

A Tool to Search and Convert Reduplicate Words from Hindi to Punjabi

Shivani Sachdeva, Dr. Ashwani Sethi

Department of Computer Science, GKU Talwandi Sabo

Professor, GKU Talwandi Sabo Department of Computer Science, GKU Talwandi Sabo

I. Introduction

1.1 Introduction

Language, the best means of communication, is not a static one. It changes from time to time. The causes for these changes may be borrowing, language contact and interference and convergence from other languages. Due to these processes many structural features may come from one language into another Language. These exchanges reflect in the phonological, morphological and lexical levels. While working with the various kind of languages there are some circumstances when there is reduplication of words that is repetition of all or the part of the word. Reduplication, the doubling or partial doubling of syllables and words, is very frequent in Hindi-Urdu. It may apply to full words of any kind: nouns, adjectives, adverbs, verbs (but not auxiliary verbs), even pronouns:

1.2 Reduplication Definition

Reduplication is the repetition of all or part of the base with or without internal change before or after the base itself. It is of two types:

1.3 Types...

Morphological & Lexical

1.3.1. Morphological Reduplication:

Morphological reduplication refers to the minimally meaningful and segmentally indivisible morphemes that are a larger number of expressions used in speech where sound and sense seems to be united. These expressions have been termed as Onomatopoeia such as:

1. Some Acoustic Noises

Monkey chattering	U? U?
Cat Chattering	Mu?Mu?

2. Noises of Natural Phenomenon

Rain pattering	tap tap
Thundering Sound	gar gar

3. Noises by Humans

Laughing	Ha! Ha!
Khick! Khick!	

1.3.2. Lexical Reduplication:

Lexical reduplication refers to the repetition of any sequence of phonological units comprising a word. Lexical reduplication, unlike morphological reduplication, is not minimally meaningful and thus can be further divided as they are formed of two identical words or two non – identical phonological words.

It can be partial or complete at this level. “Complete Lexical Reduplication is constituted of two identical (bimodal) words, Ex.: baiThebaiThe “While sitting” in Hindi. Partial reduplication, on the other hand, is constituted of partial repetition of a word either phonologically or semantically. Echo words such as khanavana “Food etc.” or compounds It consists of four types:

1.3.2.1. Echo-Formation:

An echo word is defined as a partially repeated form of the base word; partially repeated in the sense that either the initial phoneme which may be either consonant or vowel or the syllable of the base is replaced by another phoneme or another syllable. Here one thing we find that the replacer sound sequences are more or less fixed and rigid.

In Bengali repetition starts with –T, Punjabi –S and Hindi –V. The echo word has neither any individual occurrence nor any meaning of its own in the language. It acquires the status of a meaningful element only after it is attached to a word.

Here are some examples

Hindi	Gloss
nam – wam	Name
khun – Vun	Blood
asan – vasan	Easy
Phal – val	Fruit
pyar – vyar	Love
gana- -vana	Song

Echo Formation Used in Sentences:

- 1 nam - vam: nam – vam pata kar ke kya karoge? (Hindi)
- 2 khun -vun: khun – vun nikle ki nehi? (Hindi)
3. asan -vasan : asan - vasan question mat kya karo (Hindi)

4. phal -val : phal – val kha kar wo ghar chala gaya (Hindi)

5. pyar -vyar : pyar -vyar sab bekarkicheejhai (Hindi)

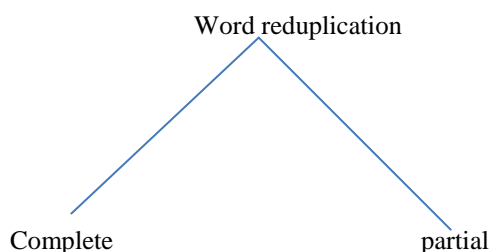
6.gana-vana: gana –vana gaya karo (Hindi)

1.3.2.2. Compounds

Compounds refer to the paired constitution in which the second word is not an exact repetition of the first but has some similarity or relationship to the first word either on the semantic or on the phonetic level. A compound may be used independently in a sentence. Here, the paired construction does not have a new meaning.

