A Novel Approach for Identification of Figurative Language Types in Devanagari Scripted Languages

Jatinderkumar R. Saini, Preety Sagar, Hema Gaikwad
Symbiosis Institute of Computer Studies and Research,Symbiosis International (Deemed University) Pune, India

***Abstract*—Poetry can be defined as a form of literary expression that uses language and artistic techniques to evoke emotions, create imagery, and convey complex ideas in a concentrated and imaginative manner. It is a form of written or spoken art that often incorporates rhythm, meter, rhyme, and figurative language to engage the reader or listener on multiple levels. There is no automated system that can identify figures of speech (FsoS) in poetry using Natural Language Processing (NLP) methods. In this research paper, the authors categorized four types of FsoS व्रत्या अनुप्रास (Type of alliteration), छेकानुप्रास (Type of alliteration), अन्तत्यानुप्रास (Rhyme), and पुनरुक्ति (Repetition) using two custom algorithms, Koshur and Awadhi (KA and AA), developed specifically for three different language corpora of poems (Koshur (K), Awadhi (A), and Hindi (H)). To evaluate the effectiveness of these algorithms, the authors conducted tests on three languages using four distinct approaches: with stopwords without optimization, with stopwords with optimization, without stopwords without optimization, and without stopwords with optimization. Authors have put lots of effort into identifying FsoS in not only one single language but in three Devanagari scripted languages. This research work is the first of its kind. The average accuracy without stopwords was not up to the mark. The authors then optimized both algorithms and again tested them on the same corpora with or without stopwords, resulting in a significant increase in accuracy.**

Keywords*—Figures of speech (FsoS); natural language processing (NLP); Koshur; Awadhi

I. INTRODUCTION

India is renowned for its rich cultural heritage and vast literary traditions. With twenty-two scheduled languages, the country boasts a remarkable diversity of literature, and poetry stands out as one of its prominent forms of expression. Indian poetry, rooted in ancient times, has flourished through the ages, reflecting the myriad influences of its diverse regions and communities.

Poetry is characterized by its unique use of language, employing words and phrases chosen for their sound, texture, and connotation, as well as their literal meaning. Poets often strive to capture the essence of an experience or emotion, using vivid imagery and metaphorical language to paint a picture or evoke sensory impressions. While poetry can take many forms and styles, ranging from traditional structures like sonnets and haikus to free verse and experimental formats, it is united by its emphasis on creative expression, aesthetic beauty, and the power of words. It explores themes such as love, nature, human experiences, social and political issues, and the mysteries of existence. To acquire a deeper knowledge of a given poetry, we might focus on numerous elements. These aspects include Emotion (Rasa), voice, diction, imagery, figures of speech, syntax, sound, rhythm, and verse meter [1]. These unfold the poem's actual theme, meaning, or feelings. The figure of speech (FoS) is one such element that creates a mesmerizing effect and produces magic. They are considered the ornaments used for the beautification of the poems.

This is a maiden attempt by the authors of this research paper to identify four different forms of figures of speech (FoS), namely व्रत्या अनुप्रास 'Type of alliteration', छेकानुप्रास 'Type of alliteration', अन्तत्यानुप्रास 'Rhyme' and पुनरुक्ति 'Repetition' (which is an implicit type of alliteration) on three different language corpora Koshur, Awadhi, and Hindi, implemented in two different algorithms Koshur and Awadhi algorithm (KA and AA) and later optimized with and without stopwords. All three languages, Koshur Awadhi, and Hindi are based on the Devanagari script.

Kashmiri or commonly referred to as 'Koshur', is a branch of the Indo-Aryan language in the linguistic landscape of India spoken by about seven million people from the state of Jammu and Kashmir, India. The Perso-Arabic script, the Sharda script, and the Devanagari script are the three main scripts used to write Koshur [2,3]. The Perso-Arabic script is considered the official script of Koshur, and Devanagari is used by Kashmiri Hindus for literature purposes [4]. The pioneers of literature, the Kashmiri poets, brought freshness and sweetness to the poems using this script [5]. The corpora considered in this research paper were Koshur poems written in Koshur Devanagari script.

Awadhi is spoken in twenty districts of India and eight districts of Nepal [6]. Awadhi is a language spoken in the Awadh region of Uttar Pradesh and some parts of north India. Awadhi is one of the dialects of Hindi, and it has 38 million native speakers [7]. But very less research work has been performed on the Awadhi language. Prominent texts like Ramcharitmanas, one of the most important religious texts in the world, is written in Awadhi. The Epic Ramcharitmanas is traditionally divided into several major kaandas or books, that deal chronologically with the major events in the life of Rama—Bala Kaanda, Ayodhya Kaanda, Aranya Kaanda, Kishkindhaa Kaanda, Sundara Kaanda, Lanka Kaanda, and Uttara Kaanda. Ramcharitmanas has seven Kaanda such as the other text like Hanuman Chalisa and Padmavat are also written in Awadhi. [8][9][10]. Today, the world is rapidly heading towards digitization. e-Books are easy to publish, handle and promote. All the important ancient books and texts are being digitized.

Hindi is the official language of India and holds a significant place in the country's cultural and literary heritage. Hindi belongs to the Indo-Aryan language family and shares roots with other languages, such as Sanskrit. Poetry in Hindi is a rich and diverse art form that has flourished for centuries. It encompasses various forms and styles, including Bhakti poetry, Ghazals, Doha, and Geet. Each form has its unique structure, rhythm, and thematic focus.

Automated identification of four types of FsoS in Devanagari scripted three major languages, Koshur, Awadhi, and Hindi, are first of its kind, and hybridization of two separate algorithms further improved the performance. No such work, as per authors' best-known knowledge, is seen with special reference to three languages. The author's contribution to this research covers the following items: -

- Identification of four different FsoS, using two different algorithms and their implementation on three different language corpora.
- Optimization of algorithms with and without stopwords.
- Comparison of both KA and AA algorithms considering accuracy and execution time taken byeach algorithm.

