Volume 14, No. 5, September-October 2023



International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

(UN) IDENTIFIED LEXEMES: A FIRST GLANCE AT PUNJABI MORPHOLOGICAL ANALYZER

Harjit Singh Dept. of Linguistics and CSTLs Indira Gandhi National Tribal University Amarkantak, India

Abstract: The paper presents a small survey on Punjabi morphological analyzer and generator, which does not give any morphological response to lexemes of Malawi dialect of Punjabi. This analyzer works well with the standard variety only. However, it fails to provide an answer when the input is colloquial. The random choice of approximately, (42 words/lexemes, e.g. ਕਾਤੋਂ/ 'why', ਉੱਈ/ 'thus', ਭੰਨਹੋਣਾ/ 'Unbearable', ਥੁੰਨੀ/ 'eating', ਅੜੀ/ 'stubborn', ਭਨਾਲੀ/ 'breaking', ਮੰਨਲਾ/ 'suppose', ਖਿਜੀ/ 'annoying' etc.) is not trained either for the analyzer nor the generator. They are unidentified and unrecognized. In this regard, the schemes of SGMA (similar guidelines for morphological analyzer) or SMGMA (similar but modified guidelines for morphological analyzer) could adopt for booming the performance of Punjabi morphological analyzer and widening its scope in the future.

Keywords: Un(dentified) lexemes, Punjabi Morphological Analyzer, SGMA

I. INTRODUCTION

Punjabi is a modern Indo-Aryan language with a beautiful story of five rivers (the Jhelum, the Ravi, the Chenab, the Vyas and the Satluj). It covers Eastern Punjab (India), Western Punjab (Pakistan), East Africa, Canada, UK and USA, etc. Gurumukhi is the script, and we usually write it from left to right. It has a (subject-object-verb) word order. However, there are four significant dialects (e.g. Majhi, Malawi, Doabi and Puadhi) of Punjabi, and there is less clarity regarding a standard variety of Punjabi. Majhi considers a standard type, and its area is Amritsar, Tarn Taran Sahib, Pathankot and Gurdaspur in the Eastern Punjab (India) [1]. See map 1.



Above Map 1 mainly shows four major Punjabi dialects (Majhi, Malawi, Doabi and Puadhi). The variations related to phonetics and grammars are available among all these dialects.

After going through the basic details of Punjabi dialects, we are more interested in studying the specific area of Punjabi

computational linguistics rather than theoretical linguistics. In this regard, we discuss the 'Research Centre for Technical Development of Punjabi Language', which

took place in February 2004 at Punjabi University, Patiala (Punjab). It has rick stock of the following online Punjabi resources.

| | Research Centro for Technical Development of Projekt Language. Literature and Calvers, Perpiet University, Prilain # Irst Garmuthi-OCR System with 975, accuracy | Outok Lanka Barmaani Umoole Tyamg Ped Infer Ogla harmanijadi orgiteksa Infer Dys karmanijadi Infer Dys karmanijadi orgi Barmani Infer Byskaminijadi orgi Infer Dys karmanijadi orgi |
|--|--|---|
| Home Page (defaulturight) | shrull and, orfer wel effectives it suchal have in the deal, shrull gener | disc), softwares |
| Abred us (Mend.leipe) | - | |
| Half Merchans (alaffroanciers aspo) | Our Online Resources | |
| Projects (A3Projects anget) | Desirability and an Descention | |
| Our Crokes Tools (or English and a segue) | Weinford Contracting Presidences Weinford Decision (Weinford Statementations) | |
| Purgato Resources on Web (pr.anjor) | Cristine Purgetic Tassifying for Weightner (http://www.isempurgetic.org/interct.asp) | |
| Publications (publications aspn) | A start in Portjeki i Azella & Text (a-book) (25a (gl.teangoargati arginortiente angel) Materia Portjeki (a Bookie) (angel) (25a (gl.teangoargati arginortiente angel) | |
| allocks (Hg:/hearqurist/ org/stocks/books and | Control (control on the control of the control | TUN (ADD, and) (AD) |
| Download Carrier (marrinal aspe) | (http://pi.loangorgati.org/weads//%20Palmenter%20Piammar_Pirat.pdf) | |
| Nexes (Datery Owner, asian) | | |
| ende sienen besate BEEDERE Proje diesen kompungste sorgebetet ausgeb | Dictionaries and Linguistic Resources • man for the Distance of the Distance | And tangen |
| Chip (Masergargeb) (cmm) | Typing and Proofing Tools . - Solte Other, Puper Typin, Petromene on SpetCheer (Hs; Instalukierpury - General University Symp Net (Hs; Instalukierpury) - Replic General States (Hs; Ingel Kenpury) | borpt |
| Particular and the Particular | Machine Transiteration | |
| | Machine Translation - Net to Poper (http://doi.intercontent.org/) - Poper Net and Deputy Annual Sector | |
| | Miscellaneous Resources - Advances Taris Resources to gain to Proget process average and applicability of - Advances Taris Resources and the second second second second second - State Second Seco | n Anada Angaro Manda Angaro |
| | INVE Museum Science for Purplet Learnings English Learning Restore Analysis Hells 147 (B); Restore 1475 (2014) 73:38472 windto English/Changeman Joint | |