1.3.2.3. Word Reduplication

Word reduplication refers to those paired constructions when a single word or a clause is repeated once in the same sentence without any phonological or morphological variation. This word reduplication is of two types:



i. Complete Reduplication:

When both the parts have meanings and are meaningful, then the reduplication is called a complete reduplication.

Examples of Complete Reduplication:

Hindi	Gloss
chalte – chalte	Running
dhire – dhire	Slowly
achche –achche	Good
safed – safed	White
chhote – chhote	Little

(ii) Partial Reduplication:

Partial reduplication in the sense that here only one word or free morpheme is meaningful other word is meaningless.

Examples of Partial Reduplication

Hindi	Gloss
pani – wani	Water
kam – wam	Work
chawye-waye	Tea
pan-wan	Betel

1.3.2.4. Discontinuous Reduplication:

It is the reduplication with inter-fusion of a syllable which could be a vowel or vowel-consonant(VC) or consonant-vowel(CV) which is termed as discontinuous reduplication.

Example of Discontinuous Reduplication

1. criss-cross: kisi ke sath criss-cross mat khela karo (Hindi)
2. zig-zag: ye rastazig-zaghai (Hindi)
3. tip-top: hamesha tip-top rahakaro (Hindi)

II. Problem Definition

The aim of the research work here is to automatically identify the echo words in hindi language that comes frequently in conversation of the day today life by using existing mechanism that is rule based approach so that we will be able to translate that particular one onto the Punjabi language. To identify these words we can use some lexical dictionary or the other methodologies to be followed.

As we know that to work on to that particular research work we have to go through various phases such as the collection of data from various sources of hindi then we have to apply some particular mechanism either rule based or the statistical approach to identify that echo reduplicated word and then finally translate them onto the other language and that is the Punjabi language.

2.1 Sources for the Data:

The data for this study are to collected by using various types of methods. The prime data for the present study are collected from various sources like the hindi texts, lexicons, daily news papers , weekly and monthly magazines, short story books, novels, etc.

III. Methodologies

3.1 Methodologies to be used:

There are the two approaches that can to be used to achieve the particular goal so as to identify the echo words and then translate that particular one onto the punjabi language. The technique that can be followed are:

- Rule based approach
- Statistical approach

In the *rule based approach*, rules are to be followed various rules according to the circumstances or depending upon the situation rules are to be applied where as in *statistical approach* various statistical tools are to be used to perform the desired task.

Here in this research work we will use the first one that is rule based approach in which various rules are to be be developed so that it can identify the echo words reduplication and then work on it as per the requirements

IV. Literature Survey

4.1 Introduction

This chapter presents a comprehensive literature survey the various types of reduplication with having its fully and partially reduplication in the form of echo words. In this work we will work on to identify the echo reduplicated words in Hindi and how to translate them into the other Punjabi language.

4.2 Previous Work

R.M.K Sinha and Anil Thakur, "Dealing with Replicative Words in Hindi for Machine Translation to English" they discuss different types of replicative words in Hindi and their syntactic and semantic characteristics to formulate rules and strategies to identify their multiple functions and mapping patterns in English for machine translation from Hindi to English. The phenomenon, in a natural language, where a word repeats itself in a sequence without any morphological variation has been referred to as "complete word reduplication". In this paper, they call this process as replication, retaining the definition given in (Abbi, 1992): "those paired constructions when a single word or a clause is repeated once in the same sentence without any phonological or morphological variation."

In Thoudam Doren Singh, Sivaji Bandy opadhyay; "Web Based Manipuri Corpus for Multiword NER and Reduplicated MWEs Identification using SVM" A web based Manipuri corpus is developed for identification of reduplicated multiword expression (MWE) and multiword named entity recognition (NER). The mode of corpus collection and the identification of reduplicated MWEs and multiword NE based on support vector machine (SVM) learning technique are reported. The SVM based reduplicated MWE system is evaluated with recall, precision and FScore values of 94.62%, 93.53% and 94.07% respectively outperforming the rule based approach. The recall, precision and F-Score for multiword NE are evaluated as 94.82%, 93.12% and 93.96% respectively.

