3.1 Introduction

The chapter discusses the different types of morphology found in Malayalam and the word compounding rules for concatenating two morphemes. The chapter discusses only concatenative morphology. This chapter also discusses the various types of models for morphological analysis. The morphological analyser requires the types of morphology in a language and the morphophonemic changes at morpheme boundary when two or more morphemes are concatenated.

3.2 Morphology of Malayalam

Malayalam is the mother language in the state of Kerala. It is one among the 22 official languages of India. The language is used by more than 35 million people spreading along the regions Kerala, Lakshadweep, and Pondichery.

Malayalam is one of the four major languages of this Dravidian family with a rich literary tradition. It is very close to Tamil, one of the major languages of the same family. This was due to the extensive cultural synthesis that took place between the users of the two languages. The origin of Malayalam as a distinct language may be traced to the last quarter of 9th Century A.D. Throughout its gradual evolution Malayalam has been influenced by the various circumstances prevailed during different periods.
Various types of morphological variations are found in Malayalam language. Morphological variations for words occur in Malayalam due to

i) Inflections
ii) Derivations
iii) Word compounding

In inflectional morphology a lexical category like noun or verb is attached to suffixes to generate words of the same category. In Derivational morphology categories like noun or verb with a suffix attached to it generates a word of new category. In word compounding new words are formed by combining a noun and a noun, noun and adjective, verb and noun, adverb and verb, adjective and noun and in some cases all the words of an entire sentence to reflect the semantics of the sentence [37].

3.2.1 Morphology for nouns

A lexical category like a noun or a verb is attached to suffixes to generate a word of the same category or of a different category. Nouns show inflectional morphology due to the addition of gender, number and case information. Nouns also show derivational morphology where modifiers and verbs are derived from nouns by adding proper suffixes.

3.2.1.1 Inflectional morphology for nouns

The morphological information that can occur with nouns are gender, number and case. Prepositions in English and postpositions in Hindi marking the cases are not attached to nouns. But in most of the Dravidian languages case suffixes are attached to nouns. There are three genders for nouns: masculine, feminine and neuter. The number can be singular or plural. For proper nouns and
abstract nouns, there are no plural forms. The case indicates the relation which a noun or pronoun has with verb of the sentence. The seven cases in Malayalam are nominative, accusative, sociative, dative, genitive and locative [38,39].

3.2.1.2 Derivational morphology for nouns

Adjectives, adverbs and verbs are derived from nouns by the addition of proper suffixes.

i) Adjectives

Modifiers are qualifiers and can be of three types. It is called *naamavisheshanam* (adjective) when it modifies a noun, *kriyaavisheshanam* (adverb) when it modifies adverb and *bhedakavisheshana* (modifier of modifier) when it modifies a modifier. The modifiers can be pure modifiers or those derived from nouns and verbs. Different types of pure modifiers are determinative adjectives, separative adjectives, interrogative adjectives, temporal adverbs, special adverbs and adverbs of manner.

Adjectives may be derived from another noun by the addition of *ulla*.

\[
\text{azham} (\text{depth}) + \text{ulla} (\text{has}) = \text{azhamulla} (\text{deep}).
\]

An adjective can be derived by suffixing *ile* (*ile / in)* to the noun,

\[
\text{kaat} (\text{forest}) + \text{ile} (\text{in}) = \text{kaattile} (\text{in forest}).
\]

An adjective can be derived by adding the suffix ‘*aththe*’ to the noun,

\[
\text{ann} (\text{that day}) + \text{aththe} (\text{on}) = \text{annaththe} (\text{that day’s})
\]
ii) Adverbs

Adverbs can be derived from a noun by suffixing adverbial markers.

The suffix – യായി /aayi is used with abstract nouns to make modifiers:

ബഹു യായി (bhangi/nice) + യായി (aayi/ly) = ബഹുയായി (bhangiayi/nicely).

To show direction, the suffix ആക്കു /ekku is added to the noun,

പ്രത്ി ആക്കു (puRathtz / out) + ആക്കു (ekku / to) = പ്രത്ിആക്കു (puRatthEkkz / to outside).

iii) Verbs

The nouns with vowel ending takes suffixes കൈയ്യന്ന് /ikkunnu, ചു / ichu and കൈക്കു /ikkum for present, past and future respectively. The nouns with vowel ending takes the suffixes കൈ /kkunnu, ശു / chu and കൈ / kkum for present, past and future tenses.

തളി (thati fat) + കൈയ്യന്ന് (kkunnu/present tense marker) = തളിക്കൈ

(thatikkunnu/becoming fat)

കല്ല് (kallz/stone) + കൈയ്യന്ന് (ikkunnu/tense marker) = കല്ലിക്കൈ

kallikkunnu (hardening)

The inflectional and derivational morphology for nouns commonly found in Malayalam are listed in Table 3.1.
### Table 3.1 Morphology for nouns

<table>
<thead>
<tr>
<th>Inflections</th>
<th>Type</th>
<th>Suffix</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>plural</td>
<td>ല്‍/kaL</td>
<td>ല്‍/kuttikaL</td>
</tr>
<tr>
<td></td>
<td>nominative</td>
<td>No suffix</td>
<td>രാമന്‍/raaman</td>
</tr>
<tr>
<td></td>
<td>accusative</td>
<td>എ‍/e, രാമന്‍/ine</td>
<td>രാമന്‍/raamane, മാധവന്‍/madhuvine</td>
</tr>
<tr>
<td></td>
<td>dative</td>
<td>രാമന്‍/Odu, രാമന്‍/inOdu</td>
<td>രാമന്‍/raivyOdu, മാധവന്‍/madhuvinOdu</td>
</tr>
<tr>
<td></td>
<td>sociative</td>
<td>രാ‍/ikku, രാ‍/INU</td>
<td>രാ‍/sathikku, രാ‍/madhuvimu</td>
</tr>
<tr>
<td></td>
<td>instrumental</td>
<td>രാ/aal, രാ/inaal</td>
<td>രാ/vatiyaal, രാ/madhuvinaal</td>
</tr>
<tr>
<td></td>
<td>genitive</td>
<td>രാ‍/inte, രാ‍/ute</td>
<td>രാ‍/seethayute, രാ‍/mOhante</td>
</tr>
<tr>
<td></td>
<td>locative</td>
<td>ല്‍/il</td>
<td>ല്‍/Kaattil, ല്‍/thaRayil</td>
</tr>
<tr>
<td></td>
<td>derivations</td>
<td>quality</td>
<td>അയക്കു/aaya</td>
</tr>
<tr>
<td></td>
<td></td>
<td>quality</td>
<td>ഉല്ല/ulla</td>
</tr>
<tr>
<td></td>
<td></td>
<td>place</td>
<td>മരതലീ/ile</td>
</tr>
<tr>
<td></td>
<td>adverb</td>
<td>manner</td>
<td>ആയി/aayi</td>
</tr>
<tr>
<td></td>
<td></td>
<td>direction</td>
<td>എക്കെ/ekkz</td>
</tr>
</tbody>
</table>

