



## Implementation of DJ Rule Based Algorithm for Dhuni- Vishleshan of Compound Punjabi Words

Deepjot Kaur\*, Navjot Kaur

Department of Computer Science & Engineering  
Sri Guru Granth Sahib World University  
Fatehgarh Sahib, Punjab, India.

**Abstract:-** Dhuni-vishleshan describes the process by which one word is broken into many words. It is the new software for Punjabi language. Various rules can be made and after that we can implement the words according to their rules. In Punjabi, words are a sequence of characters. There is a little amount of work is completed in this area. A word can be of two types-simple and compound. A simple word consists of roots. A compound word is also called as co-joined word can be broken up into two or more words. The problem to which this paper is concerned, is breaking up of Punjabi compound words into constituent words. In this paper, the rules for breaking the compound words into simple words have been applied. The problem of this paper is to break the compound word into constituent words with the help of rules of 'dhuni-vishleshan' in Punjabi.

**Keywords:-** compound words, DJ rule based algorithm, dhuni-vishleshan, phonetics, gurmukhi script.

### I. Introduction

A) Phonetics: Phonetics is the study of speech sounds of humans that appear in all human languages to represent the meanings. In phonetics we deal with different sounds neither letters. The task of phonetics is to provide brief of speech. Phonetics plays a important role in improving our communication. Alphabets and words are spelled correctly that is must [1]. Eg. A child cries and informs it mother that it is hungry. In this condition no language is used. For communication language can spoken or written. In this case sound matters. Mostly sounds are produced by air-stream from lungs through any other speech organs. It is the root of the speech sounds.

B) History of Punjabi Language: Punjabi sometimes spelled Panjabi, belong to the Indic group of the Indo-European family of Languages. Punjabi is the tonal language. Tonal being that it differentiate the words by tones [2]. Punjabi language is used in both parts of Punjab in India and Pakistan. In India and Pakistan the written standard for Punjabi is known as Majhi that is called after the Majha region of Punjab. This script was created by Guru Angad Dev Ji. This language is the mother language of more than 100 million people of Pakistan, India, Canada and America. In India it is the official Language of Punjab state, and is additionally spoken within the neighboring states of Haryana and Himachal Pradesh. The Punjabi language is closely connected with the Sikh religion. Its alphabet, recognized as Gurmukhi, was the vehicle for recording the teachings of the Sikh gurus. It was invented by the second of the gurus within the 16<sup>th</sup> century. The word "Gurmukhi" means "Guru's mouth." Gurmukhi script is used for Punjabi language and it is the 11<sup>th</sup> widely spoken language in the world. Almost 100 million people speak different accent of this language as their first language.

- 1) **Gurmukhi Consonants:** The Gurmukhi script has thirty five akhar or consonants, a twin of the Punjabi alphabet as well as 3 vowel and thirty two consonants. Each character represents a phonetic sound. The alphabetical order of the Gurmukhi script area unit classified to make a grid of 5 horizontal and 7 vertical rows. Some characters have a nasal sound [3].
- 2) **Gurmukhi Vowels:** In Punjabi language letters are joined by a line at the top. In this there is no concept of upper and lower case letters. The gurmukhi script can be separate into three zones i.e. upper, middle and lower. There are ten vowels, three semi-vowels and three half-characters are used in Punjabi language [4]. In spoken language a vowel could be a sound that is pronounced with associate vocal tract such as teeth, lips, tongue. Vowels are the affecting class of sound in any language. They play a significant role within the pronunciation of any words.

### II. Proposed Algorithm

Phonetics is the study of speech sounds of humans that appear in all human languages to represent the meanings. The work has been done in the area of English and similar languages. Punjabi is the 11<sup>th</sup> widely spoken language. There is the very little amount of work is completed in this field. Developing programs that understand a natural language is a difficult task. They contain an infinity of various sentences. The problem to which this paper is concerned, is breaking up of Punjabi compound

words into constituent words. Eg. ਰਾਮ → ਰ+ ਆ +ਮ . Sometimes a person cannot pronounce difficult word so it's a easy way to pronounce by separating the words by applying several rules. Several rules can be made according to Punjabi laga ੈ, ੈ, ਾ, ਿ, ੀ, ੁ, ੋ, ੰ, ੇ, ੁ, ੍ and Punjabi vowels are also used like ਅ ਆ ਇ ਏ ਓ ਊ ਊ ਏ ਐ ਓ ਐ . On this various rules will be made and after that we can implement the words according to rules. Some examples of 'dhunivisheshan' of Punjabi Compound Words are:

Table I: Compound Words with their Outputs (Dhuni-vishleshan)

Compound word	Dhuni-vishleshan
ਸਥਿਰੀਕਰਣ	ਸ ਥ ਇ ਰ ਏ ਕ ਰ ਣ
ਯੋਯੋਜਨਾ	ਯ ਓ ਯ ਏ ਯ ਅ ਜ ਏ
ਕੁਚੈਲਾ	ਕ ਊ ਚ ਐ ਲ ਆ

**Algorithm:-** Dhuni-vishleshan is a recently developed software for Punjabi language. It is a application that is developed in .net. The algorithm used for the implementation of this module is the DJ Rule Based Algorithm.