Annie Montaut, Inalco, Paris, Reduplication and 'echo words' in Hindi/ Urdu, enquire into the various meanings of reduplication as a linguistic operation, and not as a merely stylistic or expressive device. The theoretical frame is Antoine Culioli's 'énonciative' linguistics (notion and located occurrence, notional domain and boundary); context and intersubjectivity are taken into account as much as possible. Section deals with total reduplication, within the nominal, verbal and adjectival category: it shows that reduplication on an occurrence modifies the relation between the reduplicated term and the term syntactically associated to it by denying the occurrence any specific stable value. It thus modifies the scheme of individuation of the notion

Dr. H.S. Ananthanarayana (1976) describes "Reduplication in Sanketi Tamil". He defines the reduplication as "The repetition of all or part of a base word". To describe the reduplication he considers the onomatopoeic words vs. non-onomatopoeic words. He starts to describe reduplication by taking the onomatopoeic words which are formed from sounds which resemble those associated with the object or action. He gives examples like

1. kaTakaTa "to grind one's teeth"
2. kilikili "to laugh heartily"

Dr. Imtiaz Hasnain, Aligarh Muslim University, reviewed Dr. Anvita Abbi's book "Reduplication in South Asian Languages: An Areal Typological and Historical Study". The summary of the book is in the following: i) There is an impending danger in the identification of linguistic area on the basis of commonality of features. ii) With a view to circumventing the impending danger the author rightly prefers to talk of a particular linguistic features area with an emphasis on Reduplicated Structures (RSs). iii) RSs have been examined in the entire major and some minor languages of South Asia. iv) The language data consist of 17 tribal Tibeto-Burman and 4 Austro-Asiatic languages and the remaining five namely Juang, Kurukh, Santhali, Sora and Dakhini, which are spoken outside the Indian subcontinent but are akin to the Austric branch of the languages of India, Hokkien and Chinese structures have also been looked into, with a view to unravel their RS patterns. v) The RSs used in this study are the ones which are most commonly found in all the languages, thus constituting areal universals.

M. Arunachalam (1977) treats echo-words in Tamil. In his article he has taken up three terms iraTTaikkiLavi, aTukkuttoTar and echo-words for consideration. The Tamil grammatical works Tolkaapiyam and Nannuul mentioned iraTTaikkiLavi and aTukkuttoTar. They referred to these two forms always occurring as double. If they are occurring as single it won't give meaning. These are called as *iTaicolin* in the Tamil grammar.

Dr. Gnanasundram has written a paper entitled "Echo words in Tamil" in the year 1972. He takes up onomatopoeia and lexical doublets for his study and defines echo-words as "When an onomatopoeia is reduplicated and when reduplicated part undergoes a sound change then that is called echo words". He elaborately has collected all these kinds of words. He analyzes these words and finds out only the sound change which has taken place in these words. Because the first or the second syllable is changed in the duplicated part he considers all the above said words as echo-words. He actually lists the echo words but he does not take them into account for his

study. Most probably he does not recognize that one, as actual echo-word. But he lists many onomatopoeia words and lexical doublets as echo-words in his paper on the basis of consonant and vowel changes.

Dr. PeriBhaskarrao has published a small book in the year 1977 titled "Reduplication and Onomatopoeia in Telugu." He notes that Telugu utilizes the reduplication process for bringing out various subtle meaning differences. He includes the formation of echo-word under the reduplication. He suggests that the onomatopoeia occurring, in a reduplicated form, can be taken together..