### 3.2.2 Morphology for verbs

A verb denotes the ‘state of’ or ‘action’ done by a substance. Various admissible forms of verbs can be generated by considering their tense, mood
and aspect information. There are 57 admissible forms of a verb, taking into account various forms of tense, mood and aspect [38,39].

### 3.2.2.1 Inflectional morphology for verbs.

Inflections for verbs occur due to i) tense, ii) aspect iii) mood information. In some languages like Hindi, the verbs have inflections corresponding to the gender, number and person information of the subject. In Malayalam the inflection due to gender and person information of the subject is absent. The inflectional suffixes for verbs are given in Table 3.2.

i) **Tense**

Tense aspects considered are for present and future tenses. The morphology for past tense is not considered.

a) **Present tense**: The inflectional suffix considered is *unnu*.

\[
\text{par} (\text{tell}) + \text{unnu} (\text{present tense marker}) = \text{parayunnu} (\text{is telling})
\]

b) **Future tense**: The suffix which marks future tense is *um*.

\[
\text{poku} (\text{go}) + \text{um} (\text{future tense marker}) = \text{pokum} (\text{will go})
\]

ii) **Aspect**

The aspect factors considered are for a) perfective b) imperfective c) ingressive.

a) **Perfective**: The perfective aspect can be present perfect, past perfect, future perfect, present continuous or past continuous.
Language Morphology and Morphological Analysis

പ (pOyi / went) + കൊട്ടിരുന്ന / (kontirunnu / had been) = 

പിക്കൊട്ടിരുന്ന / (pOyikkontirunnu / had been going)

പ (pOyi/went) + നാട്ട് / (irunnu/had) = പി / (pOyirunnu / had gone)

b) Imperfective: The imperfective suffixes considered are ഉകയാണ് / unnuntu and ഉകയാണ് / ukayaaNz

ട (pokz / go) + ഉ�യാണ് (unnz/will be) = പൊക്കൻ / pokunnuntz (will be going)

c) Ingressive: The ingressive feature indicates beginning of a situation. The suffix used is അരായി / aarayi.

ഉകയാണ് / thutangz + അരായി / aaRaayi = ഉക്കരായി / thutangngaRaayi

iv) Mood

The mood features considered are a) optative b) intentional c) debitive d) ability e) permission f) degree of certainty g) authority of assertion.

a) Optative: This mood indicates a wish for something to happen. The suffix used is atte.

ടട്ടി / rakshikkz + അട്ടി / atte = ടട്ടിക്കട്ടെ / rakshikkatte(let them save)

b) Intentional: This indicates the speaker’s willingness to carry out an action. This indicates a stronger commitment to a future course of action than simple future tense form. The suffix for this mood is അട്ടി / aam. Here the subject should be first person.
c) Debititive: This expresses obligation. This is marked by suffixes anam, eetheeroo.

\[ \text{pokz/go} + \text{aam} = \text{pOkaam/will go} \]

\[ \text{pokz/go} + \text{aNam/should} = \text{pokaNam/should go} \]

\[ \text{vannz/come} + \text{Eetheero/should} = \text{vannEetheero/should come} \]

d) Ability: This indicates the physical ability to perform an action. The suffix used for these are of two types: a verb with the suffix "aam" or an infinitive followed by forms of verbs kazhiyuka, sadhikkuka and okkuka. Both cases expect a dative subject.

\[ \text{kaaN/see} + \text{aam} = \text{kaaNaam/can see} \]

\[ \text{pokz/go} + \text{aankazhinjju/could} = \text{pokaankazhinjju/could go}. \]

e) Permission: This mood grants permission to the addressee. The verbal suffix used is anam with a second person dative subject.

\[ \text{pokz/go} + \text{aam/shall} = \text{pokaam/shall go} \]

f) Degree of certainty: This indicates the degree of certainty with which the speaker makes an assertion. The suffixes used for this are: \text{Ekkam}, umaayirikkum / umaayirikkaam.

\[ \text{vannz/come} + \text{Ekkam/may} = \text{vannEkkaam/may come} \]
Table 3.2 Inflectional morphology for verbs

<table>
<thead>
<tr>
<th>Type of inflection</th>
<th>Imperative forms</th>
<th>Suffix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Past</td>
<td>ഡോ / um</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>ഡോ / unnu</td>
</tr>
<tr>
<td>Aspect</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perfect</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu</td>
</tr>
<tr>
<td></td>
<td>Present</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu</td>
</tr>
<tr>
<td></td>
<td>Past</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu, ഇട്ടു / ittuntuayirunnu</td>
</tr>
<tr>
<td></td>
<td>Future</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu, ഇട്ടു / ittuntakum</td>
</tr>
<tr>
<td></td>
<td>Present continuous</td>
<td>ഇട്ടു / ittuntu</td>
</tr>
<tr>
<td></td>
<td>Perfect</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu</td>
</tr>
<tr>
<td></td>
<td>Imperfective</td>
<td>ഇട്ടു / ittuntu, ഇരിക്കു / irikkunnu, ഇട്ടു / ittuntakum</td>
</tr>
<tr>
<td>Other Auxiliaries</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Optative</td>
<td>അടു / atte</td>
</tr>
<tr>
<td></td>
<td>Intentional</td>
<td>അടു / aam</td>
</tr>
<tr>
<td></td>
<td>Debititive</td>
<td>അടു / aNam, അടു / aEteeru</td>
</tr>
<tr>
<td></td>
<td>Debititive(-ve)</td>
<td>അടു / aNam, അടു / ikkoottu, അടു / aanpaattillu</td>
</tr>
<tr>
<td></td>
<td>Ability</td>
<td>അടു / aam + dative subject</td>
</tr>
<tr>
<td></td>
<td>Permission(+ve)</td>
<td>അരുത് / aruthu</td>
</tr>
<tr>
<td></td>
<td>Permission(-ve)</td>
<td>അരുത് / aruthu, അരുത് / Oloo (only third person)</td>
</tr>
<tr>
<td></td>
<td>Degree of certainty</td>
<td>അരുത് / aruthu, അരുത് / Oloo (only third person)</td>
</tr>
<tr>
<td></td>
<td>Authority for assertion</td>
<td>അരുത് / aruthu, അരുത് / Oloo (only third person)</td>
</tr>
</tbody>
</table>
3.2.2.2 Derivational morphology for Verbs