Picture 1: Online Punjabi Resource via Research Centre for Technical Development of Punjabi Language, <u>http://www.learnpunjabi.org/default.aspx</u>

Above Picture 1 provides a list of online Punjabi resources that the 'Research Centre has developed for Technical Development of Punjabi Language' (Punjabi University, Patiala). The green rectangle () and arrow represent that we pick up only 'Punjabi Morphological Analyzer' instead of remaining other online resources.

For Punjabi, 'Morphological Analyzer' works to analyse the submitted word classes root and grammatical information. It provides gender, number, and person and case information related to noun word classes and gives tense, aspect and modal information related to verb word classes only in a standard variety of Punjabi. While the 'Morphological Generator' does the opposite tasks, it receives root and grammatical information and then produces the word forms accordingly. Consider few instances.



Picture 2: Perform morphological analysis via Research Centre for Technical Development of Punjabi Language, <u>http://pgc.learnpunjabi.org/#Morph</u>

| Online Punjabi Resources | • | | | el. | | | | | | | | B - | | | |
|--|--|---|----------------------------------|-----------------------------|-----------------------|---------------------|---------------------------------------|--------------------------------------|------------------------------------|---------------------------|-------------------------------|--------------------------------------|---|--|--|
| Grammar Chachar Hornbohutral Anatwis POB Teoret Phrase Churber | | | * 11 = 1 | f H H H H H | 1 1 1 1 1 | 1 2 4 2 1 | H K H H I | K R R R I | - 0 N N | 1 1 1 1 1 | 「日日日日日日日 | ···································· | · * * * * | | |
| The Punjab) grammar checker can detect grammatical errors, resulting from the lack of agreement, order of words in varies of agreement, order of words in varies system include – modifier and nous system include – modifier and nous phrase, order of modifiers in a neuro phrase, order of words in a verb phrase, use | Pu oot word: | ninj na | at | | M | or | pł | 10 | 10 | gi | Part | | marp | hologic rate wa | |
| or contractions sets in the present states, the control of the con | AX553XBNO cond; Tense stional/Cass iD; WordClas 50; WordClas 50; WordClas 50; WordClas 50; WordClas 50; WordClas 50; WordClas | - Wo - Fut - O - O - O - O - O - O - O - O - O - O | raca laure) laure laure | ass) Phi Gen Class | nine) ende ende | Nor Nor Ir: F | erti; n-Pe erni erni Vert | Gen infec nine nine b; G | nder tij Tr tij Nu tij Nu | n X; amb amb amb | Nus er: 1 er: 1 Mass | mber y B Singi Singi | ri Sinu othi G ulari 4 ulari 4 ulari 4 ulari 4 | gular; Dausalit Dase; Dase; mber; uselity | |

Picture 3: Generate word forms via Research Centre for Technical Development of Punjabi Language, http://pgc.learnpunjabi.org/#Morph

The yellow colour () identifies the Punjabi lexemes such as $(\ddot{H}\vec{u}')$ /'Thinking' in picture 1. The following description shows the morphological analysis of this lexeme.

