Language in 10 minutes: Turkish

Adi Renduchintala
Outline

- Demographics
- History
- Phonology (very briefly)
- Grammar
  - Nouns, Verbs, Adjectives
  - Negation
  - Word Order
  - Unique aspects of Morphology & Syntax
- Machine Translation
  - Challenges
  - Corpora
  - State of the art
Demographics

- 63 million speakers
- Apart from Turkey, spoken in Bulgaria, Cyprus, Greece, Macedonia, Romania, and Serbia.
- 2 million speakers in Germany, and large Turkish speaking population in USA
History

- Oldest record of written Old Turkic (8th century)
  - Göktürk (Köktürk) tribe of central Asia
  - Chinese and Old Turkic Inscriptions
  - Parallel Text of verses, telling the story of the tribes rebellion against Chinese emperor (Tang Dynasty)
- Spread over central asia, eastern europe and the middle east
- “Ottoman Turkish” became the official language of the Ottoman Empire (11th century)
History

- “Ottoman Turkish” mix of Kaba, Persian, and Arabic
- Kaba associated with lower social status of the period
- 19th century saw the collapse of the Ottoman empire and the creation of the Republic of Turkey
- Kaba became the bases of modern Turkish.
- Script was also romanized.
Phonology: Vowel Harmony

- Vowel Harmony (Constraints on closely located vowels)
- Vowels: ⟨a⟩, ⟨e⟩, ⟨ı⟩, ⟨i⟩, ⟨o⟩, ⟨ö⟩, ⟨u⟩, ⟨ü⟩
  - twofold (-e/-a) the locative suffix, for example, is -de after front vowels and -da after back vowels.
  - fourfold (-i/-ı/-ü/-u): the genitive suffix
- -XeY can take the form -XeY or XaY
- -XiY can take the form -XiY, -XiY, -XüY, or -XuY
## Grammar: Nouns

<table>
<thead>
<tr>
<th>Case</th>
<th>Endings</th>
<th>Example</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Village</td>
<td></td>
</tr>
<tr>
<td>Nominative</td>
<td>köy</td>
<td>köy</td>
<td>(the) village/tree</td>
</tr>
<tr>
<td>Genitive</td>
<td>-ın (-in -ün -un)</td>
<td>köyün</td>
<td>the village's/tree's</td>
</tr>
<tr>
<td>Dative</td>
<td>-e (-a)</td>
<td>köye</td>
<td>to the village/tree</td>
</tr>
<tr>
<td>Accusative</td>
<td>-i (-i -ü -u)</td>
<td>köyü</td>
<td>the village/tree</td>
</tr>
<tr>
<td>Ablative</td>
<td>-den (-dan)</td>
<td>köyden</td>
<td>from the village/tree</td>
</tr>
<tr>
<td>Locative</td>
<td>-de (-da)</td>
<td>köyde</td>
<td>in the village/on the tree</td>
</tr>
</tbody>
</table>

Source: [http://en.wikipedia.org/wiki/Turkish_language](http://en.wikipedia.org/wiki/Turkish_language)
## Grammar: Nouns

<table>
<thead>
<tr>
<th>Turkish</th>
<th>+Morpheme</th>
<th>Surface</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>ev</td>
<td></td>
<td>ev</td>
<td>(the) house</td>
</tr>
<tr>
<td>ev</td>
<td>+ler (+lar)</td>
<td>evler</td>
<td>(the) houses</td>
</tr>
<tr>
<td>ev</td>
<td>+in</td>
<td>evin</td>
<td>your (sing.) house</td>
</tr>
<tr>
<td>ev</td>
<td>+iniz</td>
<td>eviniz</td>
<td>your (pl./formal) house</td>
</tr>
<tr>
<td>ev</td>
<td>+im</td>
<td>evim</td>
<td>my house</td>
</tr>
<tr>
<td>ev</td>
<td>+im +de</td>
<td>evimde</td>
<td>at my house</td>
</tr>
<tr>
<td>ev</td>
<td>+im +de +im</td>
<td>evimdeyim</td>
<td>I am at my house.</td>
</tr>
</tbody>
</table>

source: [http://en.wikipedia.org/wiki/Turkish_language](http://en.wikipedia.org/wiki/Turkish_language)
Grammar: Adjectives