The remaining sections of the paper are organized as follows: Section II gives a detailed literature review of the work. Section III explains the methodology used along with the algorithms. Section IV shows the interpretation of the results. Section V concludes the work and discusses its future scope.

II. Literature Review

A thorough literature review was done to gather information about the present state of the research effort in the field of Natural Language Processing (NLP) which is a branch of Artificial Intelligence and Machine Learning, and work done in FoS in poetry. Research in poetry and nearby segments like poetry generation, sentiment analysis, text classification, meter classification, etc., has been seen in many Indian and foreign languages like Marathi, Punjabi, Hindi, Arabic, Chinese, Persian, etc., but still, many Indian regional languages are either completely absent or badly represented on the NLP map. Although authors can find some work in the Hindi language, but Koshur and Awadhi language are ones in which very little orno study has been done.

According to Chopra et al. [11] NLP is a field of study in computer science, artificial intelligence, and linguistics that investigates how computers interact with human or natural languages. NLP is primarily concerned with human-computer interaction. NLP was also felt necessary because computers could access much information recorded or stored in natural language. Saini and Kaur [1] worked on poetry characteristics such as diction, rhyme, and rhythm, setting it apart from other genres of literature. These factors were experimented with, to attempt an automatic system for categorizing poetries based on emotional states tested to develop a system for categorizing poetries based on the Indian concept of 'Navrasa.' Kushwah and Joshi [12] investigated Hindi poetry based on 'Chhand', one of the properties of Hindi poems. They created an algorithm that detects the presence of 'Rola Chhand' in any poem provided as input. A few poems are available in digital form, but their poetic properties are not aimed at, and their algorithm focuses on onesuch property.

Audichya and Saini [13] worked on producing automatic metadata for 'Chhand' based on the stanzas of the poems. They also provided superior techniques for metadata creation and procedures for 'Muktak Chhands'. It was the first time that not only rules of the 'Chhands' were identified but also were confirmed and modelled from the standpoint of computational linguistics. Audichya and Saini [14] worked on identifying three Hindi figures of speech using NLP. They also created a systematic structure of types and sub-types of Hindi FoS. Bafna and Saini

The study [15] recovered tokens from two corpora using two different methodologies. To count and contrast extracted tokens, BaSa, and Zipf's law were employed. Further token comparison between the two approaches is accomplished. They used both Hindi and Marathi poems and prose. To demonstrate that Hindi and Marathi behave similarly for NLP operations, common tokens from corpora of Marathi and Hindi poetry and prose were identified. It was established that BaSa outperforms Zipf's law. Kaur and Saini [16] worked on creating a content-based classifier for Punjabi poetry. After going through the pre-processing layer, more than 2,034 poems and 31,938 tokens were separated and weighted using the term frequency (TF) and term frequency-inverse document frequency (TF-IDF).

Pal and Patel [17] classified poetry based on nine different types of Rasas like Shringar, Hasya, Rudra etc., and used a mix of part-of-speech and emotion-based features to classify poems into different types. In research by Lone et al. [18], a Kashmiri-to-English Machine Translation System was presented, as well as it highlighted various features of the Kashmiri language. Their method was built on machine intelligence, and it can learn various translation rules from a series of translated input wordsby employing Long Short-term Memory (LSTM) architecture for deep sequence learning. The paper also reports difficulties and challenges associated with the work. Mir and Lawaye [19] worked on Word Sense Disambiguation (WSD) System for Kashmir Language; they designed Sense Annotated Corpus for Kashmiri Language and WSD Data Set. Ahmad and Syam [20] developed a Parts-of-Speech tagger (POS) in Perso-Arabic scrip for the Kashmiri language with an accuracy of 80.64%. Rasool et al. [21] opened the doors to the creation of a powerful multilingual machine translation system that includes Kashmiri as one of the languages. The aim of the researcher here is to incorporate Kashmiri into UNL (Universal Networking Language) framework. In this work, a selected Kashmiri corpus is analysed to UNL using IAN, and subsequently, Kashmiri expressions are generated from UNL expressions using EUGENE. Gilkar et al. [22] proposed a POS rule-based tagger for Kashmiri written in the Nastaliq script. The authors tried to create an automatic tagger for Kashmiri corpora using a rule-based and stochastic (hybrid) tagging approach. Aabid et al. [23] presented an Automatic Recognition System (ARS) that allows computers to understand natural

speech for recognizing Kashmiri digits zero (sefar) to nine (nov) spoken in isolation by different male and female speakers. In study by Ramakrishna et al. [24] Kashmiri palatalized consonants were examined, which were related to, the i-matra vowel and palatalized in Kashmiri phonology. For the purpose of analysing Kashmiri voice data.

Firdaus et al. [25] explained that poem writing is a way to express ideas, thoughts, and feelings using artful language. Writing poetry is a skill that is taught in schools. There are hardly any students who are interested in poetry. Less appealing is the poetry writing process used in schools. As a result, the purpose of this study is to demonstrate an original way to write poetry using the "Atafora" technique by fusing students' sensory experiences with a metaphorical figure of speech. Malik et al. [26] discussed the figures of speech in the songs from Rose's album "R". The study's objective was to characterize four different types of figures of speech used in the two-track lists of the R album: (a) comparative figures of speech, (b) contradiction figures of speech, (c) affirmation figures of speech, and (d) satirical figures of speech. This study used a qualitative research design and an analytical framework. The -R- album's two songs with the title "On the Ground and Gone" that use figure of speech are the study's source of data. Krisna et al. [27] stated that poets frequently use figurative language to convey their thoughts and emotions. Poetry that captures the attention of the readers will be enhanced by figurative language. Paradox is frequently used in figurative poetry to emphasize the poem's message. This study focuses on the implications of paradox in the poem by Rudyard Kipling. The authors used the descriptive qualitative method to examine Rudyard Kipling's chosen poems' figure of speech. The study discovered 17 paradoxes, including seven rhetorical paradoxes, seven social paradoxes, two logical paradoxes, and one philosophy of science contradiction. Maula et al. [28] identified the figurative language in Lady Gaga's album titled “Always Remember Us This Way” using descriptive qualitative research. Wati et al. [29] discussed the comparative figurative language in the poems from Emi Suy's poetry collection titled Ibu Menanak Nasi hingga Matang Usia Kami. The anthology was published in 2022, served as the study's data source. Heuristic reading, a first-level semiotic reading technique, and Hermeneutic reading, a second-level structuralism-semiotic technique, was used for data collection and data analysis purpose, respectively. Naaz et al. [30] discussed different tools such as Text2Matr, RPaGen, and FoSCal. The Text2Matr tool offers a chanda-related observation, including chanda type detection and classification, rhythm determination, chanda correctness verification, etc. For an input Hindi poetry, the RPaGen tool provides rhyme pattern(s). The tool FoSCal, used for generating alankara scores, and the tool’s FoSCal uses the pattern created by RPaGen for alankara rating. Mahdi et al. [31] explained that poets use a wide range of writing strategies when creating new poems. This paper discussed the use of personification, symbolism, and figure of speech such as Simile and Metaphor in British love poems and lyrics. Nidi et al. [32] discussed the personification figure of speech and its general meaning. The two varieties of personification FsoS prosopography and prosopopoeia are used. The author has used the poems by Robert Frost.