Root = ਸੋਚ

(Tag: VBMAXSS3XBNO; Word Class: Main Verb; Gender: X; Number: Singular; Person: Second; Tense: Future; Phase: Non-Perfect; Transitivity: Both; Causality: None; Inflectional Class: O;)

(Tag: NNFSD; Word Class: Noun; Gender: Feminine; Number: Singular; Case: Direct ;)

(Tag: NNFSO; Word Class: Noun; Gender: Feminine; Number: Singular; Case: Oblique ;)

The black colour () recognizes the morphological generator that applies to the same lexeme ' $\vec{H}\vec{\upsilon}$ ' /'Thinking' and produces different word forms. Look at the following results. $\vec{H}\vec{\upsilon}$

(Tag: VBMAXSS3XBNO; Word Class: Main Verb; Gender: X; Number: Singular; Person: Second; Tense: Future; Phase: Non-Perfect; Transitivity: Both; Causality: None; Inflectional Class: O;)

(Tag: NNFSD; Word Class: Noun; Gender: Feminine; Number: Singular; Case: Direct ;)

(Tag: NNFSO; Word Class: Noun; Gender: Feminine; Number: Singular; Case: Oblique ;)

ਸੋਚਣਾ

(Tag: VBMAMSXXXBNNA; Word Class: Main Verb; Gender: Masculine; Number: Singular; Person: X; Tense: -; Phase: Non-Perfect; Transitivity: Both; Causality: None; Inflectional Class: N! A ;)

We see that ਸੋਚ/Thinking with this tag: VBMAXSS3XBNO is a primary root verb. On the other hand, ਸੋਚਣਾ/Thinking comes

with another tag like VBMAMSXXXBNNA. But, many colloquial lexemes in Punjabi are unresolved morphologically. We capture those unanalyzed or unknown cases and suggest a procedure to incorporate them with $\overline{H}\overline{U}$ /Thinking like lexemes.

Briefly, we have a total of six parts in this paper. In the first part, we discuss few virtual instances of lexemes in a standard variety of Punjabi via Punjabi Morphological Analyzer. At the same time, we see previous literature related to the morphological analyzer in other languages in the second part. Next, we elaborate on the aims and discuss the scope of the paper in the third part. In the fourth part, we mainly present 'unknown' or 'unrecognized' cases in the colloquial Punjabi. The fifth part discusses the procedure to resolve 'unknown' or 'unrecognized' lexemes. In the sixth part, we discuss the conclusions and future of the works.

II. RELATED WORKS

In 2008, a paradigm approach was applied to Hindi words to study their morphological structure [2]. Another language Kokborok (which belongs to Tripura and nearby areas of Bangladesh), has shown 80% correct results when the morphological analyzer trained around 56732 Kokborok words [3]. For Malayalam morphological analyzer, the Memory-Based Language Processing (MBLP) algorithm was adopted. The TiMBL (Tilburg Memory-based Learner) was adopted in this analyzer after annotation for training purposes [4]. Later, the Multi-Task Deep Morphological analyzer (MT-DMA) also morphologically trained Hindi and Urdu [5].

On the other hand, some scholars like Sivan have tried to compare the existing (corpus-based, rule-based, stemmer based, paradigm based, finite state automata and transducer-based) morphological analyzers and notice the reliability for Malayalam [6]. But it points out that the paradigm approach of morphological analyzer increases 84.50%, 81.50% and 80.50% accuracy for nouns, verbs and adjectives, respectively [7].

