CS6770 - Knowledge Representation & Reasoning
MALAYALAM TO CD THEORY CONVERSION

(CS13M020 - GOVIND C, CS13M044 - SANDEEP NK)

April 23, 2014

Objective of the assignment :

CD theory

1. lexicon construction ,and

2.parsing of malayalam language.

To create malayalam lexicon for the domain 'election2014' and to create a parser for processing simple malayalam sentences and convert it to CD representation.

Challenges Faced :

Choosing the correct words for creating the malayalam lexicon,corresponding to a single word in english there are multiple words in malayalam.The most difficult challenge was to derive a formal specification for grammar.Malayalam has a very vast grammar structure.Unlike english no CFG like grammar specification is available for malayalam. Malayalam is one of the languages where movement phenomena among sentence structures is very common. So while parsing malayalam sentences it will be too difficult to include all the cases that corresponds to one English sentence .

The next difficult challenge in this implementation is also attributed to the structure of malayalam language - the syntax and semantics of the language is inseparably tangled together.

Language Used :

       JAVA has been chosen as the language for implementation.

Implementation issues :

       For taking malayalam sentences as input we used unicode.Like all natural languages malayalam is a morphologically complex language.So we have many forms of a single word that fits in different contexts.Therefore the parser needs to accomodate a large number of cases.So we have implemented parser for a subset which covers the most common sentence structures in this domain.

Details :

Input :

       The input for the program is a sentence in malayalam typed using malayalam font(using either Google input tools or SCIM Keyboard input).Multiple sentences can also give as input separated by full stop (.) . New words can be added to lexicon by simply adding the word to corresponding mapping text files in format used in those files.

Output :

       The CD representation of the corresponding sentences.

Execution of the program :

Run the main.java file in the src folder in eclipse.Please go through README file for further information.

Program will take the words from files and create the lexicon according to the format given in files also we make mapping of malayalam words to corresponding type.Then we will parse the sentence in malayalam.We will take a compound sentence,break it into simple sentence which are processed independently.From the sentence we extract the verb phrase and according to the type(MTRANS,INGEST,...),We will convert it into CD form.

Testing of program:

        We tested the CD form generated from a Malayalam sentence with others(CD to English).We got sentences with same\similar meaning in English.

Some of the sample sentences and their CD representation is given below :

Input sentence: മോഡി പായസം തിന്നു .
English Translation: Modi ate payasam .

CD-FORM : (INGEST
   (ACTOR (PERSON (NAME Modi) (POSSBY NULL)))
   (OBJECT (OBJECT (NAME payasam) (POSSBY NULL)))
   (CONC_TENSE PAST)
   (USING NULL)
)

Input sentence: മോഡി പായസം സ്പൂണ്ൻ കൊണ്ട് തിന്നു .
English Translation: Modi ate payasam using spoon .

CD-FORM : (INGEST
   (ACTOR (PERSON (NAME Modi) (POSSBY NULL)))
   (OBJECT (OBJECT (NAME payasam) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING (OBJECT (NAME spoon)(POSSBY NULL)))

)

Input sentence: മോഡി സ്പൂണ്‍ കൊണ്ട് പായസം തിന്നു

English Translation: Same as above,But different usage of same sentence in malayalam .

CD-FORM : (INGEST

(ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

(OBJECT (OBJECT (NAME payasam) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING (OBJECT (NAME spoon)(POSSBY NULL)))

)

Input sentence: മോഡി മോഡിയുടെ സ്പൂണ്‍ കൊണ്ട് പായസം തിന്നു

English Translation: Modi ate payasam using modi's spoon .

CD-FORM : (INGEST

(ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

(OBJECT (OBJECT (NAME payasam) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING (OBJECT (NAME spoon)(POSSBY (PERSON (NAME Modi)(POSSBY NULL)))))

)

Input sentence: മോഡി നഗ്മയുടെ ചുവന്ന സ്പൂണ്‍ ഉപയോഗിച്ച് പായസം തിന്നു

English Translation: Modi ate payasam using Nagma's red spoon .

CD-FORM : (INGEST

(ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

(OBJECT (OBJECT (NAME payasam) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING (OBJECT (NAME spoon)(POSSBY (PERSON (NAME NAGMA)(POSSBY NULL)
))(ISA(ADJECTIVE red))))

)

Input sentence: കെജ്രിവാള്‍ ഡല്‍ഹിയിലേക്ക് പോയി

English Translation: Kejriwal went to delhi .