Setiani et al. [33] stated that poetry is a literary form that includes stanzas, lines, rhythms, and rhymes. Denotative and connotative meanings are frequently utilized in poetry. Denotative means emotional feeling of the word, and connotative means figurative language. Hutauruk et al. [34] mentioned the importance of figures of speech. The personification and apostrophe create pictures in the mind of reader or listener, and pictures help to convey the message faster than words. Sayakhan et al. [35] Figurative language is used frequently in ordinary conversation, popular music, television, and commercial topics as well as in classic works like Shakespeare and the Bible. The author has specially discussed two figures of speech such as personification and apostrophe. A person is just mentioned in an apostrophe, whereas with personification, inanimate objects are given human characteristics.

While performing the literature review the authors found few papers on FoS but no papers were found in Koshur and Awadhi.

Therefore, the gap with the authors was huge, the reason behind this gap is mainly that the languages are highly resourceless, it is tedious to work with such resourceless languages and researchers have to face lots of difficulties as no initial work has been carried out before.

To address this significant gap, the authors have embarked on developing FoS algorithms. The research question then becomes: How can the development of FoS algorithms alleviate the challenges posed by the lack of resources in these languages and contribute to narrowing this gap?

For the above mentioned research question authors have worked on the objectives like development, optimization and implementation of FoS algorithms for Devanagari scripted languages.

III. Methodology

To identify FsoS in three different languages, authors developed the logic represented in Fig. 1. The input to the flowchart was the poems in all three languages. The entered poems then need to be tokenized on the basis of sentences, words, and letters, and accordingly, rules were applied as per the type of FoS identified.

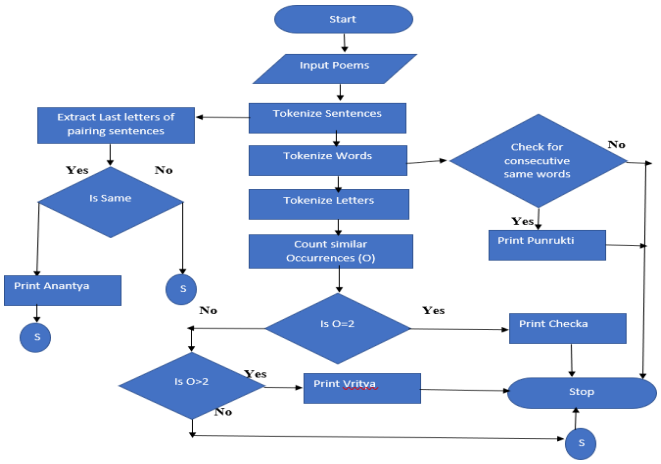


Fig. 1. Logical representation.

TABLE I. SUMMARY OF FOS, TRANSLITERATION, TRANSLATION,DEFINITION, AND EXAMPLE

FoS- N	T	RST	D	H	E	K	A
व्रत्या अनुप्रास	Vrity Anupras (VA)	Type of Alliteration	consonant word repeats more than once.	चारुचंद्र की चंचल किरणें, खेल रहीं हैं जल थल में|	Serve Fast food with fast speed	यि छु म्योन वतन यि छ् चोन वतन	बंदउ गुरु पद पदुम परागा। सुरुचि सुबास सरस अनुरागा
छेकानुप्रास	Cheka Anupras (CA)	Type of Alliteration	The same letter used repeatedly but only 2 times	फूले फले राष्ट्र दिन-रात मेरा	My country, my nation is blooming	वदुन रिवुन पॅतिम ति बार याद छुम में याद छुम	जो सुमिरत सिधि होइ गन नायक करिबर बदन
अन्तत्यानुप्रास	Anantya Anupras (AA)	Rhyme	Rhyming pattern at the end of lines	बुंदेले हरबोलों के मुँह हमने सुनी कहानी थी।खूब लड़ी मर्दानी वह तो झाँसी वाली रानी थी	On the fifteenth of May, in the jungle of nool, In the heat of the day, in the cool of the pool	सुमरन पनन्य दिचॉनम, लोलुक निशान वेसिये।रॅछरुन तोगुम न रोवुम, ओसुम न बान वेसिये	एहि महँ रघुपति नाम उदारा। अति पावन पुरान श्रुति सारा
पुनरुक्ति अलंकार	Punrukti (PU)	Repetition	one word used twice consecutively without changing its meaning	मीठा-मीठा रस टपकता	milky-milkey chocolate, covered with coffee beans.	जिंदगी हुंद मोदुर-मोदुर फरागथ.	बार बार रघुबीर सँभारी। तरकेउ पवनतनय बल भारी॥

Legends- FoS N-FoS Name,, T-, Transliteration , RST-Roman Scripted Translation, D-Definition, H Hindi, ,E-English-Koshur ,A-Awadhi

TABLE II. METADATA OF THREE DIFFERENT LANGUAGE POEMS

S.No.	Attributes	Koshur	Awadhi	Hindi
1	CT	DA	RAM	ABV
2	NoP	17	7	17
3	CS	341	340	353
4	ExT	2096	2819	1716
5	ExSW	510	264	862

Legends-Corpus Type- CT, DA-Different Authors, RAM-Ramcharitmanas, ABV-Poems of Shri Atal Bihari Vajpayee, NoP-Number of Poems, CS-CorpusSize, ExT-Extraction of tokens, ExSW-Extraction of Stop Words.