III. OBJECTIVES AND SCOPE

• To study only Punjabi Morphological Analyzer such as (online Punjabi resources)

• To find out 'unknown lexemes' related to colloquial Punjabi

• To suggest the procedure for resolving the unknown lexemes

To conclude the work and plans in future

IV. UNKNOWN/UNIDENTIFIED LEXEMES

There are two parts (morphological analysis and generate word forms) in the Punjabi morphological analyzer. Earlier, we notice that it works only for the standard Punjabi. After input of the lexeme, it separates the root and then provides the grammatical information, as seen in Picture 2. However, we argue that it does not identify any lexeme that belongs to a colloquial variety (e.g. Malawi) of Punjabi. Moreover, it also does not provide any answer if the input is so standard (like $\forall Z \delta I / Patna, \ \vec{U} \vec{C} \delta I / Stop, \ \vec{S} \vec{C} \vec{\delta} I / Stable, \ \vec{d} \vec{C} \vec{\delta} I / Memorization etc.)$ lexemes from the standard Punjabi. The following Table 1 shows the list of lexemes in Malawi that are possibly unknown/unidentified lexemes.

| | Table 1: Lis | st of unknov | vn lexemes in Malaw | vi (Punjabi) | 30 | ਸਖਾਏ | ਸਖਾਏ | Verb | Teach |
|-----------|----------------|----------------|------------------------|----------------|-----------------------|---|---|--|------------------------------------|
| Sr. No | Lexemes of | Root | Word class | English | 31 | ਛਾਬਾਛੀ | ਛਾਬਾਛ+ | Adjective | Well done |
| 1 | Malwi ਕਾਤੋਂ | ਕਾਤੋਂ | Adverb | Why | 32 | ਮੈਤੋ | ਈ ਮੈਤੋ | Pronoun | First person |
| 2 | ਉਂਈ | ਉਂਈ | Adverb | Thus | 33 | ਕੋਲੇ | ਕੋਲੇ | Post-position | Along |
| 3 | ਭੁੰਨਹੋਣਾ | ਭੁੰਨ+ਹੋਣਾ | Adjective | Unfavourable | 34 | ਐਕਨ | ਐਕਨ | Adverb | Likewise |
| 4 | ਬੁੰਨੀ | ਬੁੰਨ+ ਈ | Noun/Adjective | Eating | 35 | ਨਿਆਨਾ | ਨਿਆਨਾ | Noun | Small child |
| 5 | ਅੜੀ | ਅੜ+ਈ | Adjective | Stubborn | 36 | ਘੁਰੀਂ | ਘੁਰ+ ਈਂ | Adjective | Scary |
| 6 | ਭਨਾਲੀ | ਭਨਾਲੀ | Verb | Breaking | 37 | ਟਿਕਨਾ | ਟਿਕ+ਨਾ | Noun | Sitting |
| 7 | ਦਈ | ਦਈ | Adjective | Giving | 38 | ਬੁੱਜ | ਬੁੱਜ | Adjective | Foolish |
| 8 | ਅੜਬ | ਅੜ+ਬ | Adjective | Stubborn | 39 | ਨੀਚੋ | ਨੀਚ+ ਓ | Adverb/Adjective | Downstairs |
| 9 | ਮੰਨਲਾ | ਮੰਨਲਾ | Verb | Suppose | 40 | ਓਏ | ਓਏ | Noun | Hello |
| 10 | ਜਾਈਦਾ | ਜਾਈਦਾ | Noun | Going | 41 | ਬਚੀ | ਬਚੀ | Noun | Kid |
| 11 | ਰੁੜ | ਰੁੜ | Adjective | Falling | 42 | ਨਾਨੂੰ | ਨਾ+ਨੂੰ | Negation | Not |