In this word categories like noun or verb with a suffix attached to it generates a word of new category. The derivations considered are i) participles ii) infinitives. Derivational morphology for verbs are given in Table 3.3.

i) **Participles:** The participles considered are verbal participle, conditional participle, concessive participle and relative participle.

- പാരജിപ്പ് (paRanjja / said-relative participle), 
- പാരണു (pOyaal / if goes-conditional participle)

ii) **Infinitives:** The suffix taken by infinitives is “aan”

- പാര (paRayz / say) + അന്ത് (aan/to) = പാരയാൻ (paRayaan/to say)
- പാര (paRayz / say) + അന്ത് (aan/to) = പാരയാൻ (paRayaan/to say)

3.2.3 Word compounding

In written text any letter can follow another letter. But in spoken language, it depends on the ability of sound generating organs. It is difficult to pronounce when some sounds come together. This is called hiatus (vivruthi). In order to avoid this some changes are made to the sounds so that pronunciation becomes easy. Sometimes, for clarity of meaning or for beauty of sounds also these morphophonemic changes are made. The sound change when two words or suffixes join are called word compounding (sandhi). Malayalam is usually written in a way it is spoken. So Malayalam text contains a lot of compound
Table 3.3 Derivational morphology for verbs

<table>
<thead>
<tr>
<th>Derived form</th>
<th>Derivation</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal participle (+ve)</td>
<td>Past form + പ / u</td>
<td>പരഞ്ഞി / paRanjnu</td>
</tr>
<tr>
<td>Verbal participle (-ve)</td>
<td>Past form + പാ / aathe</td>
<td>പരഞ്ഞകെ / paRayaatthe</td>
</tr>
<tr>
<td>Conditional participle (+ve)</td>
<td>Past form + പാ / aal</td>
<td>പരഞ്ഞാല്ല / paRanjnal</td>
</tr>
<tr>
<td>Conditional participle (-ve)</td>
<td>Root + പാ / aathirunal</td>
<td>പരഞ്ഞാതിരുന / paRayaathirunaal</td>
</tr>
<tr>
<td>Concessive participle (+ve)</td>
<td>Past form + പാ / alum</td>
<td>പരഞ്ഞാലു / paRanjnalum</td>
</tr>
<tr>
<td>Concessive participle (-ve)</td>
<td>Past form + പാ / illenkilum</td>
<td>പരഞ്ഞാലൻ / paRanjnillenkilum</td>
</tr>
<tr>
<td>Relative participle (+ve)</td>
<td>Past form – പ</td>
<td>പരഞ്ഞാ / paRanjna</td>
</tr>
<tr>
<td>Relative participle (-ve)</td>
<td>Root + പ / aaththa</td>
<td>പരഞ്ഞകെ / paRayaaththa</td>
</tr>
<tr>
<td>Infinitive</td>
<td>Root + പ / aan</td>
<td>പരഞ്ഞാ / paRayaan</td>
</tr>
</tbody>
</table>

words. Formation of new words by combining a noun and a noun, noun and adjective, verb and noun, adverb and verb, adjective and noun and in some cases all the words of an entire sentence to reflect the semantics of the sentence are very common [39]. The complexity of compounding in Malayalam language can be understood from the following example:

Malayalam: നീയന്നായെയറാനയേകിട്ടു

Transliteration: njaaninnaleyoraanayekkandu

English: I saw an elephant yesterday
Such sentences are common in Malayalam language. This kind of compounding is found in plenty in media passages and in poems. Sanskrit grammarians like Panini classified word compounding on the basis of the position in which the sound changes occur. According to this classification word compounding is classified into three types: word_medial (padamadhyam), word final (padaanta) and hybrid (ubhaya). In word medial, the compounding occurs between a stem and a suffix. Word final compounding occurs between two words. In hybrid both word medial and word final are involved. Malayalam compounding rules are also classified as: vowel sandhi, vowel-consonant sandhi and consonant-consonant sandhi.

\[ \text{mazhal} (\text{rain}) + \text{alla} (\text{not}) = \text{mazhayalla} \] (vowel sandhi)

\[ \text{thaamara} (\text{lotus}) + \text{kuLam} (\text{pond}) = \text{thaamarakkulam} (\text{lotus pond}) \] (vowel-consonant sandhi)

\[ \text{nel} (\text{rice}) + \text{maNi} (\text{seed}) = \text{nenmaNi} (\text{rice seed}) \] (consonant-consonant sandhi)

Keralapanini has adopted a different classification for the compounding rules based on the changes occur during compounding. They are: \text{LOpa sandhi} (elison), \text{aagama sandhi} (addition), \text{dvitva sandhi} (germination or reduplication) and \text{aadEsa sandhi} (displacement or substitution). \text{LOpa sandhi} is that in which one of the sounds is lost, \text{aagama} is that in which new sound is added, \text{dvitva} is that in which one of the sounds geminates and \text{aadEsa} is that in which one of the sounds is displaced by another sound. The basis of all sound changes in sandhi is ease in pronunciation. It is difficult to pronounce consonants without vowels. When two consonants combine they are pronounced as one. Vowels
have well defined pronunciation. Consequently, sound changes are more essential in vowel combinations. The rules for compounding in each category are given in tables 3.4 to 3.7. The morphophonemic changes at the boundary depends on the ending vowel or consonant, the category of the first or the second word and the beginning vowel or consonant of the second word. The abbreviations used in the tables are: ss(v2)- vowel symbol corresponding to the beginning of the second word. Trans(x)- function which converts N,n,L,R etc. to Na,na,la,Ra respectively. Each entry in the substitution column has three fields. The first and second field gives the number of characters to be removed from the beginning of the first word and from the beginning of the second word. The third field gives the character to be placed at the boundary. The entry (1,1,ss(V2)) indicates when two words are compounded one character from the end of the first word, one character from the beginning of the second word are to be removed and the vowel symbol corresponding to the starting vowel of the second word has to be added at the boundary [40,41].