To detect or identify the figures of speech in Koshur, Hindi, and Awadhi, authors need detailed descriptions of all four types ofFsoS and their rules. Table I shows the definitions and examples of FsoS in all four languages.

The Table II above shows the details of corpus used. Well-known poems from three different languages were taken. Awadhi poems were extracted from different kaand of epic Ramcharitmanas where one kand represents one section. Throughout the paper, the authors used the word section in place of kand. For Hindi, poems of Shri. Atal Bihari Vajpayee ji had been taken, and poems of different authors were extracted for Koshur. The selection of Koshur poetry was made in order to include a variety of poems, including those about love, nature, patriotism, and other topics. For the implementation of the algorithm, lines of the poems were considered as Units of Measurement (UoM). In Awadhi, 40-50 lines were taken from each section. For Hindi and Koshur, seventeen poems were considered having 353 and 341 lines, respectively. Tokenization is a process of breaking sentences into words. After applying the tokenization process, the extracted words were 2819, 1716, and 2096 for Awadhi, Hindi, and Koshur, respectively. In NLP, stopwords are words that have less relevance and do not carry valuable information. The authors extracted 264, 862, and 510 stopwords in Awadhi, Hindi, and Koshur. Awadhi and Koshur are low-resource languages, and there are various challenges, such as the non-availability of linguistic documents like dictionaries, wordnet, thesaurus, stopwords list, POST etc. So, the extraction of stopwords was also a challenge for the authors.

As all three languages support the Devanagari script, the Hindi stopwords list has been taken as the benchmark. To identify Koshur stopwords, the authors have considered three lists of stopwords from three different languages (Hindi, English, and Punjabi). Some new words were also added to the list that were fetched from credible sources. For the identification of Awadhi stopwords, the authors have referred list of Hindi stopwords and modified them according to the Awadhi language. The authors also contributed new words from the literature of Awadhi.

IV. RESULT AND DISCUSSION

The authors developed two different algorithms, one for the Koshur language and another for Awadhi. Both algorithms were then implemented and tested on the corpus with stopwords and without stopwords. Optimization of algorithms was also done to improve the performance. Algorithms developed for the Koshur language were implemented in Hindi and Awadhi, and the algorithms developed for the Awadhi language were implemented in both Koshur and Hindi.

The performance of the Koshur and Awadhi algorithms on a corpus with stopwords and without optimization is shown in Table III. Table IV shows the performance of both algorithms with stopwords and with optimization.

TABLE III. IMPLEMENTATION OF KOSHUR AND AWADHI FOS ALGORITHMS IN THREE LANGUAGES WITH STOPWORDS WITHOUT OPTIMIZATION

	KA						AA					
FoS	**K**	**ET**	**A**	**ET**	**H**	**ET**	**K**	**ET**	**A**	**ET**	**H**	**ET**
VA	85.71	0.09	83.33	0.12	80.00	0.03	100.00	0.02	97.60	0.03	80.00	0.01
CA	93.33	0.08	86.66	0.25	94.11	0.06	96.40	0.01	97.10	0.04	94.80	0.02
AA	92.30	0.15	94.11	0.60	94.11	0.06	73.90	0.01	90.30	0.02	74.10	0.01
PU	91.66	0.15	92.30	0.60	92.30	0.06	96.10	0.03	100.00	0.01	100.00	0.01
Avg	90.75	0.12	89.10	0.39	90.13	0.05	91.60	0.02	96.25	0.03	87.23	0.01

Legend-ET-Execution Time

TABLE IV. IMPLEMENTATION OF KOSHUR AND AWADHI FOSALGORITHMS IN THREE LANGUAGES WITH STOPWORDS AND WITH OPTIMIZATION

	KA						AA					
FoS	**K**	**ET**	**A**	**ET**	**H**	**ET**	**K**	**ET**	**A**	**ET**	**H**	**ET**
VA	100.00	0.66	90.00	0.65	100.00	0.12	100.00	0.02	95.20	0.02	90.00	0.01
CA	93.33	0.70	100.00	0.83	94.11	0.15	98.30	0.05	98.80	0.04	96.80	0.03
AA	92.30	0.71	92.85	1.67	94.11	0.40	74.80	0.01	98.63	0.02	75.30	0.02
PU	91.66	0.70	80.00	1.67	93.30	0.40	96.20	0.02	100.00	0.01	100.00	0.01
Avg	94.32	0.70	90.71	1.21	95.38	0.27	92.33	0.03	98.16	0.02	90.53	0.02

TABLE V. IMPLEMENTATION OF KOSHUR AND AWADHI FOSALGORITHMS IN THREE LANGUAGES WITHOUT STOPWORDS WITHOUT OPTIMIZATION

	KA						AA					
FoS	**K**	**ET**	**A**	**ET**	**H**	**ET**	**K**	**ET**	**A**	**ET**	**H**	**ET**
VA	83.33	0.12	83.00	0.27	80.00	0.15	87.50	0.02	100.00	0.02	100.00	0.01
CA	53.33	0.12	86.66	0.36	58.82	0.15	98.76	0.01	97.25	0.02	100.00	0.01
AA	92.30	0.32	92.85	0.75	92.30	0.32	100.00	0.01	98.70	0.02	100.00	0.01
PU	91.66	0.31	80.00	0.74	92.30	0.31	85.71	0.01	77.78	0.01	83.33	0.01
Avg	80.16	0.22	85.63	0.53	80.86	0.23	92.99	0.01	93.43	0.02	95.83	0.01