| 12 | ਗਏਆ | ਗਏਆ | Verb | Went | The | total 42 in | nstances in | Table 1 are highly | unknown and |
| 13 14 | ਛੱਡੀਂ ਧੱਕਮ | ਛੱਡ+ ਈ ਧੱਕਮ | Adjective Verb/Noun | Giving Push | unre sing are 1 | cognized ca le instance ike as: | ases for Pun has yet bee | Jabi Morphological A n resolved [8]. The f | Analyzer. Not a ollowing cases |
| 15 | ਪੋਲਿਆ | ਪੋਲਿਆ | Noun | Polya | | | | Research Centre for Punjabi Language Technology Punjabi University, Patala | |
| 16 | ਤਾਵਾ | ਤਾਵਾ | Noun | ? | | Online Pu | njabi Resources | | |
| 17 | ਖਿਜੀ | ਖਿਜੀ | Adjective | Annoying | | Grammar Checker M POS Tagger P | lorphological Analysis hrase Chunker | ਚ ਨਾ ਇਸ ਹੁਕ ਬਰੀ ਘੁੜ ਚੁੱਝ ਚ ਛ ਜ ਝ ਵ ਟ ਨ ਡ ਢ ਣ ਓ ਨ ਤ ਬ ਦ ਧ ਨ ਪੁਛ ਬ ਲ ਮ ਨਾ ਨੇ ਯ ਰ ਲ ਵ ਤ ਸ ਮ ਨਾ ਜ ਛ ਨੇ ਨੇ | |
| 18 | ਮਗ਼ਜ | ਮਗ਼ਜ | Noun | Brain | | The Punjabi grammar of and suggest rectification grammatical errors, resu of agreement, order of phrases etc., in literary si | checker can detect is for a number of ulting from the lack f words in various tyle Punjabi texts. | Punjabi Morphological A | nalysis |
| 19 | ਚਮਲਾਈ | ਚਮਲਾਈ | Adjective | Excited | | The grammatical errors system include – m agreement, subject/ot agreement, order of m phrase, order of words in of contractions etc. In it | s covered by this Specify a to odifier and noun bject and verb lodifiers in a noun in a verb phrase, use is present state, the | 양면 analysis G | orphological enerate word forms |
| 20 | ਰਿੰਗੀਂ | ਰਿੰਗੀਂ | Adjective | Crying | | complex and compound a User can type the text onscreen keyboard or keyboard layout of Anno also paste Unicode text in | by clicking on the by typing in the ollipi font. User can n the text box. | word. | |
| 21 | ਨੀ | ਨੀ | ? | ? | | Developed by: Mandeep Singh Dr Gurpreet Singh Leh Advanced Centre for Te Devlopment of Punjabi Language, Literature an | al schnical sd Culture | | |
| 22 | ਬੁੱਥੀ | ਬੁੱਥੀ | Noun | Face | | | | | |
| 23 | ਚੱਜਦੀ | ਚੱਜ+ਦੀ | Adverb/Adjective | Right | Pictu | re 4: Unknow Centre fo | n case of ਉਂਈ/ or Technical D | Thus (morphological analy evelopment of Punjabi La | ysis) via Research nguage, |
| 24 | ਬੁੱਗੇ | ਬੁੱਗੇ | Noun | Child | | | http://pgc.le | arnpunjabi.org/#Morph | |
| 25 | ਰੜਿਆ | ਰੜ+ | Adjective | Angry | | | | | |
| 26 | ਕੁਰੇ | ।ੲਅਾ ਕੁਰੇ | Noun | Title | | | | | |
| 27 | ਸਿਹਾਂ | ਸਿਹਾਂ | Noun | Title | | | | | |
| 28 | ਗਹਾਂ | ਗਹਾਂ | Adverb | Ahead | | | | | |
| 29 | ਨਨਦਾਂ | ਨਨਦਾਂ | Noun | Sister-in-law | | | | | |