3.2.3.1 Elision (lopaSandhi)

Elision rules are given in Table 3.4. Elision of sounds occur in the following cases:

i) When followed by any Vowel, ő (unrounded u / chandrakkala) undergoes elision.

\[ \text{thaNupp} (\text{thaNupp} \text{ /chillness}) + \text{uNtz} (\text{uNtz} \text{ /is}) = \text{thaNupp\u0102tzn} (\text{thaNupp\u0102tzn} \text{ /there is chillness}) \]

\[ \text{katt} (\text{kaatt} \text{ /wind}) + \text{tikkunnu} (\text{atikkunnu} \text{ /blows}) = \text{katt\u0102tikkunnu} \]

\( (\text{kattatikkunnu} \text{ /wind blows}) \)
ii) The /u occurring finally in verbs undergoes elision when followed by Vowel. The /u found in these forms, is by origin unrounded ‘u’, added for clarity in pronunciation. Unrounded ‘u’ is pronounced as rounded ‘u’ to provide emphasis to the sentence.

a) Regular

\[
\text{കാൻതൂ} / (kaNtu/saw) + \text{/ഇല} / (illa/no) = \text{കാൻതില} / (kaNtilla/did not see)\\
\text{കാണുന്നു} / (kaaNunnu/see) + \text{/ൻട} / (uNTz/have) = \text{കാണുന്നുന്റ} / (kaaNunnuNTz/is seen)
\]

b) Irregular

\[
\text{കാൻ} / (kaNtu/saw) + \text{/} / (O / ?) = \text{കാൻത} / (kaNtuvO / kaNTO / saw?)
\]

iii) The \text{/എ} / a’ at the end of \text{/ഇല} / alla and \text{/ഇല} / illa, and the the \text{/എ} / i at the end of \text{/ഇല} / aayi and \text{/പി} / pOyi have elision when followed by Vowel.

\[
\text{അല} / (alla/not) + \text{/എന്ന} / (ennu/thus) = \text{അലൻനു} / (allennu/not thus)\\
\text{ഇല} / (illa/no) + \text{/എന്ന} / (ennu/thus) = \text{ഇലൻനു} / (illennu/not thus)
\]

iv) The mid verbal participle also means respectful persuasion. It’s final \text{/എ} / a also has elision.

\[
	ext{വരിക} / (varika/come) + \text{/} / (etO/you) = \text{വരികെ} / (variketO/ you come)
\]
v) The final ‘എ / e’ of the permissive suffix തടു / atte, verbal participle തടീ / aathe, possessive case marker തടു/ute and conjunction തുടെ / uute occasionally have lision when followed by vowel.

(ബ്രിട്ടിഷ് / swamy’s) + (ബ്രിട്ടിഷ്)

(aniyan / brother) = (ബ്രിട്ടിഷ് / swamy’s brother)

### Table 3.4 Elision rules

<table>
<thead>
<tr>
<th>word1 ending</th>
<th>word2 beginning v2</th>
<th>Substitution</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>വ / z</td>
<td>Vowel</td>
<td>(1,1,ss(v2))</td>
<td>കാട്(kaat / forest)+ ഇല്(i/An) = kaattil</td>
</tr>
<tr>
<td>Word1=ഇല്ല / alla, ഇല്ല / illa</td>
<td>Vowel</td>
<td>(-1,ss(v2))</td>
<td>എല്ലാ(alla/not) + എന്ന്(ennz/that) = എല്ലാന്ന്(allaennz)</td>
</tr>
<tr>
<td>Word1=ഇല്ല / alla, ഇല്ല / illa</td>
<td>Vowel</td>
<td>(1,1,ss(v2))</td>
<td>എല്ലാ(alla/not) + എന്ന്(ennz/that) = എല്ലാന്ന്(allaennz)</td>
</tr>
<tr>
<td>Word1= മരു / oru</td>
<td>Vowel</td>
<td>(1,1,ss(v2))</td>
<td>ഒരു(oru/a)+നുമു(oru/um) = ഒരു (oru/um)</td>
</tr>
<tr>
<td>Word1= ഉ / u, ഉ / um, Word1 category = verb</td>
<td>Vowel</td>
<td>(1,1,ss(v2))</td>
<td>ഉന്നു(vannu/came) + ഇല്ലി(illa/not) = ഉന്നില്ല (vannilla)</td>
</tr>
</tbody>
</table>

### 3.2.3.2 Addition (aagamasandhi)

It is inconvenient to pronounce two vowels together as the vowels have independent pronunciation. The difficulty is overcome by inserting യ / ya or വ / va in between them. If the preceding vowel is palatal സ, യ, എ, എൻ, നി, നി,
\( a, \, aa, \, i, \, ii, \, e, \, ee \, or \, ai. \) \( /a, \, o, \, e, \, ee \) or \( /a, \, uu, \, o, \, O \) \( /ya \) is inserted and if it is labial \( (o, \, u, \, oo, \, /u, \, uu, \, o, \, O) \) \( /va \) is inserted. Addition rules are given Table 3.5.

i) Compounding in Palatals

\( (thii/fire) + (aattz/dance) = tiiyaattz \) (fire dance)

\( (kai/hand) + (untz/has) = kaiyuntz \) (has hand)

ii) Compounding in Labials

\( (thata/rub) + (unnu/does) = thatavunu \) (rubs)

\( (chaal/die) + (unnu/does) = chaavunu \) (dies)

ii) In verbs \( /E \) may be substituted by \( /na \). \( /na \) can also occur for making the utterance more pleasing.

\( (kaatti/showed) + (En/I) = kaattiyEn \) or \( kaattinEn / I \) showed

\( (karuthi/think) + (a)=karuthina/karuthiya/(that was thought)

iii) The demonstratives \( a, \, aa, \, i, \, ii \) and \( e \) are referred to as \( chuttezhuththu \) in Tamil. Though these three vowels are palatals \( /v \) instead of \( /y \) is added when followed by a vowel, thus providing exception to the earlier rule.

