Chapter 6
Conclusion

A word is not a crystal, transparent and unchanged, it is the skin of a living thought and may vary greatly in colour and content according to the circumstances and the time in which it is used.
- Oliver Wendell Holmes

6.0 Introduction
The present study has the observations and findings about the grammar of Gujarati language. The focus is on the structure of the verb and verb phrase. In order to use the grammar for NLP applications, the verbs have been categorized into morpho-phonetic frames.
The past tense suffix behaves differently in Gujarati. The past tense suffixes are attached according to the morpho-phonetic features of the verb.

6.1 Gujarati Language
Gujarati language belongs to the Indo-Aryan family and has its roots in Sanskrit. Modern Gujarati has come through the stages of Prakrit, Apbhransh and other Middle and Neo Gujarati stages.

6.1.2
Gujarati is written in Gujarati script. It is a variant of Devanagari script differentiated by the loss of the characteristic horizontal line running above the letters and by a small number of modifications in the remaining characters.

6.1.3
Gujarati has a syllabic alphabet and all consonants have an inherent vowel. Vowels can be written as independent letters, or by using a variety of diacritical marks which are written above, below, before or after the consonant they belong to.
6.1.4

Gujarati is a head-final, or a verb final language. Adjectives precede nouns, direct objects come before verbs, and there are postpositions. The word order of Gujarati is SOV, and there are three genders and two numbers. There are no definite or indefinite articles. A verb is expressed with its verbal root followed by suffixes marking aspect and agreement in what is called a main form, with a possible proceeding auxiliary, marking tense and mood, and also showing agreement. Causatives (up to double) and passives have allomorphic bases.

6.1.5

Gujarati enjoys the unique status of being the first language of the father of nation of two countries- India and Pakistan. Gujarati was the first language of Mahatma Gandhi and Mohammad Ali Jinnha, the founder of Pakistan. Gujarati is the first language of the iron-man of India- Sardar Vallabhbhai Patel and Dada Bhai Navroji; they played a vital role in the freedom struggle of India and also in shaping the future of independent India. The doyen of Industrial revolution in modern India, Sir Jamshedjee Tata, was a Gujarati Parsi. This is the historical importance of Gujarati language. In the present scenario Gujarati is the first language of Ambanis and Tatas the global magnates of India. From history to the present times Gujarati people have made a noteworthy contribution in the progress of the country. And this is a techno-savvy generation. The language needs to progress with the ever progressing technology.

NLP is the product of the computer technology. NLP is the first step in this direction. As we know the verb is the ruling planet in the sentence, the analysis is based on the structure of verbs and verb phrases.
6.2 Verbs

6.2.1

Verbs are generally categorized on transitivity and agentivity. Semantically verbs can be divided in to three categories- action, process and state. They can be further divided on syntactic bases. Action verbs can be intransitive, transitive and ditransitive. Process verbs can be with one or two objects (natural and mental too) and Static verbs are neutral. This categorization is for human beings to understand. The computer can not understand it. It needs a concrete division, which is directly visible in terms of alphabets. It has to be marked in such a way the computer understands it. Computer dose not understand the worldly knowledge, it can not distinguish between action, process or state verbs. It needs clear signs. These clear and neat signs are possible only if the verbs are divided on morpho-phonetic basis, viz the syllable structure.

6.2.2

The diagram below shows the general syntactic classification based on morpho-phonetic level of the verb roots. The verb roots are broadly divided according to their syllabic structure as monosyllabic and multi-syllabic. Monosyllabic has regular and irregular forms; with reference to its behaviour for the past tense. Multi-syllabic is categorized as per the position of vowels in the verb root.

The mono-syllabic verb roots are divided as regular and irregular, on the basis of the behaviour with the past tense suffix. The regular ones follow the general past tense suffix /ya/ and the irregular ones follow other various past tense suffixes like - /dha/, /Tha/, /ta/ and /na/. The next division is of multi-syllabic verb roots. It is based on the position of the vowels in the verb roots. The first division, in it, is of verb roots with long penultimate vowel. Example- /chAla/, /Ug/ etc this is further divided according to their behaviour. Another category has the verb roots whose first syllable is long. In it, one category shortens the long vowel, second undergoes no change and third one inserts an infix.
This division is presented diagrammatically in fig- 6.1. But this classification did not prove apt for the analyzer. The analyzer needed more neat and simplified classification of the verb roots. So in another try the verb roots were divided according their endings. The detailed explanation is provided in 6.2.3 and it has been diagrammatically explained in the figure 6.2.

6.2.3

In the present work, the categorization is based on the morpho-phonetic features for NLP applications. And it has been tested on the tense and voice suffixes. Gujarati verb roots need to be divided as per their endings for the computational purpose. Like verb root ending in ‘a’ (અ) and verb root ending in other than ‘a’. The verb roots ending in ‘a (અ)’ can have first syllable long or short e.g. ચાલ/ chAla/, એરાથ/ Ara~bha/ રમ/ rama/ The one that has long syllable might or might not undergo change; when TAM suffixes are attached. For example – chAla become chala but Ara~bha remains Ara~bha. And those with the short syllable dose not undergo any change; when TAM suffixes are attached e.g. રમ/ raml/, કર/ kar/. The verb roots ending in vowels
which are other than ‘a’ can be divided as- regular and irregular. This division is made on their behaviour, with the past tense suffixes. Past tense suffixes behave differently and there are four suffixes representing it. The general ones are placed in regular ones. They take the /ya/ past tense suffix. And the different ones are placed into irregular. And it is further divided in ‘a( અ )’ ending and other than ‘a’ ending. The regular ones are \textit{dho} (ધો) \( gA(ગા) \). The irregular ones ending only in ‘a’ (અ) are \textit{besa} (બેસા), \textit{pesa} (પેસા). The ones other than /a/ are \textit{su} (સુ), \textit{jA} (જા), \textit{khA}(કા), \textit{pl}(પલ), \textit{de}(ડે), they take different past tense markers. The following figure shows the division used in the analyzer.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig-6-2.png}
\caption{fig- 6.2}
\end{figure}

