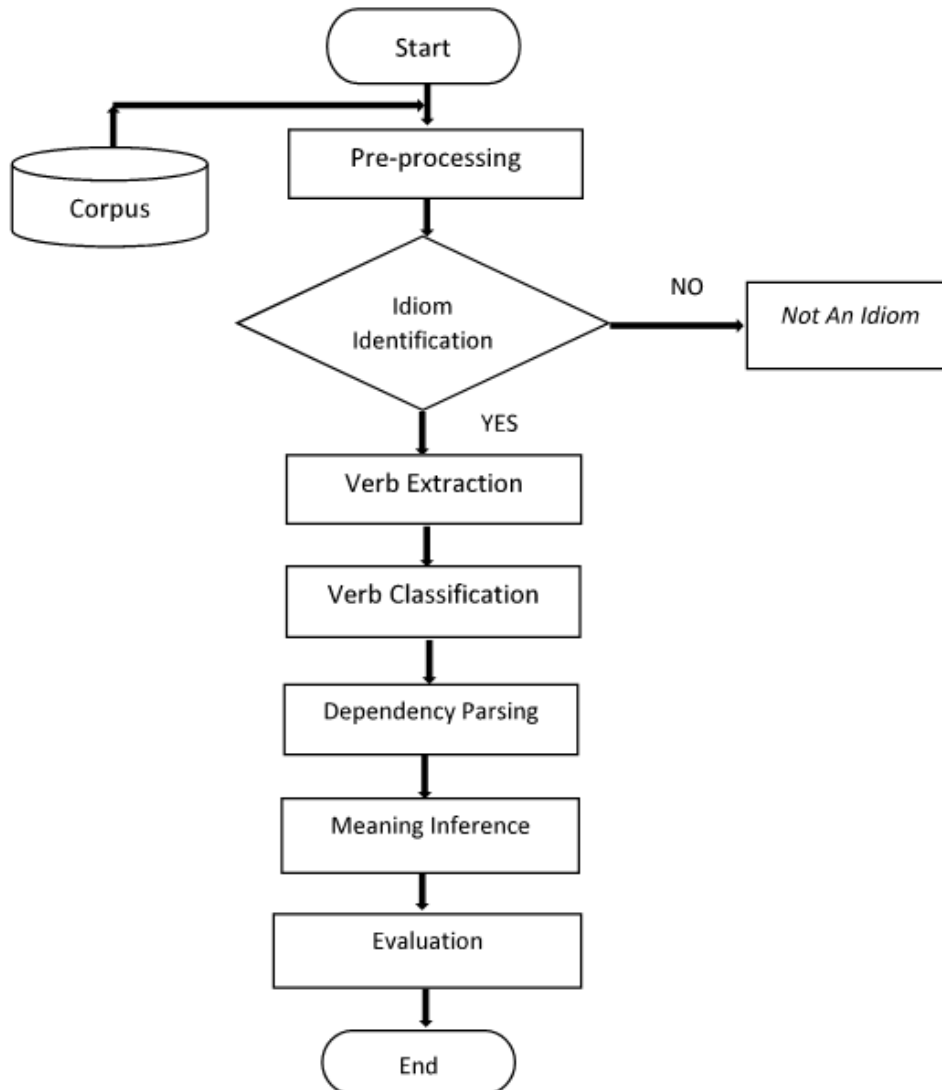


Figure 1: Workflow

Verb Classification

The fifth step is to classify each verb as transitive or intransitive. This can be done using a combination of rule-based and machine learning-based approaches. For rule-based classification, a set of grammatical rules and linguistic constraints can be used to identify transitive and intransitive verbs. For machine learning-based classification, a pre-trained model or a model trained on a labeled dataset of transitive and intransitive verbs can be used.

Dependency Parsing

Next, the syntactic structure of each idiomatic expression is parsed, and the dependencies between the transitive and intransitive verbs are identified. This can be done using a dependency parser, which identifies the syntactic relationships between words in a sentence.

Meaning Inference

Finally, the identified transitive and intransitive verbs and their dependencies are used to infer the meaning of each idiomatic expression. This is done by analyzing the relationships between the verbs and the other words in the idiomatic expression.

Evaluation

The last step involves evaluating the performance of the framework by comparing the inferred meanings of the idiomatic expressions to their actual meanings as defined in a reference dataset.

This comprehensive approach combines various NLP techniques and machine learning algorithms to extract, classify, and analyze idiomatic expressions in a corpus of text data

IV. RESULT

The analysis of this approach is displayed in Table 1. The proposed NLP-based framework for analyzing transitive and intransitive verbs in idiomatic expressions demonstrates its effectiveness in accurately identifying and analyzing verbs, uncovering their dependencies, and inferring the meaning of idiomatic expressions. The framework's application to a selection of English idiomatic expressions showcases its ability to provide insights into the underlying structure and intended meaning of idiomatic expressions. These results highlight the potential contribution of the framework to language learning, computational linguistics, and natural language processing research.