Picture 5: Unknown case of ਉੱਈ/Thus (Generate word forms) via Research Centre for Technical Development of Punjabi Language, <u>http://pgc.learnpunjabi.org/#Morph</u>

| Online Punjabi Resources | | | R | search | Cert | tor P | | | Tech | ncioz | • | |
|--|----|----------|---------|--------|------|------------|------|-----|--------------|--------------|-------------|-----------------|
| | 8 | т | + | f | 7 | 3 1 | 6 | * | 2 | 21 | | |
| | 8 | 8 | ж | E | Ħ | ਹ ਰ | ਰ ਖ | ਗ | ъſ | 22 | R | 8 |
| Grammar Checker, Marshelenical Applicate | × | ਚ | z | न | 8 | 2 2 | | 15 | E | 3 | G | ж |
| POS Tagger Phrase Chunker | | न | H | | π | × 1 | 1 2 | R | 8 | H | ыт | 2 |
| | | 77 | 7 | x | 7 | H 1 | TH | जा | 7 | 77 | 37 | ÷. |
| | | - | 1 | - | ÷. | mit | a ne | 301 | - | T | for | 5 |
| and suggest rectifications for a number of grammatical errors, resulting from the lack of agreement, order of words in various | Pu | Inj | at | oi I | Mo | orp | ho | lo | gio | al | A | nal |
| and suggest restifications for a number of grammatical errors, resulting from the lack of agreement, order of words in various phrases stee, in literary style Punjak basts, system include - modifier and noun agreement, subject/object and verb agreement, order of modifiers in a noun phrase, order of words in a verb phrase, use | Pu | nj ਐব | at T | Di | Mo | brp | ho | lo | e F analy | erfo (sis | A m O | morph Genera |

Picture 6: Unknown case of ਐਕਨ/Likewise (morphological analysis) via Research Centre for Technical Development of Punjabi Language, <u>http://pgc.learnpunjabi.org/#Morph</u>

| Online Punjabi Resources | | | Reser | rch Ce | ntre for Angel | Purje | di Lar nah, F | | Techno | | |
|---|----------------|------------|-------|--------|-------------------|-------|------------------|----|---------------|-----------------|--|
| | | T | † f | 1 | 10 | 3 | 10 | 5 | 9 | | |
| | 2 | ष्ठि | ਅ ਬ | ਸ | ਹ | ਕ | ਖ | ਗ | ม 5 | e e | B |
| Grammar Checker Morphological Analysis | 8 | ਚ | 8 7 | 8 | स | 2 | σ | ड | च ट | 6 | 201 |
| POS Tagger Phrase Chunker | 2 | 3 | म स | ч | চ | ਪ | ढ | в | ਭ ਮ | N N | ' ਅਤੇ |
| | 9 | ज | ਰ ਲ | द द | 3 | ਸ਼ | Ħ | ज | ਜ ਫ਼ | พี | हे |
| | 10 | | | | ਖਾਲ | ो वां | | | 5 | 5 ਇ | ਈ |
| nd suggest rectifications for a number of rammatical errors, resulting from the lack f agreement, order of words in various hrases etc., in literary style Punjabi texts. Specify a ror | Pu | INJ | abi | M | or | ph | ol | og | pica | al A | Analysis |
| nd suggest rectifications for a number of rammatical errors, resulting from the lack f agreement, order of words in various he granmatical errors covered by this specification of the second second second ystem include - modifier and noun greement, order of modifiers in a noun greement, order of modifiers in a noun f contractions etc. In its present state, the | Pu ot word: | nj भेवत | abi | M | or | ph | ol | |) Pe nalys | rform is Cle | Analysis morphological Generate word for |

Picture 6: Unknown case of ਐਕਨ/Likewise (morphological analysis) via Research Centre for Technical Development of Punjabi Language, <u>http://pgc.learnpunjabi.org/#Morph</u>



Picture 7: Unknown case of औवत/Likewise (Generate word forms) via Research Centre for Technical Development of Punjabi Language, http://pgc.learnpunjabi.org/#Morph

| Online Punjabi Resources | | | Res | earch | Centre | for Pur | | | e Tech | nolog) | + | |
|---|---------|------------|-----|-------|--------|---------|----|-----|--------|--------------|------|--------------------------------|
| | 8 | т | ÷ | f | 1 | 3 | CI | 3 | 011 | 9 | | |
| | 8 | 8 | ж | B | ਸਫ | ਹ ਕ | ਖ | ਗ | খ | W I | 8 | 81 |
| Grammar Checker Morphological Analysis | 8 | ਚ | 2 | ਜ | ष्ठ र | 5 1 | δ | 3 | ਚ | ह । | 6 | m |
| POS Tagger Phrase Chunker | 2 | 3 | Ħ | ਦ | य र | 5 थ | ਫ | я | ਭ | ਮ 🕨 | NT : | R |
| | 9 | ज | ਰ | ਲ | र ३ | য় ম | Ħ | ,ता | ਜ਼ | ਡ 1 | Ŵ | 8 |
| | 20 | | | | SL. | ਲੀ ਬ | ŧ | | | ਲੀ | ਇ : | ਈ |
| nd suggest rectifications for a number of rammatical errors, resulting from the lack f agreement, order of words in various hrases etc., in literary style Punjabl texts. Specify a word: he grammatical errors covered by this Specify a word: | Pu | nj | ab | i I | Мо | rpl | 10 | lo | gic | al | A | naly |
| nd suggest rectrications for a number of rammatical errors, resulting from the lack if agreement, order of words in various present and the supervised by this Specify a word: system include - prodifier and noun greement, subject/object and verb trass, order of words in a verb in ass, use from the subject object and set of the subject object for ontractions etc. In its present state, the performance of the subject object and the subject object for ontractions etc. In its present state, the | Pu C | nj fafi | ab | cal / | Mo | rpl | 10 | | e p | erfor sis | | naly: norpholog Senerate |