Sangamitra has written a thesis on "Reduplication in Bengali, Mundari and Telugu: A Linguistic Study". In it, her aim is to study not only the reduplication but also the other reduplicative structures like echo – words, expressives and copulative compounds with focus on semantic analysis in Bengali, Mundari and Telugu. She concludes her thesis on three grounds, namely, structural pattern, findings on semantic grounds and future scope. For structural pattern she follows Moravisk's proposition and categorizes the reduplication under two heads, namely, monomorphemic model and bimodel. Bengali, Mundari and Telugu languages have identical, discontinuous and partial reduplication

Dr. A. Usha Devi (2001) has prepared a small dictionary for Telugu onomatopoeic forms. In the introduction she mentions that the language is operating on two systems, viz., grammatical and lexical. The grammatical system analyses the phoneme, morpheme, sandhi, compound and sentence. The lexical system deals with the lexical categorization on the basis of sense and the relation between the common, opposite and the synonymous words

V. Objectives

5.1 Objectives of the Project Work

- i. To thoroughly study of the basic concepts of REDUPLICATION echo words in hindi.
- ii. To automatically identify the reduplicated echo words in hindi language.
- iii. To use various resources such as lexical dictionaries or other methodologies.
- iv. To translate the particular words onto the Punjabi language that is to be required

VI. Work to be carried out

6.1 Work to be carried out in Fourth Semester

Reduplication is very common in our day today language conversation and the different languages have different properties related to it. Such that these languages have their own combination of vowel/consonants, consonant/vowels and so on .but the problem here to be discussed is that the reduplicated words that comes out while working

with the hindi language have to be identified automatically by the system.

The framework of rule based approach is being followed here to identify such words, as we know the rule based approach uses various rules depending upon their sentences and all Following work will be carried out during this semester.

- Study the problem of reduplication
- Study various past and recent researches on the topic reduplication.
- Study the previous implementation of reduplicated echo words translation.
- Translate reduplicated echo words of hindi to punjabi
- Thoroughly study about rule based approach.
- Comparative analysis of proposed research existing research.

References

- [1] M.D SohelRana (P hd)nov 2010, "reduplication in bengali language"; Volume 10:11 ;pp 1930-2940.
- [2] Kishorjit, N. &Sivaji, B., (2010) *Identification of MWEs Using CRF in Manipuri and Improvement Using Reduplicated MWEs, In the Proceedings of 8th International Conference on Natural Language (ICON-2010), IIT Kharagpur, India, pp51*
- [3] Thoudam Doren Singh, Sivaji Bandyopadhyay, (august 2010)" *Web Based Manipuri Corpus for Multiword NER and Reduplicated MWEs Identification using SVM*" *Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), the 23rd International Conference on Computational Linguistics (COLING), Beijing, pages 35-42*
- [4] A.PARIMALAGANTHAM,(8 august 2009)"*A STUDY OF STRUCTURAL REDUPLICATION IN TAMIL AND TELUGU*": LANGUAGE IN INDIA ,*Strength for Today and Bright Hope for Tomorrow "Volume 9 : ISSN 1930-2940*
- [5] JANA NOVOTNA,(2000)"*REDUPLICATION IN SWAHILI*", AAP 64 : *Swahili Forum VII • 57-73*
- [6] Rajendra singh (2005)"*Reduplication in Modern Hindi and the Theory of Reduplication*", *In Hurch (ed.), 263-81.*
- [6] Jeffrey Lidz," (1999) "*Echo Reduplication in Kannada*":*Implications for a Theory of Word-formation*" *U. Penn Working Papers in Linguistics, Volume 6.3.*

- [7] Akinlabi, A.M. & E. E Urua (1996) “A Prosodic Account of Reduplication in Ibibio”. *Nigerian Language Studies*. “A Journal of the National Institute for Nigerian Languages, Vol. 4” 1-15
- [8] Giannakis, Geokrgios, 1992. “Reduplication as a Morphological marker in the Indo-European Languages” *Word, vol.43. No 2*.
- [9] Al Mtenje, 1988, “On tone and transfer in Chichew: A Reduplication” *Linguistics*26. pp. 125-155.
- [10] Anvita Abbi.(1997)”Reduplicated Adverbs in Hindi Indian Linguistics”*Journal of the Linguistic Society of India vol:38;1977:pp: 125-135*.