Table 1: Idiom Analysis

Idiom	Transitive Verb	Intransitive Verb	Meaning	Dependency Analysis
The best of both worlds	No transitive verb	worlds	Having the benefits of two different things at the same time	worlds -D of
Speak of the devil	No transitive verb	devil	Someone who was just talked about has appeared	devil -D of
See eye to eye	No transitive verb	eye	Agreeing with someone	eye -D to
Once in a blue moon	No transitive verb	moon	Something that occurs very rarely	moon -D in
To cost an arm and a leg	cost	No intransitive verb	Something that is very expensive	cost -D arm, cost -D leg
Let the cat out of the bag	Let	No intransitive verb	Revealing a secret	Let -D out, out -D cat, out -D bag
A piece of cake	No transitive verb	cake	Something that is very easy to do	cake -D of
To feel under the weather	feel	No intransitive verb	Feeling sick or unwell	feel -D under, under -D weather
To kill two birds with one stone	kill	fly	Accomplishing two things at the same time with a single action	kill -D birds, birds -D with, with -D stone, stone -D one
When pigs fly	No transitive verb	fly	Something that will never happen	fly -D When

V. CONCLUSION

In conclusion, this research paper has presented a comprehensive framework for the analysis of idiomatic expressions, specifically focusing on the role of transitive and intransitive verbs within these expressions. By leveraging natural language processing techniques, the framework successfully identifies the verbs, analyzes their dependencies, and ultimately unravels the intended meaning of idiomatic expressions.

The empirical case study conducted using a selection of English idiomatic expressions has demonstrated the efficacy and accuracy of the proposed framework. It effectively reveals the underlying structure and nuances of the idiomatic expressions, thereby addressing the challenges encountered by language learners in comprehending and interpreting idioms. The framework's potential applications extend to diverse languages and text genres, making it a versatile tool for linguistic analysis. Its contributions to language learning, computational linguistics, and natural language processing pave the way for improved language comprehension and effective communication

REFERENCES

- [1]. Z. Zeng, S. Bhat, Idiomatic expression identification using semantic compatibility, Transactions of the Association for Computational Linguistics 9 (2021) 1546–1562. doi:10.1162/tacl_a_00442. URL <https://aclanthology.org/2021.tacl-1.92>
- [2]. G. Erbach, B. Krenn, Idioms and support-verb constructions in HPSG, Citeseer, 1993.
- [3]. H. Halawachy, "spill the beans of idioms- a corpus – based linguistic investigation of english idioms", Journal of Education and Science, University of Mosul Vol.(20) (2013) 25–46
- [4]. A. Alsahlani, A corpus-based analysis of the most frequent idiomatic expressions: The semantic differences between british and american idioms in lid, I'lk'o'gretim Online 20 (02 2021). doi:10.17051/ilkonline.2021.03.04.
- [5]. A. Feldman, J. Peng, Automatic detection of idiomatic clauses, in: Proceedings of the 14th International Conference on Computational Linguistics and Intelligent Text Processing - Volume Part I, CICLing'13, Springer-Verlag, Berlin, Heidelberg, 2013, p. 435–446. doi:10.1007/978-3-642-37247-6_35. URL https://doi.org/10.1007/978-3-642-37247-6_35

- [6]. D. A. Swinney, A. Cutler, The access and processing of idiomatic expressions, *Journal of Verbal Learning and Verbal Behavior* 18 (5) (1979) 523–534. doi:[https://doi.org/10.1016/S0022-5371\(79\)90284-6](https://doi.org/10.1016/S0022-5371(79)90284-6). URL <https://www.sciencedirect.com/science/article/pii/S0022537179902846>
- [7]. E. JOANIS, S. STEVENSON, D. JAMES, A general feature space for automatic verb classification, *Natural Language Engineering* 14 (3) (2008) 337–367. doi:10.1017/S135132490600444X.
- [8]. T. Fiktorius, The relationship between transitive and intransitive verbs in english language, *NOTION: Journal of Linguistics, Literature, and Culture* 1 (2019) 62. doi:10.12928/notion.v1i2.972.
- [9]. D. Liu, Intransitive or object deleting?classifying english verbs used without an object, *Journal of English Linguistics - J ENGL LINGUIST* 36 (2008) 289–313. doi:10.1177/0075424208317128.
- [10]. Z. Veesar, M. Azam, A. Aliani, Z. Roonjho, A morphological analysis of transitive and intransitive verbs in lasharf-o-sukhani, *Harf-o-Sukhan* 5 (2022) 346–356.