Picture 8: Unknown case of पिनी/Annoying (morphological analysis) via Research Centre for Technical Development of Punjabi Language, http://pgc.learnpunjabi.org/#Morph

| Online Punjabi Resource | es | | Reser | ch Ca | | Panja | bi Lar | | Tech | nclog | | |
|---|---|----------------------|-------|-------|----|-------|--------|---|---------------------|--------------|------|-----------------|
| | 1 | r | † f | 7 | 3 | 3 | 8 | 3 | 10 | 2 | | |
| | 0 | 6 | M E | ਸ | ਹ | व | ਖ | ਗ | ш | 10 | ĝ | <u>ğ</u> |
| Grammar Checker Morphological Analysis | 10 | ਚ | R H | 8 | य | S | 2 | 3 | स | 3 | G | m |
| Pos lagger Pirase cristiker | - | З | ष र | ч | δ | ч | a. | я | 8 | н | m | m |
| | - | <i>ы</i> | 3 8 | .4 | 3 | B | Ħ | ज | H | - | M | 8 |
| he Punjabi grammar checker can detect nd suggest rectifications for a number of rammatical errors, resulting from the lack ' agreement, order of words in various vrases etc., in literary style Punjabi texts. Specify | Pu | nj | abi | м | or | ph | ol | 0 | gic | al | | Anal |
| he Punjabi grammar checker can detect di suggest recifications for a number of rammatical errors, resulting from the lack agreement, order of words in various vases etc., in literary style Punjabi becks. Specify sem include – modifier and noun preement, subject/object and verb preement, subject/object and avoin | Pu a root word: | nj | abi | M | or | ph | ol | | gic O P analy | erfo | | morph |
| he Purjabi grammar checker can detect ammatical entrols, resulting from the lack ammatical entrols, resulting from the lack here and the second second second second here and the second second second second parameters that the second second second parameters of the second second second second second second se | Pu a root word: Genera m word. | nj विज्ञी te W | abi | M | or | ph | ol | | gic O P analy | erfc rsis | Cles | morph Genera |

Picture 9: Unknown case of धिनी/Annoying (Generate word forms) via Research Centre for Technical Development of Punjabi Language, http://pgc.learnpunjabi.org/#Morph.

From (Picture 4 to Picture 9) related to $\frac{\partial}{\partial t}$ Thus, $\frac{\partial}{\partial a \delta}$ /Likewise, and $\frac{\partial}{\partial t}$ Annoying lexemes are absolutely 'unknown'/ 'unidentified' cases of Malawi dialect. After performing the morphological analysis, the light blue rectangle () shows 'no results' available for such lexemes. Similarly, the green rectangle () also indicates that there is no chance to generate any word form based on these lexemes.

V. PROPOSAL FOR UNKNOWN LEXEMES

In practice, unknown lexemes are nothing but a collection of the Punjabi dialect (e.g. Malawi). Phonetics is a significant factor of the total 42 lexemes, and they are orthographically finding less attractive. Even it is a challenging task to define them linguistically also. So far, we are not sure whether any work (in on-line mode) is available related to the dialect or not. But regarding the unknown lexemes, we approach the following figure 1.