TABLE VI. IMPLEMENTATION OF KOSHUR AND AWADHI FOSALGORITHMS IN THREE LANGUAGES WITHOUT STOPWORDS WITH OPTIMIZATION

	KA						AA					
FoS	**K**	**ET**	**A**	**ET**	**H**	**ET**	**K**	**ET**	**A**	**ET**	**H**	**ET**
VA	100.00	0.14	83.33	0. 06	100.00	0.06	87.50	0.02	100.00	0.01	100.00	0.01
CA	93.33	0.15	94.11	0.13	94.11	0.11	97.93	0.01	99.09	0.01	100.00	0.01
AA	92.30	0.31	94.11	0.12	94.11	0.11	100.00	0.01	99.34	0.02	100.00	0.01
PU	91.66	0.30	92.30	0.12	92.30	0.12	85.71	0.01	77.78	0.01	83.33	0.01
Av g	94.32	0.23	90.96	0.11	95.13	0.10	92.79	0.01	94.05	0.01	95.83	0.01

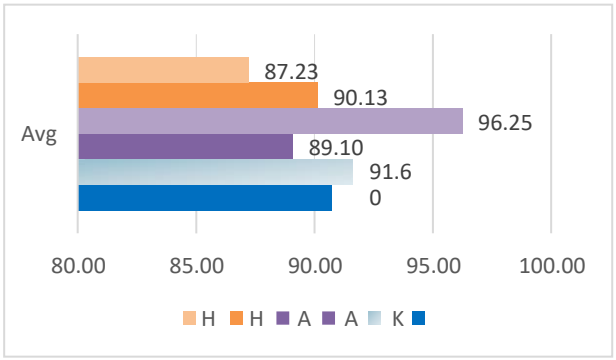
The performance of the Koshur and Awadhi algorithms on a corpus without stopwords and without optimization is shown in Table V and Table VI shows performance without stopwords with optimization.

Fig. 2 shows the average accuracy of two algorithms with stopwords, with and without optimization. Fig. 4 show the result of the Koshur and Awadhi algorithms without stopwords, with and without optimization. The dark shades represent the result of Koshur algorithm and light shades represent the resultof Awadhi algorithm.

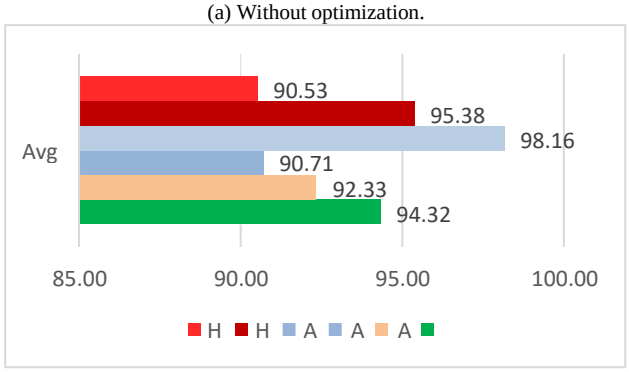
The authors' observation on Fig. 2 is that both algorithms have increased the performance after optimizing the existing code and removal of the extra loops and unwanted instructions from the code. Fig. 3 displays the time taken by the algorithms.

There are four types of approaches considered by the authors the names are WSWWO-With Stopwords With Optimization, WSWWOO-With Stopwords Without Optimization, WOSWWO-Without Stopwords With Optimization, WOSWWOO-Without Stopwords Without Optimization. The graph shows the relation between ET and the approaches. The time taken by the algorithms before optimization was little high than the time taken after optimization. The orange colour line shows with stopwords with optimization, and blue colour represents with stopwords

without optimization.



(a) Without optimization.



(b) With optimization.

Fig. 2. (a,b) Koshur and Awadhi algorithm accuracy with stopwords.

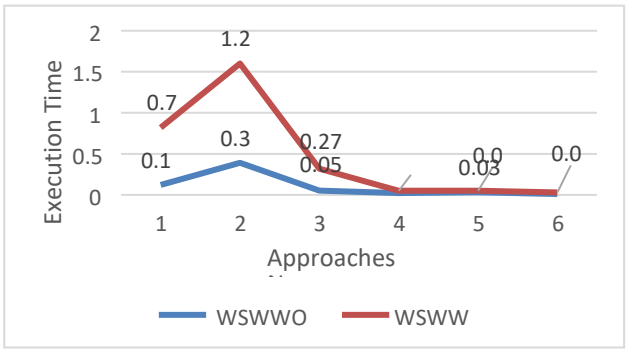
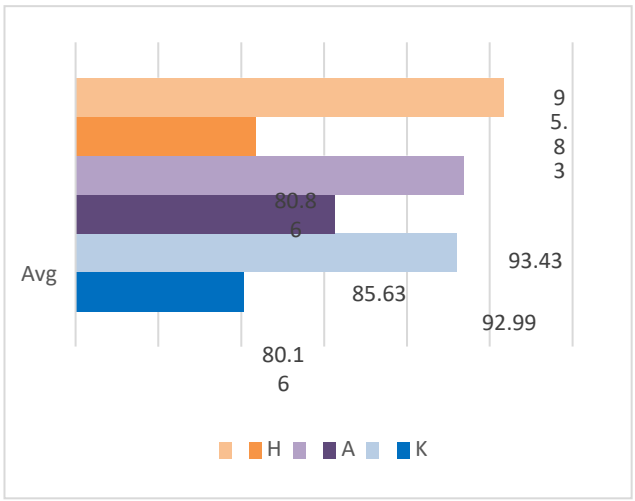


Fig. 3. ET taken by algorithms with stopwords.