**Step 1:** Load data from database.

**Step 2:** Select the word from the database or whether enter manually.

**Step 3:** Splitting the string into character by character.

**Step 4:** Now, comparing the characters:-

- a) If character = ਾ  
Replace it with ਆ
- b) Else If character = ਿ  
Replace it with ਇ
- c) Else If character = ੀ  
Replace it with ਏ
- d) Else If character = ੁ  
Replace it with ਊ
- e) Else If character = ੋ  
Replace it with ਊ
- f) Else If character = ੇ  
Replace it with ਏ
- g) Else If character = ੈ  
Replace it with ਐ
- h) Else If character = ੋ  
Replace it with ਓ
- i) Else If character = ੰ  
Replace it with ਐ

- j) Else If character = ਂ  
Replace it with ਨ
- k) Else If character = ੀ  
Replace it with ਨ
- l) Else If character = ੱ  
Replace it with ਅ
- m) Else If character = ੜ  
Replace it with ਰ
- n) Else If character = ੜੁ  
Replace it with ਹ
- o) Else character = ੜੁਵ  
Replace it with ਵ

**Step 5:** Concatenate the final character for final output

### III Experiment & Result

Dhuni-vishleshan is a recently developed software for Punjabi language. It is a application that is developed in .net. With in which the work is done on MS Excess at the back-end tool and front-end tool is .net. The algorithm used for the implementation of these module is the Rule Based Algorithm. Accuracy is the significant issue to be examined. So, to measure the accuracy of our algorithm we implement experiments on number of different words. Whenever the application is started, the window shown in figure 1 will appear which contains the text area, where the user can enter the text. In this, we can choose a word from the database or enter a word manually.



Fig.1: Snapshot of main screen.

In the following snapshot Fig.2 shows the working of Dhuni-vishleshan. First the user will choose the word from database or whether enter manually. After that they will get the output on clicking button ਆਉਟ ਪੁਟ ਦਿਉ. Fig. 3 shows its word to sound rule.





Fig. 4: Dhuni-vishleshan for the word “ਕਿਰਾ-ਸੁਣਿਆ”

If the user choose the word ਕਿਰਾ-ਸੁਣਿਆ from the database then it will show the output as:

ਕ ਇ ਰ ਆ - ਸ ਉ ਣ ਇ ਆ and it gives correct output.

We perform testing on different words corresponding to our algorithm. To measure the accuracy of our algorithm. We perform testing on 24231 words. After testing we obtain 99.9% accuracy.

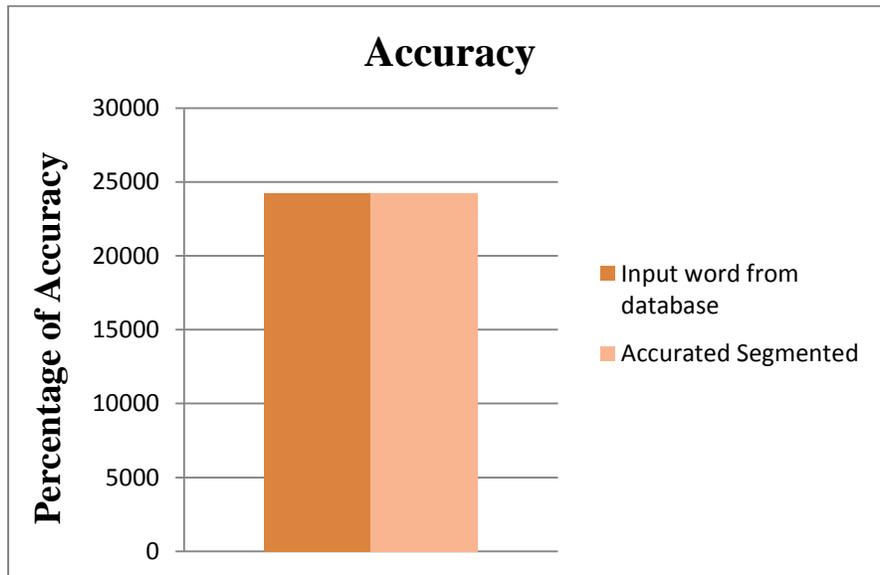


Fig.5: Histogram showing the accuracy for Punjabi Word

**Results:-** We have tested the system by first giving the input from database that contain approximately twenty four thousand words where our system has given no error. After that we have entered the words manually that also gives no error. So we can say that our system has good accuracy. Following is the part of the document.