Fig 1. Steps of filtering the unknown lexemes

Word/lexeme such as 'ਭੁੰਨਹੋਣਾ' / 'Unfavorable' breaks down into 'ਭੁੰਨ+ਹੋਣਾ' two root parts that mean both are independent forms. The ending – \mathcal{M}^{T} of ਹੋਣਾ represents singular, masculine/feminine, third-person pronoun etc. information. When they come all together, they function as a compound. Thus, we could quickly analyze the remaining lexemes also. For more details, please go to Table 1 and see the root column carefully.

After the analysis, our next task is to provide a tag to each word class accordingly. The whole procedure defines as,





Following the description, we can resolve almost all the unknown cases at this initial level. Then, SGMA (similar guidelines for morphological analyzer) might be adopted for tagging purposes. See below some instances.

'ਭੁੰਨਹੋਣਾ'/ 'Unbearable' [Compound word]

Tag: ADVXSS3P; Word Class: Adverb, Gender: Male, Number: Singular, Person: Third, Tense: Past, Phase: Perfect etc.

'냄нी'/ 'Annoying' [Complex word]

Tag: ADJXM/FS3P; Word Class: Adjective, Gender: Male/Female, Number: Singular, Person: Third, Tense: Past, Phase: Perfect etc.

'ਕੋਲੇ'/ 'Along' [Complex word]

Tag: PPXM/FS3P; Word class: Post-position, Gender: Male/Female, Number: Singular, Person: Third, Tense: Present, Phase: Perfect etc.

We have successfully but manually annotated the above instances. We used the tag-set of the standard Punjabi. In this way, we suggest that the existing tag-set of 'Punjabi Morphological Analyzer' could apply to all 42 unknown lexemes in future.

VI. CONCLUSION

Like other languages, Punjabi is also rich with many online computational resources. 'Punjabi morphological analyzer' is an essential resource freely available on the web (<u>http://pgc.learnpunjabi.org/#Morph</u>) to analyze the words morphologically. But it is not trained to diagnose the dialectal

vocabulary so far. Malawi is a good example here, and its total of 42 words is unknown. In this regard, we suggest an (SGMA) scheme to minimize the rate of unknown results in the future.

VII. REFERENCES

- [1] Bhatia, K.T. (2013). Punjabi: Cognitive-Descriptive Grammar. Routledge: Taylor and Francis Group. New York, USA.
- [2] Debbarma, K. et al. (2012). Morphological Analyzer for Kokborok. In Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing (SANLP). COLING, pages 41-52, Mumbai, India.
- [3] Goyal, V. and Lehal, S.G. (2008). Hindi Morphological Analyzer and Generator. In Proceedings of the 1st International Conference on Emerging Trends in Engineering and Technology, (ICETET), pages 1156-1159, IEEE, Nagpur, India,. https://doi.org/10.1109/ICETET.2008.11.

- [4] Jha, S. et al. (2018). Multi Task Deep Morphological Analyzer: Context Aware Neural Joint Morphological Tagging and Lemma Prediction. arXiv preprint arXiv:1811.08619.
- [5] Nisha, M. et al. (2015). Malayalam morphological analysis using MBLP approach. In Proceedings of the International Conference on Soft-Computing and Network Security (ICSNS). IEEE, Coimbatore, India. https://doi.org/10.1109/ICSNS.2015.7292387.
- [6] Research Centre for Technical Development of Punjabi Language. (2004). Our online tools: Punjabi Morphological Analyzer. Punjabi University, Patiala. http://pgc.learnpunjabi.org/#Morph.
- [7] Shah, N.D. and Bhadka, H. (2019). Paradigm Based Morphological Analyzer for Gujarati Language. In Proceedings of the Intelligent Communication, Control and Devices (ICICCD). Advances in Intelligent Systems and Computing, vol. 989. Springer, Singapore, pages 469-481. https://doi.org/10.1007/978-981-13-8618-3_50.
- [8] Sivan, R. (2019). Study on Morphological Analyzer and Generator for Malayalam. J. IJESI, 8 3:73-77.