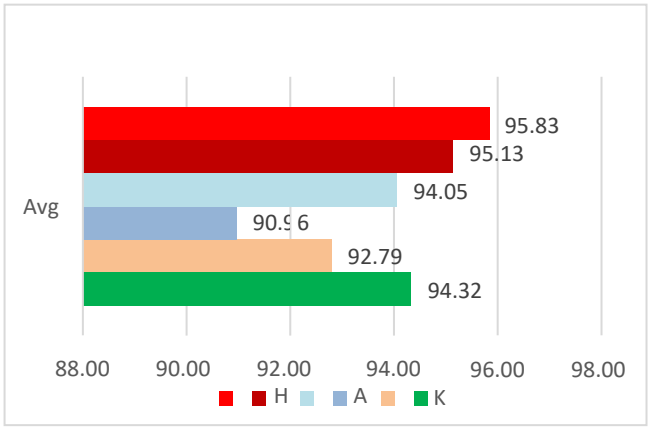
The observation related to Fig. 4 is that after removing the stopwords, the accuracy reduces noticeably. The stopwords are words that do not make semantic sense and are unwanted for processing. As we know that poets use figurative language for writing poems and create a magical effect, but when we remove the stopwords, this effect breaks the rhythm of the poem, and hence code is not able to catch the particular FoS. So our contribution is improvised algorithms that increased the accuracy of the previous algorithms without stopwords up to a great extent. Fig. 5 displays the execution time taken by both the algorithms.

The algorithms after optimization showed the difference in timing. The optimized algorithms were taking less time than the algorithms without optimization.

Fig. 6 show the accuracy of individual FoS named VA, CA, AA, and PU. The blue, pink, and green colours represent Koshur, Awadhi, and Hindi languages. The Koshur algorithm with stopwords was giving 100% accuracy in Hindi and Koshur languages for VA FoS. The same algorithm was giving 100% accuracy in Awadhi in CA FoS. The accuracy for AA and PU FoS in Hindi was 94.11% and 93.30, which was the highest among all three languages. The same algorithm, when used on corpus without stopwords, was giving 100% results in Koshur and Hindi for VA FoS. The CA and AA FoS scored 94.11 % accuracy in Awadhi and Hindi. The 92.30% accuracy was received in Awadhi and Hindi languages for PU FoS.



(a) Without optimization.

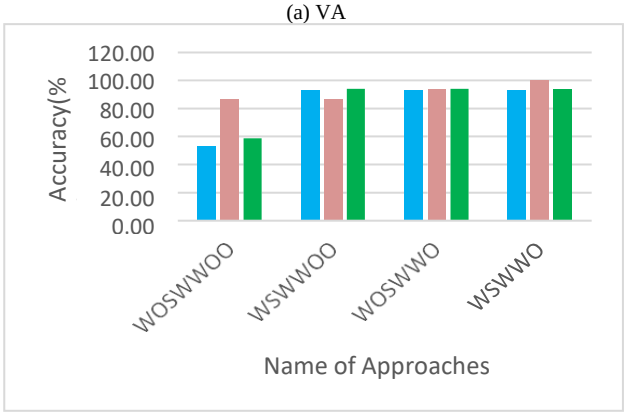


(b) With optimization.

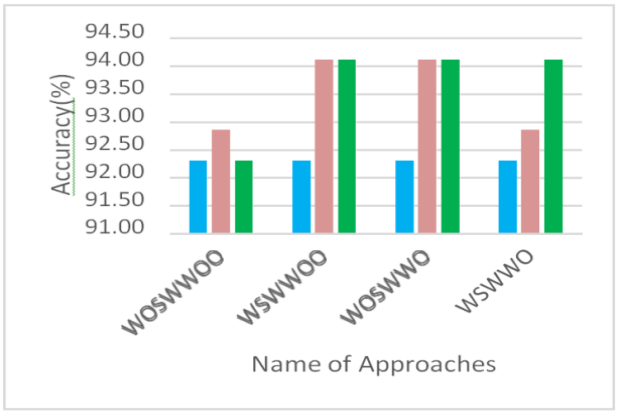
Fig. 4. (a,b) Koshur and Awadhi algorithm accuracy without stopwords.

Fig. 5. ET taken by algorithms without stopwords.