\( (i/this) + (an/he) = ivan/he \)

It is \( v \) that gets added when this pronounce are used as qualifiers as well.

\( (a/that) + (itam/place) = avitam/place \)
iv) The relative participle suffix വാ / va also is demonstrative. So there too ഔ / va should be added.


t്തി (cheythath/which was done) + അന (an/he) = തത്തവൻ

(cheythavan/he who did)

v) In words ending in the long vowels aa, ii, uu, ee, ai, oo, y or v is generally added. This is confined to a few words. There is no specific grammatical rule regarding this phenomenon.

ഡ ക ് ത (kaay/fruit)

ഡ പ ് യ (paay/mat)

vi) യ/a/ya is added to palatal vowels then followed by suffixes beginning with ക/ka. There is no addition of യ/a/ya if the ക/ka is not the beginning of a suffix or the preceding vowel is not palatal.

ഡ ഥ ല (thal/ka/head) + ് (kku/to) = തലയ്ക്കു (thalykk/to head)

ഡ വ ന (chaati/jump) + ഗ ത ന (katann/crossed) = ചാതിക്കനന (chaatikkatann/crossed jumping)

### 3.2.3.3 Reduplication (dvitvasandhi)

Vowels do not have cluster information. Each vowel is capable of free pronunciation with no obstructions by the tongue in the points of articulation. Even in combination, it is not marked as in the case of consonants. Instead of germination they have length; instead of clustering they have diphthongs. So
Table 3.5 Addition rules

<table>
<thead>
<tr>
<th></th>
<th>Word1 ending/word1</th>
<th>Word2 beginning v2</th>
<th>substitution</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Vowel</td>
<td>Vowel</td>
<td>(-1, ya+ss(v2))</td>
<td>നപയാല (panayola)</td>
</tr>
<tr>
<td>2</td>
<td>ഉ/U, ഊ/uu</td>
<td>Vowel</td>
<td>(-1, va+ss(v2))</td>
<td>മധുവാണ് (madhuvaanz)</td>
</tr>
<tr>
<td>3</td>
<td>Word1=അ /a, ഇ/i, എ/e (chuttezhuththz)</td>
<td>Vowel</td>
<td>(-1, va+ss(v2))</td>
<td>അവര് (avar/they)</td>
</tr>
<tr>
<td>4</td>
<td>ക/ a</td>
<td>Consonant</td>
<td>(-7, ss(aa))</td>
<td>കലാപമള (kalaamELa)</td>
</tr>
</tbody>
</table>

the rules of germination are restricted to consonants. In Malayalam, germination is more in tense consonants and less in lax consonants.

i) When two words combine in which the first is the qualifier and the qualified, the tense consonants initial to the second word geminates.

ഠാ/ (thala/head) + കെറ്റ് (kettu/tie) = ഠാലക്ക്ട് (thalakkettu/turban)

ഠാമര (thaamara/lotus) + കുളം (kuLam/pond) = ഠാമരക്ക്കുളം (taamarakkkuLam /lotus pond)
ii) *dvandvasamasam* does not involve combination of the qualifier and the qualified. So this gemination does not occur in it.

\[
\text{kai} (\text{hand}) + \text{kaal} (\text{leg}) = \text{kaikaal} (\text{hand and leg})
\]

\[
\text{raama} (\text{rama}) + \text{krishNanmaar} (\text{krishna}) = \text{raamakrishNanmaar} (\text{Rama and Krishna})
\]

iv) There will be gemination in sentences also, namely in the tense consonant after the three verbal participles, *mun*, *tan* and *paaksikam*, the pseudo case *e* used in the locative sense and the case suffixes *aal*, *il*, *kal*.

\[
\text{pOyi} (\text{went}) + \text{pRanjnju} (\text{said}) = \text{pooyippaRanjnju} (\text{went and said})
\]

\[
\text{manassaal} (\text{by will}) + \text{kotuththu} (\text{gave}) = \text{Manassaalkkotuththu} (\text{gave by will})
\]

v) The pronouns *a*, *i* and *e* are called *chuttezhuththu*. The term is meaningful, they being the base for words that point out persons. *a-van* (*a* /that), *i-van* (*i* /this) and *ee-van* (*ee* /who). All consonants that follow *chuttezhuththu* are geminated.

\[
\text{a} (\text{that}) + \text{kaalam} (\text{period}) = \text{akkaalam} (\text{that period})
\]

\[
\text{i} (\text{this}) + \text{kanta} (\text{which was seen}) = \text{ikkanta} (\text{this which was seen})
\]
vi) The nasals as well as /y and /l occurring finally in *Ekamaathra* (single beat) bases will geminate. There is no single beat word in Malayalam.

\[\text{ney} \text{/ghee} + \text{aaRu} \text{/river} = \text{neyyaaRu}\]

vii) To words ending in consonants, unrounded /u has to be added. If it is a chillz, this addition is optional.

\[\text{vayas} \text{/age} + \text{kuRanj} \text{nu/ reduced} = \text{vayasukuRanj} \text{nu/age reduced}\]

### 3.2.3.4 Substitution (aadeesasandhi)

i) When consonants of the t-class combine with those of the t-class they become t-class. With the c-class and class they become c-class and # class (ncha, nta, ntha) etc. respectively. Substitution rules are given in Table 3.6.

\[\text{thaN} \text{/cool} + \text{thaar} \text{/flower} = \text{thantaar/cool flower}\]

ii) When l followed by consonants of the t-class, it is substituted by /tt. Similarly when l followed by t-class consonants it is substituted by s/t. This change is called *vinaamam* which means lowering in position.

\[\text{vit} + \text{tu} = \text{vittu/ sold}\]

\[\text{keet} + \text{tu} = \text{kEttu/heard}\]
iii) The ന്/n in the words നുണ/mun, പന്/pin and പന്/pon changes to ല് when followed by voiceless stops.

    പന് (pin/back) + പാട് (paattu/song) = പില്ലപാട് (pilppaattu
    background song)

    പന് (pon/gold) + കട്ട് (kutam/pot) = പൊല്ലകട്ട് (polkkutam/
    golden pot)

iv) The voiceless stops initial to a suffix is substituted by a Nasal except when the suffix begins with ക/ka or ട/ta. The substitution by Nasal according to this rule is called Nasal Assimilation.

