

Decoding Human Language

CMPUT 331

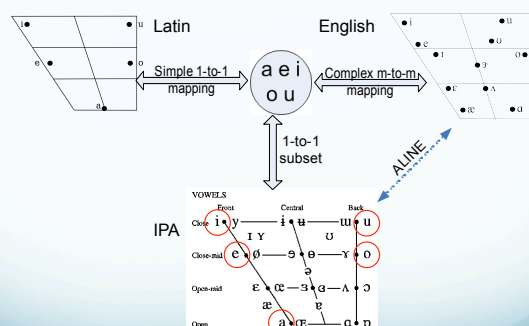
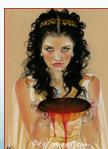
Transliteration

- The task of converting a word from one writing script to another.
- Usually based on phonetics.
- Differs from Romanization and translation:
 - e.g. خالد
 - Romanization: khald
 - Translation: eternal
 - Transliteration: Khalid

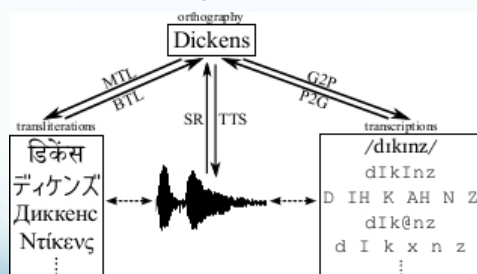
Slide 2

How was Latin pronounced?

- English: Cicero */ˈsɪsɪroʊ/*
 - Greek: Κικέρων
 - Japanese: キケロ
 - Russian: Цицерон
- Greek: Κίρκη
 - English: Circe */ˈsɜːsiː/*
 - Japanese: キルケー
 - Russian: Кирка
- Greek sounds in Latin: y k th ph



Word Representations











Transliteration and Decipherment

- Is transliteration between scripts or languages?
- Со чи = Sochi Sotchi Sotschi Sotsji Szocsi ... ?
- Can we do **unsupervised** transliteration?
- Serbian: признавање урођеног
- [Hauer et al, 2014] recover character mappings from a small sample of Serbian using a Croatian language model and decipherment techniques.
- Can this be done on unrelated or extinct languages?

Deciphering Egyptian Hieroglyphs

- The enclosed text is a name.
- In 1814, a breakthrough:

Thomas Young's decipherment of  the cartouche of **Ptolemaios** (standard version) from the Rosetta Stone.

Hieroglyph	Young's sound value	Actual sound value
	P	P
	t	t
	optional	o
	lo or ole	l
	me or m	m
	l	l or y
	osh or os	s



Linear B



- A syllabic script in an unknown language
- Found on excavated Cretan tablets in 1900
- Mycenaean culture destroyed ca. 1375 BC
- Initially believed to represent a lost language
- 1948: Alice Kober identified inflected forms of words.
- Kober developed a method for relating different script syllable symbols by shared consonants and vowels.

Deciphering Linear B

- Michael Ventris matched ancient town names to Linear B syllabic signs.
 - Knossos (ko-no-so)
 - Amnisos (a-mi-ni-so)
 - Tulisos (tu-li-so)
- Ventris used Kober's method to decipher several words.
- This allowed him to identify the language of the tablets as Ancient Greek.
- Linear A remains undeciphered.



Kober's Method

	Akkadian
Case 1	sa-da-nu
Case 2	sa-da-ni
Case 3	sa-du

	Word A	Word B
Case 1	25 26 37 57	70 52 41 57
Case 2	25 26 37 36	70 52 41 36
Case 3	25 26 05	70 52 12

	Vowel I	Vowel II
Consonant I	37	05
Consonant II	41	12