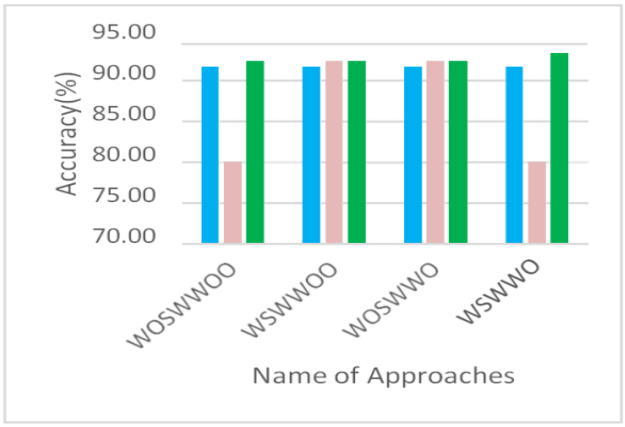
(a) VA



(b) CA

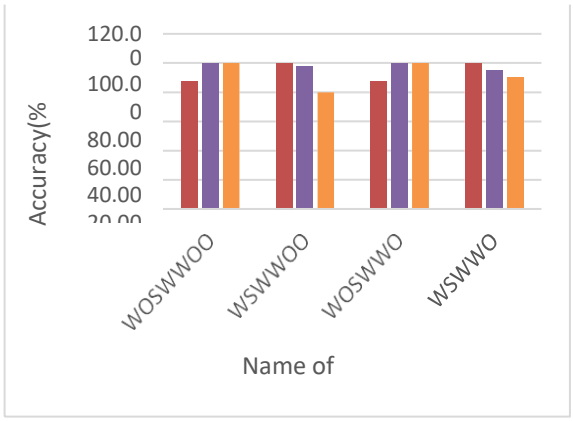


(c) AA

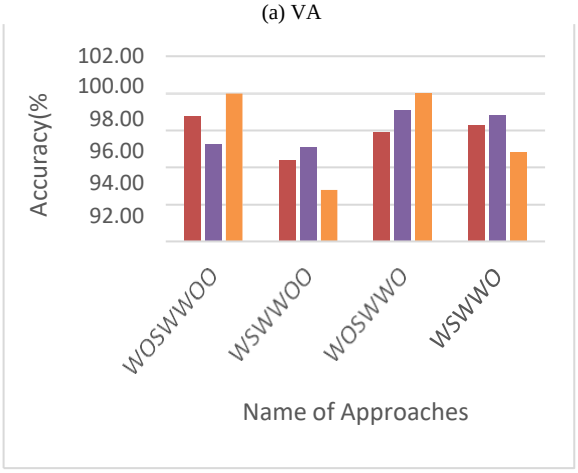


(d) PU

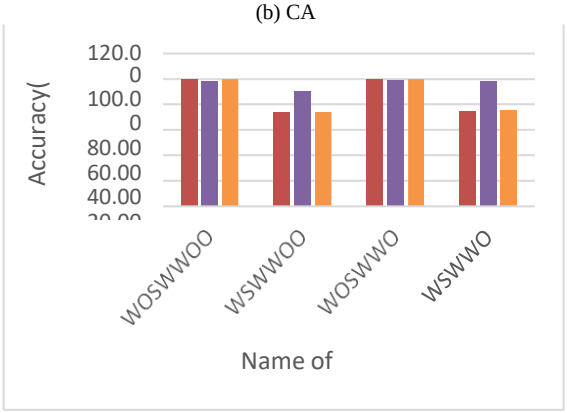
Fig. 6. Application of Koshur algorithm for four FsoS.



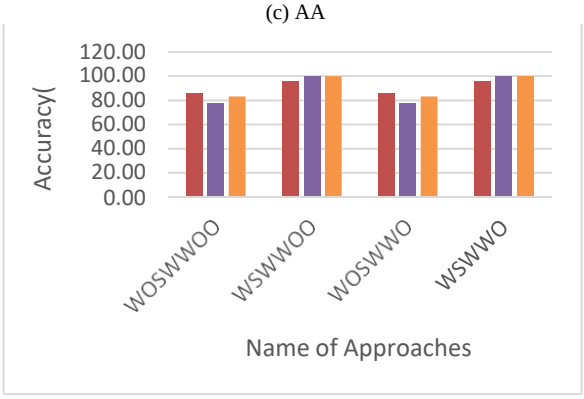
(a) VA



(b) CA



(c) AA



(d) PU

Fig. 7. Application of Awadhi algorithm for four FsoS.

Similarly, Fig. 7 show the individual accuracy of FoS named VA, CA, AA, and PU. The dark red, voilet and orange colours are used for Koshur, Awadhi, and Hindi. While implementing the Awadhi algorithm on corpora with stopwords, the authors recorded 100% accuracy in VA and PU FoS in Koshur, Hindi, and Awadhi languages, respectively. For CA and AA, the accuracy was 98.80 % and 98.63% in Awadhi. The Awadhi algorithm without stopwords provided 100% accuracy in VA in Awadhi and Hindi. The same algorithm gives 100 % results in Hindi for CA and AA FoS. For the Koshur language, 85.71% accuracy was scored in PU FoS.

After analyzing the results the authors derived that the initial implementation of both algorithms performed less accurately on datasets without stopwords compared to datasets that included stopwords. The disparity in accuracy was quite noticeable. However, after applying optimization techniques to both the KA and AA algorithms, there was a significant enhancement in accuracy for both types of datasets – with and without stopwords. The optimization process played a crucial role in enabling the authors to achieve higher accuracy levels for their algorithms, even when stopwords were removed from the text. This underscores the effectiveness of the optimization in improving the algorithms' performance across different types of textual data.

V. Conclusion And Future Scope

The Koshur, Awadhi, and Hindi corpora were used by the authors of this research study to identify four different forms of figures of speech. Two algorithms, KA and AA, were applied by the authors on a corpus with and without stopwords. Authors then improved the algorithms' performance and applied them again on a corpus with and without stopwords.

Both the algorithms before optimization, when implemented on corpora without stopwords, showed low accuracy compared to corpora with stopwords. The difference in the accuracy was quite noticeable. After implementing optimized algorithms (both KA and AA), a significant increase in accuracy can be seen on the corpus without stopwords and with stopwords as well. Optimization helped authors to increase the accuracy of algorithms even after removing the stopwords.

This accomplishment offers a tangible solution to the stark resource limitations these languages confront. The optimized algorithms exhibit a capacity to extract meaningful insights from text, compensating for the lack of extensive linguistic resources. Consequently, the development of these FoS algorithms not only provides a means to navigate the complexities of resource-scarce languages but also contributes significantly to narrowing the existing accuracy gap between corpora with and without stopwords.

In the future, the work will be extended for other Devanagari scripted languages, and the identification of more FoS can be explored and tested.

References

[1] Saini, J. R., & Kaur, J. (2020). Kāvi: An Annotated Corpus of Punjabi Poetry with Emotion Detection Based on 'Navrasa'. Procedia Computer Science, 167,1220-1229. https://doi.org/10.1016/j.procs.2020.03.436

[2] https://prsindia.org/billtrack, "The-jammu-and- kashmir-official-languages-bill-2020", available online, accessed on (07-07-23).

[3] https://www.indiacode.nic.in/, "The Jammu and Kashmir official Language Act, 2020" available online, accessed on (07-07-23).

[4] Why we demand Devanagari as additional/co- script for Kashmiri language, by Dr. R.K. Bhat http://ikashmir.net/rkbhat/script.html

[5] Handoo, J. (1979). Contemporary Kashmiri Poetry. Indian Literature, 22(5), 145-154.

[6] https://www.omniglot.com/writing/awadhi.html, "online encyclopedia of writing systems & Languages", available online, accessed on (10-7-23)

[7] https://archive.org/details/in.ernet.dli.2015.3210 6/page/n23/mode/2up

[8] http://www.ramcharitmanas.org/,"About Ramcharitmanas", available online, accessed on (02- 7-23)

[9] http://aumamen.com/stotra/hanuman-chalisa,"Aumamen", available online, accessed on (11-07-23)

[10] https://kavitakosh.org, "Kavita Kosh", available online, accessed on (11-07-23).