    പാറന് (paRan/say) + ച് (chu/past tense) = പാറന്ന്ന് (paRanjnju/said)

    താറ് (Un/to eat) + ഥ് (thu/did) = തുറ (uNtu/ate)

v) 1) Anusvāram(ə / am) becomes clear when vowel follows. In combination with any vowel it changes to ത/ma.

    മരം (maram/tree) + അല്ല് (alla/not) = മരമ്മൽ (maramalla/nottree)

    തെ (uN/to eat) + ഥ് (thu/did) = തുറ (uNtu/ate)

2) When anusvāram is followed by class sounds changes to ക, ചന, ത, മ (nka, ncha, nta, nth, mpa) respectively.
3) **anusvaaram** changes to \( \Omega / v \) when it is followed by *samuchchaya* (additive) *nipaatham* \( \xi_0 \) /um. **anusvaaram** in the future tense marker \( \xi_0 / \text{um} \) also changes to \( v \) when followed by suffix.

\[
\xi_0 \text{(kulam/family)} + \xi_0 \text{(um/and)} = \xi_0 \text{(kulavum/family and)}
\]

\[
\xi_0 \text{(vaaram/week)} + \xi_0 \text{(um/and)} = \xi_0 \text{(vaaravum/week and)}
\]

4) **anusvaaram** changes to \( \Theta / \text{ththa} \) when followed by any suffix beginning with a vowel. If the suffix following is the case marker \( \Theta \gamma / \text{Otu} \) this substitution is not regular.

\[
\xi_0 \text{(dhanam/wealth)} + \xi_0 \text{(e/of)} = \xi_0 \text{(dhanaththe/of wealth)}
\]

\[
\xi_0 \text{(thulaam/support)} + \xi_0 \text{(inte/of)} = \xi_0 \text{(thulaaththinte/support of)}
\]

5) \( \omega / \) am changes to \( \varepsilon / \text{n} \) when followed by plural marker \( \xi_0 / \text{kaL} \), due to assimilation.

\[
\xi_0 \text{(maram/tree)} + \xi_0 \text{(kaL/plural)} = \xi_0 \text{(marankaL/marangngaL/trees)}
\]
The assimilation is found also in

\[ \text{kuLam} / \text{pond} + \text{kara} / \text{shore} = \text{kuLangngara} / \text{pond’s shore} \]

<table>
<thead>
<tr>
<th>Word1 ending</th>
<th>Word2 beginning</th>
<th>Substitution</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \infty ) / am</td>
<td>Vowel(v2)</td>
<td>(1,1, 1 (/ \text{ma} + ss(2)))</td>
<td>( \text{maram} + \text{alla} = \text{maramalla} )</td>
</tr>
<tr>
<td>( \infty ) / am</td>
<td>Vowel(v2), word class=case marker</td>
<td>(1,1, 1 (/ \text{tha} + ss(2)))</td>
<td>( \text{dhan} + \text{e} = \text{dhanaththe} )</td>
</tr>
<tr>
<td>( \text{N} ) / N</td>
<td>( \text{nta, ntha, na} )</td>
<td>(1,1, 1 (/ \text{nta, ntha, nNa} ) resp.)</td>
<td>( \text{thN+thaar=thaNthaar} )</td>
</tr>
<tr>
<td>( \infty ) / am</td>
<td>( \text{nka, ncha, nta, nth, mpa, nnga} )</td>
<td>( \text{maram+kaL=maranggaL} )</td>
<td></td>
</tr>
<tr>
<td>( \text{N, n, L, l, r} )</td>
<td>Vowel(v2)</td>
<td>(( \infty ), 1 (/ \text{Na, na, La, la, ra} ) resp.+ss(2))</td>
<td>( \text{avar +il= avaril} )</td>
</tr>
</tbody>
</table>

### 3.2.3.5 Sanskrit compounding

Since Malayalam had adopted many words from Sanskrit many of the compounding rules found in Sanskrit are common in Malayalam sentences. Sanskrit compounding is classified as vowel compounding (swarasandhi) which happens when two vowels join and consonant compounding (vyanjanasandhi) which happens when two consonants join. Vowel sandhi is
further classified into i) deerghasandhi, ii) guNasandhi, iv) vridhisandhi, and iii) yaNsandhi. Sanskrit compounding rules are given in Table 3.7.

i) Deerghasandhi

When the vowels /a/, /i/, /u join with /a, /aa, /i, /ii, /u, /uu respectively, the vowels change to /aa, /ii, /uu respectively. Most common for compound nouns which join two nouns or between an adjective and a noun.

\[
\text{ಕದha} (kadha/story) + \text{ഞ്യം} (anthyam/end) = \text{കദha�്യം/story ending}
\]

\[
\text{കഥ} (katam/loan) + \text{ഞ്യം} (aasvaasam/help) = \text{കഥ�്യം/relief loan}
\]

ii) GuNasandhi

When /a, /aa/ is followed by /i, /ii, /u, /uu it changes to /E, /O respectively.

\[
\text{മഹa} (maha/great) + \text{ഇണ്ട} (indran/indra) = \text{മഹഇണ്ട/mahEndran/great Indra})
\]

\[
\text{കാ} (kala/art) + \text{ഉപാസ} (upaasana/devotion) = \text{കാുപാസാ/devotion for art}
\]

ii) Vridhisandhi

a) When /a, /aa/ is followed by /E, both sounds join to form /ai.
b) When ഐ / a, ആ / aa is followed by ഓ / oo, both sounds join to form ഔ / ou.

നെയ്യ (sada/always) + Evam (all) = സദൈവം (sadaivam /always)

iii) YaN sandhi

When ഐ / e is followed by ഐ / a, ആ / aa, ഇ / y is added. When ഐ / e is followed by ഐ / u, ഊ / uu, ഑ / va is added. It is a form of addition compounding (aagamasandhi) in Malayalam.

ാസ് (athi/very) + വാസ്യം (aavaSyam/need) = ആസ്യാസ്

(athyavaSyam/most needed)

<table>
<thead>
<tr>
<th>Table 3.7 Sanskrit compounding rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wor d1 end</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

47
3.3 Morphological analysis

Morphological analysis is the first phase in any natural language processing application. The output of this phase is used by the syntax analysis phase following it. Morphological analyser should accept as input a surface form of a word in a language such as spies and return an underlying form divided into morphemes, namely spy+s, plus a gloss string such as N+PLURAL [37]. Here, N and PLURAL are the category information regarding the morphemes spy and s, the constituents of the input.