Word	Output	Comment
ਅਭਿਭਾਵਕ	ਅ ਭ ਇ ਭ ਆ ਵ ਕ	Correct
ਅਣਹੋਂਦ	ਅ ਣ ਹ ਓ ਨ ਦ	Correct
ਸਤਰੰਗਾ	ਸ ਤ ਰ ਨ ਗ ਆ	Correct
ਤਬੋਲਾ	ਤ ਬ ਏ ਲ ਆ	Correct
ਲੋਂਗ	ਲ ਔ ਨ ਗ	Correct
ਅਭਿਜੀਤ	ਅ ਭ ਇ ਜ ਈ ਤ	Correct
ਸਿਧਾਂਤ	ਸ ਇ ਧ ਆ ਨ ਤ	Correct
ਇੰਟਰਨੇਸ਼ਨਲ	ਇ ਨ ਟ ਰ ਨ ਏ ਸ਼ ਨ ਲ	Correct
ਸਪੁੱਤਰ	ਸ ਪ ਉ ਅ ਤ ਰ	Correct
ਹਮਦਰਦੀ	ਹ ਮ ਦ ਰ ਦ ਈ	Correct
ਕੈਸ਼	ਕ ਐ ਸ਼	Correct
ਗੋਖਲੇ	ਗ ਓ ਖ ਲ ਏ	Correct
ਲੂੰ-ਕੰਡੇ	ਲ ਉ ਨ - ਕ ਨ ਡ ਏ	Correct
ਗਾਲੀਗਲੈਜ	ਗ ਆ ਲ ਈ ਗ ਲ ਔ ਜ	Correct
ਝਾਂਕ	ਝ ਆ ਨ ਕ	Correct
ਤਜਰਬੇਕਾਰ	ਤ ਜ ਰ ਬ ਏ ਕ ਆ ਰ	Correct
ਪੁੰਨ-ਭੂਮੀ	ਪ ਉ ਨ ਨ - ਭ ਉ ਮ ਈ	Correct

#### IV Conclusion

In this work, we have develop the DJ Rule-Based algorithm on words according to their rules. With the help of this algorithm we have noted an accuracy of 99.9% depending upon the number of rules that are implemented. As future work, we can use the sound button for pronouncing the word and further implementation can be done on the line or paragraph also. This software can be beneficial for those people who are learning punjabi. With this software one can learn about the very important aspect of Punjabi Grammar i.e. 'Dhunivishleshan' is in an straightforward and interesting way that can give entirely new dimension that add new way to traditional approach to Punjabi Teaching. This can also be used to solve and test the problems related to Punjabi Grammar.

#### References

- [1] Deepjot Kaur, Navjot Kaur, "A Review: An Efficient Review of Phonetics Algorithms", International Journal of Computer Science & Engineering Technology (IJCSSET), ISSN : 2229-3345 Vol. 4 No. 05 May 2013.
- [2] Meenu Bhagat, "Spelling Error Pattern Analysis of Punjabi Typed Text", Thesis report, Thapar University, Patiala (2007).
- [3] Parminder Singh and Gurpreet Singh Lehal, "Text-To-Speech Synthesis System for Punjabi Language".
- [4] Gurmukhi Vowels <http://sikhism.about.com/od/learntoreadgurmukhi/ig/Gurmukhi-Vowels-Illustrated/>
- [5] Rakesh Chandra Balabantaray, Sanjaya Kumar Lenka, "An Automatic Approximate Matching Technique Based on Phonetic Encoding", IIIT Bhubaneswar, International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012.
- [6] Sheilly Paddal, Nidhi, "Punjabi Phonetic: Punjabi Text to IPA Conversion", Department of Computer Science & Engineering, SVIET Banur, Punjab, International Journal of Emerging Technology and Advanced Engineering Issues, Vol.2, Oct. 2012.
- [7] Priyanka Gupta and Vishal Goyal, "Implementation of Rule Based Algorithm for Sandh-Vicheda of Compound Hindi Words", Department of Computer Science Punjabi University Patiala, International Journal of Computer Science Issues, Vol. 3, 2009.

- [8] Kare Sjolander, "Automatic alignment of phonetic segments", Centre for Speech Technology, Department of Speech, Music (2001).
- [9] Walter D. Andrews, Mary A. Kohler and Joseph P. Campbell, "Phonetic Speaker Recognition", Department of Defense Speech Processing Research.  
<http://jcarreras.homestead.com/rrphonetics1.html>.
- [10] David Pinto, Darnes Vilariño, Yuridiana Alem, "The Soundex Phonetic Algorithm Revisited for SMS-based Information Retrieval", Department of computer science, Mexico.
- [11] Contractor, D., Kothari, G., Faruque, T.A., Subramaniam, L.V., Negi, S.: Handling noise queries in cross language FAQ retrieval. In: Proceedings of the 2010 Conference on Empirical Methods of phonetics in Natural Language Processing. EMNLP '10, Stroudsburg, PA, USA, Association for Computational Linguistics (2010) 87–96.
- [12] Gurpreet Singh Lehal, "A Survey of the State of the Art in Punjabi Language Processing", Language in India, Vol. 9, no, 10, pp. 9-23, 2009.
- [13] Bodo Winter, "Pseudoreplication in Phonetic Research", Department of Linguistics, Germany, August 2011.
- [14] Ashby, "New Directions in Learning, Teaching and Assessment for Phonetics", Estudios de Fonética Experimental in 2008, XVII, 19-44.
- [15] Rajkovic, P., Jankovic, D.: Adaptation and Application of Daitch-Mokotoff Soundex Algorithm on Serbian names. In: XVII Conference on Applied Mathematics (2007).