- Adjectives are not declined*
- Unless they are used as nouns (in which case they are declined)
- Can be placed before or after a noun
  - mavi ev - the blue house
  - ev, mavi - the house is blue (statement of fact)

*Declension: the inflection of nouns, pronouns, adjectives, and articles to indicate number, case, and gender.
Grammar: Verbs

- Can be inflected to indicate:
  - tense
  - mood
  - aspect
  - negation
    - -değil (to be)
      - Türküm = i am Turkish,
      - Türk değilim = i am not Turkish.
    - -me (all others)
Grammar: Word Order

- Turkish is mostly SOV
- Some constraints
  - Definite article (the) precedes the indefinite (a)
  - for eg:
    - hikâyeyi bir çocuğa anlattı
    - "she told the story to a child"
    - “she told a story to the child"
**Unique Aspects: Mirativity**

Mirativity: shows how familiar a speaker is with a piece of information.

<table>
<thead>
<tr>
<th>Turkish</th>
<th>Kemal</th>
<th>gel-di</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>Kemal</td>
<td>came</td>
</tr>
<tr>
<td>Translation</td>
<td>Kemal came.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turkish</th>
<th>Kemal</th>
<th>gel-mls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>Kemal</td>
<td>came, MIRATIVE</td>
</tr>
<tr>
<td>Translation</td>
<td>Kemal, surprisingly, came</td>
<td></td>
</tr>
</tbody>
</table>

**Unique Aspects: Causativity**

Morphological causatives: increase valence (change the argument structure of a verb so that it takes additional arguments)

<table>
<thead>
<tr>
<th>Turkish</th>
<th>Hasan</th>
<th>öl</th>
<th>dü</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>Hasan</td>
<td>died</td>
<td>CAUSATIVE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turkish</th>
<th>Ali</th>
<th>Hassan-t</th>
<th>öl-dür-dü</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td>Ali</td>
<td>killed</td>
<td>Hassan</td>
</tr>
</tbody>
</table>

Unique Aspects: Existentials

Existentials refer to the existence of something.

- special verb for existentials
- special verb for negating existentials

<table>
<thead>
<tr>
<th>Turkish</th>
<th>kösede</th>
<th>bir</th>
<th>kahve</th>
<th>var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>on:corner</td>
<td>a</td>
<td>book</td>
<td>EXIST</td>
</tr>
<tr>
<td>Translation</td>
<td>There is a book on the corner.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turkish</th>
<th>kösede</th>
<th>bir</th>
<th>kahve</th>
<th>yok</th>
</tr>
</thead>
<tbody>
<tr>
<td>Morphosyntax</td>
<td>on:corner</td>
<td>a</td>
<td>book</td>
<td>LACK</td>
</tr>
<tr>
<td>Translation</td>
<td>There isn’t a book on the corner.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MT: Challenges

- Language Modeling: naive LM models will produce high OOV rate.
- Alignment is challenging when source and target have large difference in lengths.
- “BLEU will kill you if you get a single morpheme wrong”
  -K Oflazer
Corpora

- The Swedish-Turkish Parallel Corpus and Tools for its Creation (LREC?)
- Turkish English parallel text from Kemal Oflazer (COLING 08)
- Turkish Wordnet
- TS Corpus
- LDC: ECI Multilingual Text
- OPUS: KDEdoc (~226 bitexts)
- OPUS: KDE (~1800 bitexts)
- OPUS: PHP (~230 bitexts)
State of the art

- Eyigoz et al (ACL 2013)
  - 50K sentences from Turkish Ministry of Foreign Affairs Documents
  - English to Turkish : 22.52
  - Turkish to English : 29.98
- Any other published results of open corpora?
Thank you!
Teşekkür ederiz!