3.3.1 Two level morphology

Kimmo Koskenniemi in 1983 proposed a two level morphology for word form recognition and generation [42,43]. The model was based on the traditional distinction that linguists make between morphotactics and morphophonemics. For example, the word chased is analyzed morphotactically as the stem chase followed by the suffix -ed. The addition of the suffix -ed apparently causes the loss of the final ‘e’ of chase; thus chase and chas are allomorphs or alternate forms of the same morpheme.

Two level morphology is based on three ideas:

   i) Rules are symbol-to-symbol constraints that are applied in parallel, not sequentially like rewrite rules.

   ii) The constraints can refer to the lexical context, to the surface context, or to both contexts at the same time.

   iii) Lexical lookup and morphological analysis are performed in tandem.
In Koskenniemi's two level model a word is represented as a direct, letter-for-letter correspondence between its lexical or underlying form and its surface form. For example, the word chased is given this two-level representation (where + is a morpheme boundary symbol and 0 is a null character):

Lexical form: c h a s e + e d
Surface form: c h a s 0 0 e d

Koskenniemi’s two-level morphology was the first practical general model in the history of computational linguistics for the analysis of morphologically complex languages. Karttunen et.al completed the project and published a collection of papers on the topic, along with Lisp code. They called it the KIMMO system and it inspired many other KIMMO implementations. The most popular of these is PC-KIMMO, a free C implementation from the Summer Institute of Linguistics [44,45,46,47].

The KIMMO parser had two analytical components: the rules component and the lexical component, or lexicon. First, the rules component consisted of two-level rules that accounted for regular phonological or orthographic alternations, such as chase versus chas. Second, the lexicon listed all morphemes (stems and affixes) in their lexical form and specified morphotactic constraints. The Generator would accept as input a lexical form such as spy+sand return the surface form spies. The Recognizer would accept as input a surface form such as spies and return an underlying form divided into morphemes, namely spy+s, plus a gloss string such as N+PLURAL.

But it had a serious deficiency: it could not directly determine the part of speech of a word or its inflectional categories. For example, given the word enlargements, PC-KIMMO could tokenize it into the sequence of morphemes
en+large+ment+sand gloss each morpheme, but it could not determine that the entire word was a plural noun. This meant that PC-KIMMO was not adequate to act as a morphological front end to a syntactic parser its most desirable application.

3.3.2 Computational models for morphological analysis

Various NLP research groups have developed different algorithms and data structures for morphological analysis. Some of the algorithms are language dependent and some of them are language independent. The various techniques are:

1) Finite State Transducers (FST).
2) Suffix stripping approach
3) Corpus Based Approach
4) Paradigm Based Approach.

3.3.2.1 Finite state transducers

An FST is represented as a two tape automaton. We can combine lexicon, orthographic rules and spelling variations in the FST to build a morphological analyzer. The simplest finite state machine is a finite state automaton (FSA), which recognizes (or generates) the well-formed strings of a regular language.

A Mealy machine extension FST can be formally defined as:

\[ Q \] : a finite set of \( N \) states \( q_0,q_1,q_2,\ldots,q_n \)

\[ \Sigma \] : a finite alphabet of complex symbols. Each complex symbol is composed of an input output pair \( i:o \); one symbol \( i \) from an input
alphabet I and one symbol o from an output alphabet O. I and O may each include ε.

Q0: the start state

F : the set of final states, F С Q

δ (q,i:o): the transition function between states. δ is a relation from Q X Σ to Q

A sample transducer which can work as a morphological parser for English nominal number inflection is shown in Figure 3.1. When the states are traversed from start state to final state with an irregular noun like geese the output will be goose + N + PL.

Figure. 3.1 Transducer for English nominal number inflection

A word grammar based morphological analyzer has been discussed [48]. The model integrates the two level method and a unification based formalism. The
system was tried for Basque texts. It separates the sequential and non sequential constraints. Sequential morphotactic constraints are applied in the segmentation phase and the non sequential constraints in the final feature recombination phase. The two level formalism takes care of the morphographemics and sequential morphotactics. Unification based word grammar combines the grammatical information defined in morphemes to tackle complex morphotactics. This design allowed full coverage analyzer that processes efficiently unrestricted texts in Basque.

A Unification based method has been discussed [49,50] for morpho syntactic parsing of agglutinative and inflectional languages. The system called Humor 99 has already been integrated with a variety of industrial applications. It can handle almost all agglutinative languages like Hungarian, Turkish, Estonian etc. and inflectional languages like Polish, Czech, German etc. very effectively.

Ramaswamyveerappan et. al. has developed a Kannada Morphological analyzer and generator using Trie [51,52,53]. The tool uses a trie for storing roots and suffixes. It combines paradigm approach with suffix stripping. No full fledged Morphological analyzer for Kannada has been developed and the performance of the tool is encouraging.

Girinath Jha has developed an inflectional Morphology analyzer for Sanskrit [54]. It uses sandhi free Sanskrit Unicode text as input. The system labels each word as subanta, tinanta and avyaya. There after each of the category is further categorized using a stem and suffix database.

Kemal Oflazer has discussed about a lenient morphologic analyser [55]. The system augments the two level morphology rule so that word forms with
violations of some of the two level constraints can be analysed and ranked. The problem was motivated by the languages Turkish and Basque. In Turkish, the imported words violate the assumption of one correspondence between pronunciation and orthography and Basque words usually have errors.

### 3.3.2.2 Corpus Based Approach

Corpus is a large collection of written text belong to a particular language. Raw corpus can be used for morphological analysis. It takes raw corpus as input and produces a segmentation of the word forms observed in the text. Such segmentation resembles morphological segmentation.

Morfessor1.0 developed in Helsinki University is a corpus based language independent morphological segmentation program.

The LTRC Hyderabad successfully developed a corpus based morphological analyzer [56]. The program combines paradigm based approach as well corpus based approach. It uses a low coverage morphological analyzer and combines unsupervised learning of paradigm of unknown words using a corpus. When the morph analyzer cannot analyze a given word the morph guessing package returns a pair of stem and paradigm. The guessing is based on suffixes. Thus it improves the performance of a morphological analyzer from 32% to 63%.

