1. INTRODUCTION
1.1 Natural Language Processing –
Natural Language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. In theory, natural-language processing is a very attractive method of human-computer interaction. Natural language understanding is sometimes referred to as an AI-complete problem, because natural-language recognition seems to require extensive knowledge about the outside world and the ability to manipulate it. NLP has significant overlap with the field of computational linguistics, and is often considered a sub-field of artificial intelligence. Natural Language Processing (NLP) [1] deals with the human understandable language. NLP is an important field of computer science. It is an area of research that explores how computer can be used to understand and manipulate natural language text.

1.2 Reduplication –
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:

1.2.1 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!

1.2.2 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.

Abstract- Natural Language processing (NLP) is a field of computer science and linguistics concerned with the interactions between computers and human (natural) languages. In theory, natural-language processing is a very attractive method of human-computer interaction. Punjabi, often spelled Panjabi, belong to the Indic group of the Indo-European family of Languages. It is spoken in Punjab, the remarkable district partitioned between India and Pakistan. Reduplication is the repetition of all or part of the base with or without internal change before or after the base itself. The aim of the paper is to explain the methods of identifying the echo words in natural language that comes frequently in conversation of the day today life.
Reduplication Words In Natural Language Processing - A Survey

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:

(a) 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

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>nam – wam</td>
<td>Name</td>
</tr>
<tr>
<td>khun – Vun</td>
<td>Blood</td>
</tr>
<tr>
<td>asan – vasan</td>
<td>Easy</td>
</tr>
<tr>
<td>Phal – val</td>
<td>Fruit</td>
</tr>
<tr>
<td>pyar – vyar</td>
<td>Love</td>
</tr>
<tr>
<td>gana- -vana</td>
<td>Song</td>
</tr>
</tbody>
</table>

Echo Formation Used in Sentences:
1. nam - vam: nam – vampatakarkeyakaroge? (Hindi)
2. khun - vun: khun - vunnikkleinehi? (Hindi)
3. asan -vasan : asan - vasan question mat kyakaro (Hindi)
4. phal - val : phal - valkharedkarwogarchalagaya (Hindi)
5. pyar - vyar : pyar - vyar sab bekarkicheejhai (Hindi)
6. gana-vana: gana - vanagayakaro (Hindi)

(b) 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.

(c) 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:

Word reduplication

Complete Partial

(i) Complete Reduplication
When both the parts have meanings and are meaningful, then the reduplication is called a complete reduplication.

Examples of Complete Reduplication:

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>chalte – chalte</td>
<td>Running</td>
</tr>
<tr>
<td>dhire – dhire</td>
<td>Slowly</td>
</tr>
<tr>
<td>achche – achche</td>
<td>Good</td>
</tr>
<tr>
<td>safed – safed</td>
<td>White</td>
</tr>
<tr>
<td>chhote – chhote</td>
<td>Little</td>
</tr>
</tbody>
</table>

Complete Reduplication Used in Sentences
chalte – chalte : bas chaltechalte ho gaya (Hindi)
dhire – dhire : bus bohutdhiredhirechalraahai. (Hindi)
achche –achche: SufiaachcheachchePakorabanatihai (Hindi)
safed – safed: safed safed kaprapahnahakaro (Hindi)
5. chhoté – chhoté : chhotechotethealu le lo (Hindi)
These examples are called complete reduplication because here single word or a clause is repeated once in the same sentence
without any phonological and morphological change.

(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

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>pani – wani</td>
<td>Water</td>
</tr>
<tr>
<td>kam – wam</td>
<td>Work</td>
</tr>
<tr>
<td>chawye-waye</td>
<td>Tea</td>
</tr>
<tr>
<td>pan-wan</td>
<td>Betel</td>
</tr>
</tbody>
</table>

Partial Reduplication Used in Sentences
1. pani - wani: pani - wani hameshasafpyakaro (Hindi)
2. kam - wam: kam - wam khatamkarophirgharchalo (Hindi)
3. chawye-waye: chawye-waye ho jaye? (Hindi)
4. Pan-van pan-van khakekyakaroge? (Hindi)

(d) 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

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>criss - cross</td>
<td>kisikesathcriss-cross mat khelakaro (Hindi)</td>
</tr>
<tr>
<td>zig - zag</td>
<td>ye rastazig-zaghai (Hindi)</td>
</tr>
<tr>
<td>tip - top</td>
<td>hamesha tip-top rahakaro (Hindi)</td>
</tr>
</tbody>
</table>