CD-FORM : (PTRANS

(FROM NULL)

(TO (PLACE (NAME delhi)))

(OBJECT (PERSON (NAME Kejriwal) (POSSBY NULL)))

(ACTOR (PERSON (NAME Kejriwal) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING NULL)

)

Input sentence: മോഡി ഡല്‍ഹിയിലേക്ക് കാര്‍ ഉപയോഗിച്ച് പോയി

English Translation: Modi went to Delhi using car .

CD-FORM : (PTRANS

(FROM NULL)

(TO (PLACE (NAME delhi)))

(OBJECT (PERSON (NAME Modi) (POSSBY NULL)))

(ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

(CONC_TENSE PAST)

(USING (OBJECT (NAME car)(POSSBY NULL)))

)

Input sentence: മോഡി ഡല്‍ഹി വരെ പോയി

English Translation: Modi went to Delhi .(using different usage)

CD-FORM : (PTRANS

  (FROM NULL)

  (TO (PLACE (NAME delhi) (POSSBY NULL)))

  (OBJECT (PERSON (NAME Modi) (POSSBY NULL)))

  (ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: ബോബിലി നിന്നും മോഡി ഡല്‍ഹി വരെ പോയി

English Translation: Modi went to Delhi from Bobbili .

CD-FORM : (PTRANS

  (FROM (PLACE (NAME Bobbili) (POSSBY (PERSON(NAME Modi) (POSSBY NULL)))))

  (TO (PLACE (NAME delhi) (POSSBY NULL)))

  (OBJECT (PERSON (POSSBY (PERSON(NAME Modi) (POSSBY NULL)))))

  (ACTOR (PERSON (POSSBY (PERSON(NAME Modi) (POSSBY NULL)))))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: മോഡി നാമനിര്‍ദ്ദേശപത്രിക സമര്‍പ്പിച്ചു .

English Translation: Modi submitted nomination .

CD-FORM : (ATRANS

  (FROM (PERSON (NAME Modi) (POSSBY NULL)))

  (TO (PERSON (NAME Modi) (POSSBY NULL))))

  (ACTOR (PERSON (NAME Modi) (POSSBY NULL))

  (CONC_TENSE PAST)

  (USING NULL))

Input Sentence: കെജ്രിവാള് ഡല്ഹി പര്സംഗിച്ചു .

English Translation: kejriwal delivered a speech in Delhi .

CD-FORM : (SPEAK

  (FROM (PERSON (NAME Kejrival)(POSSBY NULL)))

  (TO (PERSON (NAME delhi)(POSSBY NULL)))

  (OBJECT NULL)

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: റിക്ഷാഡൈര്വര് കെജ്രിവാള് അടിച്ചു .

English Translation: Rickshaw-driver hit Kejriwal .

CD-FORM : (PROPEL

  (FROM (PERSON (NAME rikshawdriver) (POSSBY NULL)))

  (TO (PERSON (NAME Kejriwal) (POSSBY NULL)))

  (PHYSCONT ((PERSON (NAME Kejriwal) (POSSBY NULL))))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: മോഡി ജനങ്ങള് നേരെ കൈ വീശി

English Translation: Modi waved his hands against people .

CD-FORM : (PROPEL

  (FROM (PERSON (NAME Modi) (POSSBY NULL)))

  (TO (OBJECT (NAME people))

  (OBJECT (OBJECT (NAME hand) (POSSBY NULL)))

  (PHYSCONT ((OBJECT (NAME people) (OBJECT (OBJECT (NAME hand) (POSSBY NULL)

))))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: മോഡി കൈ വീശി.

English Translation: Modi waved his hands .

CD-FORM : (PROPEL

  (FROM (PERSON (NAME Modi) (POSSBY NULL)))

  (TO)

  (OBJECT (OBJECT (NAME hand) (POSSBY NULL)))

  (PHYSCONT ((OBJECT (NAME hand) (POSSBY NULL))))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentences: മോഡി മൈക്ക് കൈക്കലാക്കി.

English Translation: Modi grabbed the mic .

CD-FORM : (GRASP

  (ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

  (OBJECT (OBJECT (NAME mic) (POSSBY NULL)))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: അടികിട്ടിയ കെജ്രിവാള് കരഞ്ഞു .