Unsupervised learning of morphology for building Lexicon for Assamese, a highly inflectional language [57,58,59,60]. It tries to create a dictionary and a morphological rule base using a text corpus which was developed for the purpose.
A statistical method using N gram approach has been tried for modeling morphologically rich languages using split words and unstructured dependencies [61,62]. The method assumes that n-1 tokens that determine the probability of a given token can be chosen anywhere in the sentence rather than preceding n-1 positions. The system reduces the perplexity of the standard N gram model by 24%. It was designed for Turkish.

3.3.2.3 Paradigm Approach

A paradigm defines all the word form of a given stem and also provides a feature structure with every word form. The paradigm based approach is efficient for inflectionally rich languages. The ANUSAARAKA research group has developed a language independent paradigm based morphological compiler program for Indian Languages.

This or a variant of this scheme has been used widely in NLP. The linguist or the language expert is asked to provide different tables of word forms covering the words in a language. Each word-forms table covers a set of roots which means that the roots follow the pattern (or paradigm) implicit in the table for generating their word forms. Almost all Indian language morphological analyzers are developed using this method. Based on paradigms the program generates add delete string for analysis. Paradigm approach rely on findings that the different types of word paradigms are based on their morphological behavior.

Words are categorized as nouns, verbs, adjectives, adverbs and postpositions. Each category will be classified into certain types of paradigms based on their morphophonemic behavior. For example noun ‘maram’(tree) belongs to a
paradigm class is different form ‘vanitha’ (lady) which belongs to a different paradigm class as they behave differently morphophonemically.

- *maram, maraththe, maraththil, maraththinte, maraththOtu, maraththilninnum*
- *vanitha, vanithaye, vanithayil, vanithayute, vanithayOtu, vanithayilninnum*

There is a Dictionary of Roots, where along with the roots the types and grammatical information that is common to all the associated endings (that is, word forms), can be stored. This leads to efficient storage because there is only one paradigm table for a class of roots rather than separate word forms table for each root. A sample dictionary of roots is shown in Table 3.8. The paradigm for the word given in first column is given in the second column and its gender is given in the third column.

**Table 3.8 Dictionary of roots**

<table>
<thead>
<tr>
<th>Root</th>
<th>Type</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>aabharaNam</td>
<td>(n, maram)</td>
<td>N</td>
</tr>
<tr>
<td>katam</td>
<td>(n, maram)</td>
<td>N</td>
</tr>
<tr>
<td>rama</td>
<td>(n, vanitha)</td>
<td>F</td>
</tr>
</tbody>
</table>

Every language has its own dictionary of indeclinable words. Using this dictionary, paradigm tables and dictionary of roots, one can generate a word form when a root and desired feature values are given. The dictionary of roots
should be kept sorted since searching for an item in a sorted list is much faster. It is set up so that given a root and the desired features; one can locate the right table and then lock up the right entry.

### 3.3.2.4 Suffix Stripping

Suffix Stripping is another one method used for analyzing the words in a language. In highly agglutinative languages such as Malayalam, a word is formed by adding suffixes to the root or stem. Morphologically highly complex words exist in such languages, which are formed by continuously adding suffixes to the stem. Suffix Stripping method make use of this property of the language, i.e., having complex suffixes attached to the stem. Once the suffix is identified, the stem of the whole word can be obtained by removing that suffix and applying proper morphophonemic rules.

Thus the general format of the morphological analyzer of Malayalam is Word stem + suffixes. The two main grammatical category of Malayalam are Noun and Verb. Stem is either a verb stem or a noun stem. The suffix stripping method makes use of a stem dictionary (for identifying a valid stem), a suffix dictionary, containing all possible suffixes that nouns/verbs in the language can have (to identify a valid suffix), morphotactics rules and morphophonemic rules or sandhi rules. Nouns can have case markers as their suffixes. Normally verbs inflect for tense, aspect and mood. Thus the verbal forms are stripped into suffixes which denote different tenses, moods and aspects. For example, the verbal form paaTikkoNTirunnu will be analysed as follows:

- **i** - ‘conjunctive particle’
- **koNtur** - ‘continuous aspect’
- **nnu** - ‘past tense’
- **paaT** - ‘sing’
C-DAC, Thiruvananthapuram (former ER&DC), has developed a morphological analyzer for Malayalam using suffix stripping approach.

Sumam, et. al., Cochin University of Science and Technology also has developed a morphological analyzer using the same approach. It has a morphological analyzer which categorises the words based on the suffixes. For nouns the morphology due to gender, number and case are considered. For verbs, the morphology due to tense, aspect and mood are considered. The system mainly handles inflectional morphology for verbs and nouns. Some derivational morphology for nouns and verbs are also considered [38].

It use separate tables for the lexical categories nouns, pronouns, verbs and modifiers. The information stored with each category are:

- **Noun**: root, gender, property (human, non human), type (abstract noun, collective noun, common noun)
- **Pronoun**: root, gender, number, person, type (personal, reflexive, interrogative, indefinite)
- **Verb**: root, kaaritham / akaaritham

The system has not considered compound nouns. It fails to represent linguistic generalization. Use of a separate table for suffixes saves lexicon space.

Rajeev R.R. et. al., Kerala University, has developed a morphological analyzer for Malayalam as part of a Malayalam- Tamil translator. They also have used a suffix stripping approach. It does not have a sandhi splitter to split into
morphemes when more than one lexical category is joined to form a word. It takes as input a Malayalam word and finds the root and suffix from the given word. It then converts the root and suffix into Tamil using a bilingual dictionary and runs a morphological generator for Tamil to find the corresponding word in Tamil. The output is generated in Tamil script using Unicode [63].

Other existing morphological analysers and generators are: TelMOre for Telugu [64], Bangla morph generator [65], Tamil morphological analyzer [66] and Arabic Morph generator from interlingua [67].

3.4 Conclusion

This chapter discussed the various approaches used for morphological analysis for natural language sentences. The survey revealed that very few works are reported on morphological analysis for Malayalam language. The morphological variations found in Malayalam language also were discussed in this chapter. The word compounding rules identified were used in the morphological analysis phase of the prototype machine translation system. The same set of compounding rules is used in forming inflections and derivations of a word. The following chapter discusses the computational models for parsers and the various activities of the parsing phase of a machine translation system.