Computational Linguistics 1
Mid-Semester Exam
Spring 2022

Total Marks: 50
Time: 90 minutes

Answer all the questions

1. What is morph analysis? Why is it important in NLP applications like Machine Translation or Information Retrieval? Explain with suitable examples.
[5 marks]

2. (a) Give a letter-based finite state automaton to recognize the following words (and only the following words): [5 marks]

Geeta, Gita, Githa, Geetha, Sita, Sitha, Seeta, Seetha, Geet, Rita

(b) Give an FSA to generate the numbers 'one' to 'fifty'. The inputs may be words and need not be single letters. The FSA should not generate any other number.
[5 marks]

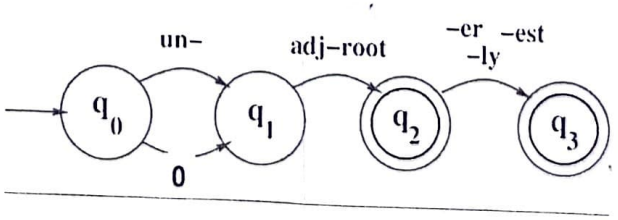
3. (a) Give the lexical forms for each of the 5 wordforms for the following verbal lexemes:
[3 marks]

PULL, RUN, DRINK

Eg. pull +V +prog = pulling

(b) Give an FST to generate all the 5 wordforms of the verbs PULL, RUN, DRINK i.e. generate the surface forms for the lexical forms you gave in (a)
[6 marks]

4. See the FSA to handle adjectives in English. (a) Give any 8 examples of overgeneration by this which are unacceptable (eg. unsad, unsadly).
[2 marks]



(b) Modify the machine to avoid generation of wrong adjectival forms.
Hint: Use lists for types of adjectives. Each list should have 3-4 adjectives
[4 marks]

5. (a) Give 4 different paradigm classes (with 2 members in each class) for nouns in your mother tongue. Write them in WX notation. [4 marks]

Eg. Hindi: ladakA – ladakA, kamarA
Telugu: Avu – Avu, kukka

(b) Give the paradigm table (singular, plural, direct and indirect cases) for one member from each of the 4 classes as exemplified below:
[4x2=8 marks]

Example in Hindi:

ladakA 'boy'	**Singular**	**Plural**
Direct case	ladakA	ladake
Indirect case	ladake	ladakoM

Example in Telugu:

Avu 'cow'	**Singular**	**Plural**
Direct case	Avu	Avulu
Indirect case	Avu	Avula

(c) Give add-delete rules for the paradigm table generated above to get the base form.
Example: [0.5x16=8 marks]

ladakA [0,0,0] = ladakA
ladake [1, 1, A] = ladakA
ladake [1, 1, A] = ladakA
ladakoM [2, 1, A] = ladakA