[11] Chopra, A., Prashar, A., & Sain, C. (2013). Natural language processing. International journal of technology enhancements and emerging engineering research, 1(4), 131-134.

[12] Kushwah, K. K., & Joshi, B. K. (2017). Rola: AnEqui-Matrik Chhand of Hindi Poems. Int J Comput Sci Inf Secur (IJCSIS), 15(3).

[13] Audichya, M. K., & Saini, J. R. (2021). Stanza type identification using systematization of versification system of Hindi poetry. International Journal of Advanced Computer Science andApplications, 12(1).

[14] Audichya, M. K., & Saini, J. R. (2021). Towards natural language processing with figures of speech in Hindi poetry. International Journal of Advanced Computer Science And Applications, 12(3).

[15] Bafna, P. B., & Saini, J. R. (2020). AnApplication of Zipf's Law for Prose and Verse Corpora Neutrality for Hindi and Marathi Languages. International Journal of Advanced Computer Science and Applications, 11(3).

[16] Kaur, J., & Saini, J. R. (2018). Automatic classification of Punjabi poetries using poetic features. International Journal of Computational Intelligence Studies, 7(2),124-137.https://doi.org/10.1504/IJCISTUDIES.2018.094901.

[17] Pal, K., & Patel, B. V. (2020). Model for classification of poems in Hindi language based on ras. In Smart Systems and IoT: Innovations in Computing:Proceeding of SSIC 2019 (pp. 655-661). Springer Singapore. https://doi.org/10.1007/978-981-13-8406- 6_62

[18] Lone, N. A., Giri, K. J., & Bashir, R. (2022). Machine Intelligence for Language Translation from Kashmiri to English. Journal of Information & Knowledge Management, 2250074.https://doi.org/10.1142/S0219649222500745.

[19] Mir, T. A., & Lawaye, A. A. (2020, December). Word Sense Disambiguation For Kashmiri Language Using Supervised Machine Learning. In Proceedings of the 17th International Conference on Natural Language Processing (ICON) (pp. 243-245), Indian Institute of Technology Patna, Patna, India. NLP Association of India (NLPAI).

[20] Ahmad, A., & Syam, B. (2014). Kashmir part of speech tagger using CRF. Computer Science, 3(3), 3.

[21] Rasool, N., Nabi, S., & Dar, Y. R. (2019). Machine Translation with UNL: A Study of Kashmiri, Advances in Computer Science and Information Technology (ACSIT), Volume 6, Issue 2; April-June, 2019, pp. 111-115.

[22] Gilkar, S., Nahida, A. H., Nabi, S., & Farooq, V. (2014). Tagging: A Case Study of Kashmiri. Interdisciplinary Journal of Linguistics, 7, 248-261.

[23] Aabid Rashid Wani, Er. Tabish Gulzar and Er. Mamoon Rashid (2015) "Kashmiri Speech Recognition System using Linear Predictive Coding and Artificial Neural Networks" International Journal of Advanced Trends in Computer Applications (IJATCA) Volume 2, Number 6,2015, pp. 20-23.

[24] Thirumuru, R., Gurugubelli, K., & Vuppala, A. K. (2018). Automatic Detection of Palatalized Consonants in Kashmiri. The 6th Intl. Workshop on Spoken Language Technologies for Under-Resourced Languages. https://doi.org/10.21437/SLTU.2018-25

[25] Firdaus, E., Syahit, A. U., & Sukmawan, S. (2023, June). Poetry Writing Method Innovation Using "Atafora" Technique in Indonesian Language Learning. In Proceedings of the 2nd International Conference on Advances in Humanities, Education and Language, ICEL 2022, 07–08 November 2022, Malang, Indonesia.

[26] Malik, N. A., Mustofa, A., & Munir, A. (2023). Unearthing the Figure of Speeches Used in the-R- Album by Rose to Deliver the Messages. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 10(2), 1787-1805.

[27] Krisna, I. P. H. A., Jayantini, I. S. R., & Resen, I. W. (2023). The Types and Meaning of Paradox Found in the Poems of Rudyard Kipling's. Journal of Language and Applied Linguistics, 4(1), 11-16.

[28] Maula, I. K., Sodiq, J., & Nugrahani, D. (2023, March). Analysis of figures of speech as used in songlyrics always remember us this way by ladygaga department of english education faculty of language and arts education universitas pgri semarang. In proceeding of english teaching, literature and linguistics (eternal) conference (Vol. 3, No. 1, pp. 124-134).

[29] Wati, M. L. K., Supriyanto, R. T., & Baehaqie, I. (2023). The Comparative Figure of Speech in a Poetry Collection entitled Ibu Menanak Nasi hingga Matang Usia Kami by Emi Suy. Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia, 12(1), 43-52.

[30] Naaz, K., & Singh, N. K. (2022). Design and development of computational tools for analyzing elements of Hindi poetry. IEEE Access, 10, 97733- 97747.

[31] Mahdi, A. L. G. S., Mohammed, A. A. J., & Alsalmani, M. S. A. (2022). Writing Techniques in the English Love and Lyric Poems. Eurasian Journal of Research, Development and Innovation, 7, 18-21.

[32] Nidi, V., Utami, N. M. V., & Maharani, P. D. (2022). An Analysis of Personification In The Some Selected Poems By Robert Frost. Journal of Humanities, Social Science, Public Administration and Management (HUSOCPUMENT), 2(2), 107-112.

[33] Setiani, R. (2020). Denotative and connotative meaning used in writing poetry. Edukasi Lingua Sastra, 18(2), 85-92.

[34] Hutauruk, B. S. (2019). The Use of Figurative Languages on the Students' Poetry Semester V at FKIP Universitas HKBP Nommensen. Journal of English Language and Culture, 9(2).

[35] Sayakhan, N. I. (2019). The use of personification and apostrophe as facilitators in teaching poetry. Journal of Language Studies, 1(4), 98-106.