

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

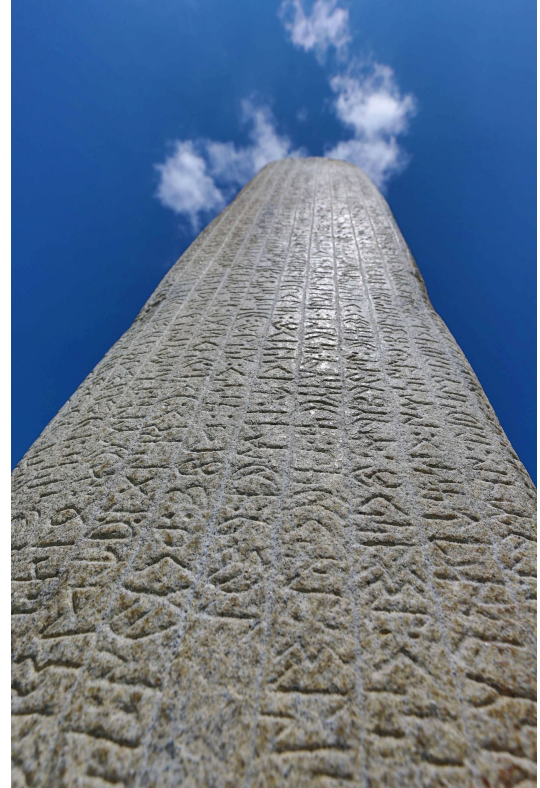
Demographics

- 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 , -XıY , -XüY , or -XuY

Grammar: Nouns

Case	Endings	Example		Meaning
		<i>Village</i>	<i>Tree</i>	
Nominative		<i>köy</i>	<i>ağaç</i>	(the) village/tree
Genitive	-in (<i>-in -ün -un</i>)	<i>köyün</i>	<i>ağacın</i>	the village's/tree's of the village/tree
Dative	-e (<i>-a</i>)	<i>köye</i>	<i>ağaca</i>	to the village/tree
Accusative	-i (<i>-i -ü -u</i>)	<i>köyü</i>	<i>ağacı</i>	the village/tree
Ablative	-den (<i>-dan</i>)	<i>köyden</i>	<i>ağaçtan</i>	from the village/tree
Locative	-de (<i>-da</i>)	<i>köyde</i>	<i>ağaçta</i>	in the village/on the tree

Grammar: Nouns

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

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)

Turkish	English
<i>gel-</i>	(to) come
<i>gelebil-</i>	(to) be able to come
<i>gelme-</i>	not (to) come
<i>geleme-</i>	(to) be unable to come
<i>gelememiş</i>	Apparently (s)he couldn't come
<i>gelebilecek</i>	(s)he'll be able to come
<i>gelmeyebilir</i>	(s)he may (possibly) not come
<i>gelebilirsen</i>	if thou can come
<i>gelinir</i>	(passive) one comes, people come
<i>gelebilmeliydin</i>	thou shouldst have been able to come
<i>gelebilsedydin</i>	if thou could have come
<i>gelmeliydin</i>	thou shouldst have come

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.

Turkish	Kemal	gel-di
Morphosyntax	Kemal	came
Translation	Kemal came.	

Turkish	Kemal	gel-mls
Morphosyntax	Kemal	came, MIRATIVE
Translation	Kemal, surprisingly, came	

Unique Aspects: Causativity

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

Turkish	Hasan	öl	dü
Morphosyntax		die-NON CAUSATIVE	
Translation	Hasan died.		

Turkish	Ali	Hassan-t	öl-dür-dü
Morphosyntax			die - CAUSATIVE
Translation	Ali killed Hassan		

Unique Aspects: Existentials

Existentials refer to the existence of something.

- special verb for existentials
- special verb for negating existentials

Positive Existential

Turkish	kösede	bir	kahve	<i>var</i>
Morphosyntax	on:corner	a	book	EXIST
Translation	There is a book on the corner.			

Negative Existential

Turkish	kösede	bir	kahve	<i>yok</i>
Morphosyntax	on:corner	a	book	LACK
Translation	There isn't a book on the corner.			

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!