English Translation: Kejriwal got beaten and he cried .(in malayalam we can use 'got beaten' as adjective using one word.)

CD-FORM : (EXPEL

  (FROM (PERSON (NAME Kejriwal) (POSSBY NULL)))

  (OBJECT (NAME WATER) (FROM EYES))

  (CONC_TENSE PAST)

  (USING NULL)

  (ISA (ADJECTIVE beaten)))

Input sentence: മോഡി കരഞ്ഞു .

English Translation: Modi cried .

CD-FORM : (EXPEL

  (FROM (PERSON (NAME Modi) (POSSBY NULL)))

  (OBJECT (NAME WATER) (FROM EYES))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: മോഡി ഡല്ഹി കണ്ടു .

English Translation: Modi saw Delhi .

CD-FORM : (ATTEND

  (ACTOR (PERSON (NAME Modi) (POSSBY NULL)))

  (ORGAN EYE)

  (OBJECT (PERSON (NAME delhi) (POSSBY NULL)))

  (CONC_TENSE PAST)

  (USING NULL)

)

Input sentence: ജനങ്ങള് മോഡിയുടെ പര്സംഗം ശര്ദ്ധയോടെ കേട്ടു .

English Translation: people carefully heard modi's speech .

CD-FORM : (ATTEND

   (ACTOR (PERSON (NAME people) (POSSBY (PERSON(NAME Modi) (POSSBY NULL)))

))

  (ORGAN EAR)

  (OBJECT (OBJECT (NAME speech) (POSSBY NULL)))

  (CONC_TENSE PAST)

  (USING NULL)

  (ISA (ADVERB carefully)))

Input: കെജ്രിവാള് ഡല്ഹിയിലേക്ക് പോയി അതുകൊണ്ട് കെജ്രിവാള് ഡല്ഹി കണ്ടു

English Translation: Kejriwal went to Delhi so he has seen Delhi .(so,because,that,and all these are represented by same word in malayalam 'Athu Kondu' (അതുകൊണ്ട്).)

CD-FORM : (LEADSTO

 (PTRANS

  (FROM NULL)

```
  (TO  (PLACE (NAME delhi )  ))
  (OBJECT (PERSON (NAME Kejriwal ) (POSSBY NULL) ) )
  (ACTOR (PERSON (NAME Kejriwal ) (POSSBY NULL) ))
  (CONC_TENSE PAST)
  (USING NULL)
  )


 (ATTEND
  (ACTOR (PERSON (NAME Kejriwal ) (POSSBY NULL) ))
  (ORGAN EYE)
  (OBJECT (PERSON (NAME delhi ) (POSSBY NULL) ))
  (CONC_TENSE PAST)
  (USING NULL)
  )


 (CONC_TENSE PAST)
 (USING NULL))
```

Input sentence: റിക്ഷാഡൈർവർ മോഡി അടിച്ചു അതുകൊണ്ട് മോഡിയുടെ ചോരവാർന്നുപോയി.

English Translation: Modi got beaten by rickshaw driver and he bleed .

```
CD-FORM : (LEADSTO
 (PROPEL
  (FROM (PERSON (NAME rickshawdriver ) (POSSBY NULL) ))
  (TO (PERSON (NAME Modi ) (POSSBY NULL) ))

  (PHYSCONT ((PERSON (NAME Modi ) (POSSBY NULL) ) ))
  (CONC_TENSE PAST)
```

```
 (USING NULL)
  )

 (LEADSTO
 (EXPEL
  (FROM (PERSON (NAME Modi ) (POSSBY NULL) ))
  (OBJECT (NAME BLOOD) (FROM BODY))
  (CONC_TENSE PAST)
  (USING NULL)
  )
 (CHANGE (OBJECT (PERSON (NAME Modi ) (POSSBY NULL) )) (STATE PHYSICAL_STATE
(INITIAL 1009) (FINAL 21) ) (CONC_TENSE PAST) (USING NULL))
 (CONC_TENSE PAST)
 (USING NULL))

 (CONC_TENSE PAST)
 (USING NULL))
```

Further Improvements:

- We have implemented parser for handling only some commonly used sentence structure in malayalam,that too easy to represent and understand in CD representation.
- Lexicon are also limited,we can add some more complex verbs and handle it.
- As malayalam is a morphologically complex language,its not possible to implement all the possible cases in a limited time.
- As in English,meaning of Malayalam words depends on the context.we are handling most general cases.