

11-737 Multilingual NLP

Lang in 10: Hindi

Example of 10 minute presentation on a language



Carnegie Mellon University

Language Technologies Institute

Hindi हिन्दी

- Hindi spoken in Northern India
 - 320mn native speakers
 - 270 L2 speakers
 - 3rd or 4th most spoken language in world (after Mandarin, English, Spanish)
- Lingua franca for Northern India
 - Taught in schools throughout India
- Hindi != Urdu
 - But mutually intelligible
 - Writing script is different (but writing style is different too)
 - Hindi more Sanskrit words, Urdu more Persian/Arabic words
 - Sometimes both are called “Hindustani”
 - Share phonology, grammar

Hindi

- Indo-European →
 - Indo-Iranian → Indo-Aryan → Western Hindi → Hindustani → Hindi
- Has common cognates with European languages
 - महा रजा (maha raja) → magnus royal (great king)
 - Numbers ek do teen chaar paanch
- Script is Devnagari देवनागरी
 - Brahmi script (as many Indian subcontinent languages are)
 - Also used for nearby languages (e.g. Marathi, Nepali)
 - Urdu uses Perso-Arabic (Nasta'liq)
 - Often in social media contexts Hindi is written in latin script

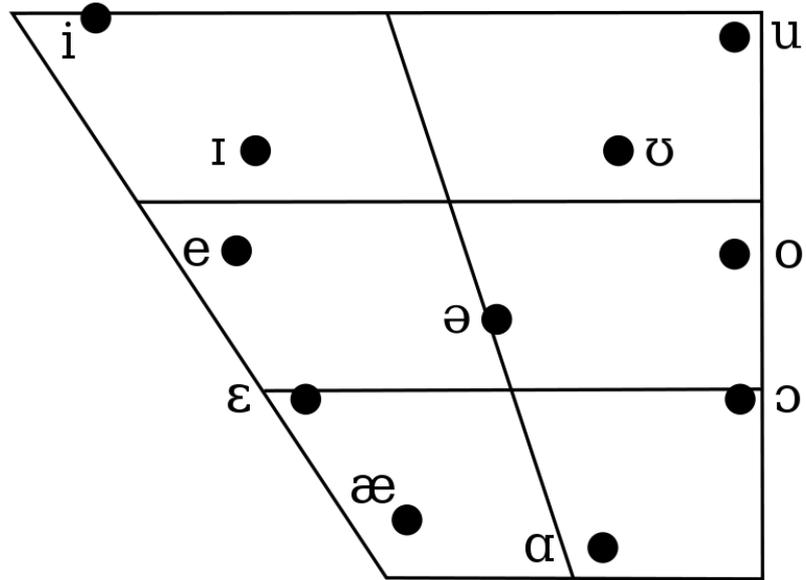
Grammar

- Default order: S O V
- But often pro-dropped, subject
- Gender for all nouns
- Inflection morphology for agreement
- Ergative marking
 - In transitive verb (sometimes) agreement is object/verb not subj/verb



Phonology

- Vowels: schwa, vowel length (plus borrowed vowel *æ* from English)



From Ohala 1999 via Wikipedia

Phonology – consonants

		Labial	Dental/ Alveolar	Retroflex	Palatal	Velar	Uvular	Glottal
Nasal		m	n	ɳ	(ɲ)	ŋ		
Plosive/ Affricate	voiceless	p	t	ʈ	tʃ	k	(q)	
	voiceless aspirated	p ^h	t ^h	ʈ ^h	tʃ ^h	k ^h		
	voiced	b	d	ɖ	dʒ	g		
	voiced aspirated	b ^h	d ^h	ɖ ^h	dʒ ^h	g ^h		
Flap and trill	plain		r	ɽ				
	voiced aspirated			ɽ ^h				
Fricative	voiceless	f	s		ʃ		(x)	ɦ
	voiced		z		(ʒ)		(ɣ)	
Approximant		ʋ	l		j			

From wikipedia: Hindi Phonology



Hindi vs Hinglish

- Most educated Hindi speakers are fluent in English
- Code-switching: mixing of two languages in one utterance
 - Common in multilingual environments
 - Typically more common in casual speech (online text)
 - Borrows phonology, morphology, grammar from **both** languages
- In last 50 years more and more English borrowing in Hindi
 - Hinglish may be the new Hindi.



Lang in 10

- History, geography, social position
- Linguistic: morphology, grammar, phonology
- Examples of something (linguistically) interesting about the language
- Status with respect to resources
- Influences, social use, issues that may affect collection/access