2. RELATED WORK
Annie Montaut [2] has written a paper entitled “Reduplication and ‘echo words’ in Hindi/ Urdu” in the year 2010. He explains the paper in two sections. In this paper the first 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 (its actualization into an occurrence). The second section, dealing with partial reduplication or echo constructions, whether formed with a v- substitution to the initial consonant or with other forms of alliteration, shows that it modifies the notion itself by de-centring it, and reshapes it by taking into account various forms of heterogeneity, particularly the conflicting viewpoints of speaker and hearer.
Meladel Mistica et.al.[3] describes reduplication in Indonesian language. In this it basically deals with morphophonemic fact relating to sound changes in morphemes and how the construction of word formation rules are done to create these derived words exhibiting reduplication. It describes various tools to construct the word. It also gives description about morphological analyzer.
In this paper Omkar N Koul [4] discusses the reduplication in Kashmiri language. Here it shows the reduplication in Kashmiri at expression level in which noun, pronoun, adjective, verbs, clauses and phrases are defined. It also shows the reduplication at semantic level, word level, echo words and onomatopoeic words.
In this paper Le Hong Phuong et.al. [5] describes reduplication in Vietnamese language. Vietnamese is a language where words are created with the combination of multiple syllables whose phonics is similar. It gives description about various types of reduplication in Vietnamese and makes use of optimal finite state devices in particular minimal sequential string to string transducers to build computational model for every efficient recognition and production of these words. At last it gives applications of this model.
In this paper Jan Daciuk et.al [6] discusses the a new method for constructing minimal, deterministic, acyclic finite state automata from a set of strings. Traditional methods consist of two phases: the first to construct a tree, the second one to minimize it. The approach is to construct a minimal automaton in a single phase by adding new strings one by one and minimizing the resulting automaton on-the-fly. They present a general algorithm as well as a specialization that relies upon the lexicographical ordering of the input strings. Proposed method is fast and significantly lowers memory requirements in comparison to other methods.
In this Thoudam doren Singh et.al [7] discusses about the Manipuri language. It is developed for reduplication multiword expression (MWE) and multiword named entity recognition (NER). Here they collect a news corpus from a Manipuri news
website. They have collected four and half a million words. Corpus collection and reduplication is based on support vector machine (SVM) learning techniques are reported.

Shanthi Nadarajan [8] explains in the paper “A cross linguistic Study of Reduplication” that reduplication is an important phenomenon in language studies. Its linguistic forms have long been studied in terms of various formalist theories but there is a general consensus that the role of reduplication in English has been widely ignored. Recent corpus based studies have indicated that the usage of reduplication among English speakers is more widespread than assumed to be and needs in depth study. This paper look at the traditional forms of reduplication and current views on how they occur. It also define reduplication and repetition found in Malay and Tamil and compares them with English forms, in order to establish a pattern. Wendell Kimper [9] in June 2006 write a paper “Syntactic Reduplication and the spell out of movement chains” The central topic in this paper has been the spell out of movement chains; under a copy theory of movement, traces are not grammatical primitives, but rather unpronounced copies of the moved element. At issue here are the factors restricting the realization of phonological form to a single link in the chain, and the conditions under which multiple chain links can be spelled out. This paper has approached that issue from the viewpoint of syntactic reduplication, a process by which a particular meaning is manifested via syntactic copying. This paper also explains that in syntactic reduplication, movement results in two copies of the reduplicated item, both of which are spelled 28 out. Multiple-copy spellout is accomplished via the same mechanism driving movement in these cases, a reduplicative morpheme that combines with a moved copy to form a reduplicant — the reduplicant is a morpheme in its own right, associated with the phonological form of the copy but the meaning of the reduplicative morpheme. In this way, syntactic reduplication is parallel to morphological reduplication.

Manijeh Mirmokri et.al. [10] explains in the paper “On the Reduplication in Kurdish language” that there are two basic forms – full reduplication and partial reduplication that each of them includes several forms. In this research, purpose is identifying and introducing a variety of reduplication patterns in Kurdish language. For this purpose they explored more than 700 samples of reduplications in Kurdish language and categorized them in several sets. They noticed that frequentative, increased action, and plural action as the most common semantic categories marked by reduplication in the samples. The studies show that this morphological process especially partial type causes some changes in the meaning of resulting combinations. It also seems this kind of words especially partial ones are used in informal language and an ironic meaning is inferred by them.

3. CONCLUSION

Language, [11] 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. Our research paper explains the issue , challenges and types of reduplicate words in natural language processing.

4. REFERENCES

[7] Thoudam doren Singh and Sivaji Bandyopadhyay “Web based manipuri corpus for multiword NER and reduplicated MWEs identification using SVM” proceedings of 1st workshop on south and southeast asian natural language processing, pp 35-42.2010

Reduplication Words In Natural Language Processing-A Survey