TURKISH LANGUAGE

- Server Çimen

–

- Middle East Technical University
- Department of Computer Engineering

Outline

- Basic characteristics of Turkish
- Turkish Language Reform
- Current status of Turkish in GF
- Future work

Basic Characteristics of Turkish

- Family:
  - Altaic
    - Turkic
      - Southwestern Turkic (Oghuz)
        - Western Oghuz
        - Turkish
- Writing system: Turkish variant of Latin Alphabet
  - Extra letters: Ç,Ğ,İ,Ö,Ş,Ü
  - Excluded letters: Q,W,X

Basic Characteristics of Turkish

- Vowel Harmony: Two dimensional
    - Front/Back Harmony (a,ı,u,o vs. e,i,ö,ü)
    - Rounding Harmony (a,ı,e,i vs. o,u,ö,ü)
- Extensive Agglutination
    - Almost only by suffixes
        - Inflecting words
        - Contructing new words
    - The longest Turkish word is Çekoslovakyalılaştıramadıklarımızdanmışsınız meaning "You are said to be one of those that we couldn't manage to convert to a Czechoslovak"

Basic Characteristics of Turkish

- Normal Word Order: Subject Object Verb
  - It is possible to arrange words in different orders to stress a certain word or phrase.
- No Gender
- Noun has cases.
  - Nominative, Accusative, Genitive, Dative, Ablative, Locative
  - There is no definite article, defineteness is implied with accusative case.
- Adjectives
  - precedes noun
  - has no declension, but they can be used as nouns.

Basic Characteristics of Turkish

- Verb
  - Person
  - Number: Singular Plural
  - Tense: Present, Past, Inferential, Future, Aorist
  - Mood: Conditional, Imperative, Necessitative, Optative
- T-V distinction

Turkish Language Reform

- *Heavy influence of Arabic and Persian during Ottoman Empire (until 1932)*
- *Script Reform (1928)*
- *Language reform (1932):*
  - Scan the Turkish Literature for pure Turkish words and grammar rules
  - Collect pure Turkish words from daily language
  - Replace the Arabic and Persian borrowings with Turkish equivalents

Turkish Language Reform

- Vocabulary problem: Today both old word and neologism (and sometimes even a western borrowing) exist:
  - word = kelime = sözcük
  - old   = ihtiyar = yaşlı
  - active = faal   = etkin   = aktif

Current Status of Turkish in GF

- Vowel Harmony,
- Numbers,
- Inflection of regular verbs,
- Inflection of regular nouns and lexicon (partially) are currently implemented.

Current Status of Turkish in GF

- But functions that implement vowel harmony suffer performance problems
- Solution:
  - Adding a new field in nouns to store harmony type so that harmony type won't be computed more than once and
  - Introducing a new type for suffixes to store all inflection of a suffix for every harmony type to avoid unnecessary computations, again.

Future Work

- After solving the performance problem and implementing more general noun morphology, I am planning to convert Zemberek's (a spellchecker for Turkish) lexicon to GF code.

Questions