6.2.4 Gujarati letters have inherent vowels. Hence, all words in written form end in a vowel. The dominant vowel is ‘a’. It is not directly visible. All other vowels are visible with dependent Matras (Vowel signs). Another important division is made on the bases of the first syllable. When the first syllable is long it is shortened in passive and causal forms. This happens with majority of verb roots. There is an exception too. Some verb roots with first syllable having long vowel, do not undergo such change. For the verb roots which have first syllable does not
undergo any change. They end in ‘a(-addon)’ or ‘A(Addon)’. So this division is made on the bases of its reaction with passive and causal forms.

The verb roots which end in other than ‘a’ vowel are categorized for past tense. The verb roots which behave in a normal way take the general past tense suffix – ‘ya(-addon)’ and irregular ones can take ‘dha(Addon)’, ‘Tha (Addon)’, ‘ta(Addon)’ and ‘na(Addon)’. Here, too the verb root ending in ‘a(Addon)’ will take ‘Tha (Addon)’.

This division results in to nine verb groups, which leads to nine types of paradigms.

1. chAla चाला
2. Arambha अरांभा
3. kara करा
4. gA गा
5. besa बेसा
6. KhA खा
7. su सू
8. bi बी
9. jA जा

Each verb groups have verb roots that behave same as the head verb; they follow the same paradigm. Verb is an open category. New verbs keep on pouring in. I have prepared a sample analyzer but I have tried to put all the verbs found in the standard dictionary. The software can analyze 2361 verbs in total. The highest numbers of verbs are found in kar group – 1328. The second highest is in Arambha group – 732. The third highest is in chAla group – 244. Then comes gA verb group with 45 verbs. KhA verb group has 6 verbs, besa has 3 verbs. And su, jA, bina has one in each group.
<table>
<thead>
<tr>
<th>Sr.No.</th>
<th>Verb Group</th>
<th>Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>kara</td>
<td>1307</td>
</tr>
<tr>
<td>2.</td>
<td>Arambha</td>
<td>711</td>
</tr>
<tr>
<td>3.</td>
<td>chAla</td>
<td>244</td>
</tr>
<tr>
<td>4.</td>
<td>gA</td>
<td>44</td>
</tr>
<tr>
<td>5.</td>
<td>khA</td>
<td>6</td>
</tr>
<tr>
<td>6.</td>
<td>besa</td>
<td>4</td>
</tr>
<tr>
<td>7.</td>
<td>su</td>
<td>1</td>
</tr>
<tr>
<td>8.</td>
<td>jA</td>
<td>1</td>
</tr>
<tr>
<td>9.</td>
<td>bina</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>2319</strong></td>
</tr>
</tbody>
</table>

*table 6.1*

Each verb has approximate 120 verb forms. This is approximate because every verb does not take all possible verb forms. Some verbs do not get passive and causal. And not all verbs take case markings or become participles. So the maximum forms of verbs a verb can wear are 130, for a diverse verb like chAla.

The approximate number of verb forms available in Gujarati language can be $2,319 \times 120 = 278,280$. This number does not include the tense auxiliaries; if they are included then the number can rise to 2,78,300 approximate. The software can identify and analyze 22,78,360 verb phrase constituents, approximate, including – negators, adverbs and particles.

6.2.5

Along with verbs and adverbs, verb phrase also have particles, tense auxiliary and negations belonging to a closed category. The analyzer has a special closed class category which has all the tense auxiliaries, particles, adverbs and negations. Adverbs belong to open class but their structure is unclear, generally they are derived. And they are identified from the reference. In a way it is logically impossible to identify the real adverbs. I have selected few frequent adverbs and placed them in the closed class along with negations,
particles and tense auxiliaries. These words do appear in Verb Phrase and it is necessary to identify them. So they have been placed in to a special class. When ever the computer encounters any of them, it will identify them, without any process. By process here I mean is the various stages (set of rules) undergone by the computer to identify the verb and components of VP. (This process has been explained in detail in the previous chapter along with the samples from the analyzer)


At the end an appendix has been given, it has the verb lists and grid of the verb *chAl* *(walk)* with all the rules. I have taken *‘chAl’* because it is simple & widely used an main verb. *‘kar’* is very productive & highly used an explicator so I have taken *‘chAl’* as a perfect example.

6. 3 Scope of the work

Natural language processing is an emerging field for Gujarati language. Not much work has been done in this computational field. My thesis is a first concrete step in this direction. Along with the traditional Gujarati grammar my work is a step towards a synthesis of tradition and modernity. I have tried to provided a modern computational grammar (with specific importance to verb), which can be perceived as the stepping stone for NLP applications. NLP leads to the huge areas of machine translation, corpus building, corpus cleaning, information retrieval, error recognition, language teaching and many more upcoming fields. MT and corpus are gigantic advancements in the field of information technology and they are inevitable for the coming centuries.

I see a sparkling future for Gujarati language in NLP and I heartily wish that an NLP castle will be constructed and my work will serve as its firm base.

References –

1. A list of Gujarati adverbs provided by Dr. Arvind Bhandari (Head of the Department, Linguistics. School of Languages. Gujarat University) from his own research papers.