11-324/11-624/11-724 Human Language for AI

Phonemes and Underlying Representations

David R. Mortensen

September 21, 2022

Language Technologies Institute
Carnegie Mellon University

Where We Left off

When we left off last time, we were talking about phonemic analysis.

Ordering

We will view allophonic rules as being INTRINSICALLY ORDERED. This means that the order in which the rules apply is governed by how specific the environment in the rule is (from most specific to most general). There is always an "elsewhere" rule that gives the realization of the phoneme if none of the more specific rules applies.

- Korean
  1. /l/ → [ɾ] / V_V (between vowels)
  2. /l/ → [l] / elsewhere
- American English
  1. /t/ → [tʰ] / at the beginning of stressed syllables
  2. /t/ → [t̚] / _# (word-finally)
  3. /t/ → [t] / elsewhere

Algorithm for Phonemic Analysis

1: **function** PHONEMICIZE(C: string of phones, n: integer)
2: $G \leftarrow \text{ngrams}(C, k)$
3: $P \leftarrow \emptyset$ ▷ Map from string to tuple
4: **for each** $g \in G$ **do**
5: $l \leftarrow g[: (k-1)/2]$ ▷ index starts at zero
6: $r \leftarrow g[-(k-1)/2 :]$
7: $p \leftarrow g[(k-1)/2]$
8: $P[p] \leftarrow P[p] \cup \{\langle l, r \rangle\}$

Different Forms of the Same Morpheme

Singular	Phonemic	Plural	Phonemic
dog	/dɑg/	dogs	/dɑg-z/
cat	/kæt/	cats	/kæt-s/
horse	/hɔɹs/	horses	/hɔɹs-əz/

Different Forms of the Same Morpheme

Infinitive	Phonemic	3SG	Phonemic
take	/tejk/	takes	/tejk-s/
give	/gɪv/	gives	/gɪv-z/
watch	/wɑt͡ʃ/	watches	/wɑt͡ʃ-əz/

Active	Passive	Gerund	Gloss
hopu	hopukia	hopukaŋa	‘catch’
aru	arumia	arumaŋa	‘follow’
tohu	tohuŋia	tohuŋaŋa	‘point out’
maatu	maaturia	maaturaŋa	‘know’

ACTIVE — I *catch* the fish.

PASSIVE — The fish *is caught*.

GERUND — Her *catching* the fish was good.

Active	Passive	Gerund	Gloss
hopu	hopuk-ia	hopuk-aŋa	‘catch’
aru	arum-ia	arum-aŋa	‘follow’
tohu	tohuŋ-ia	tohuŋ-aŋa	‘point out’
maatu	maatur-ia	maatur-aŋa	‘know’

Phonology as a Computational System

Maori II

UR of Root	Active	Passive	Gerund	Gloss
/hopuk/	hopu	hopuk-ia	hopuk-aŋa	‘to catch’
/arum/	aru	arum-ia	arum-aŋa	‘to follow’
/tohuŋ/	tohu	tohuŋ-ia	tohuŋ-aŋa	‘to point out’
/maatur/	maatu	maatur-ia	maatur-aŋa	‘to know’

Phonemic Representations Versus Underlying Representations

Phonemic representations and underlying representations both represent normalizations over the sounds of speech. They are both enclosed in slashes. However, they are not the same.

Phonemic Representations	Underlying Representations
Phonemes	Underlying segments
Normalization over **phones**	Normalization over **allomorphs**
Each **phoneme** always has the same representation	Each **morpheme** always has the same representation

$$\left\{\begin{array}{c} p \\ t \\ k \\ m \\ n \\ ŋ \\ r \\ \dots \end{array}\right\} \rightarrow 0 \ / \ \_ \ \#$$

Voicing Assimilation

$$z \rightarrow s \; / \left\{ \begin{array}{c} p \\ t \\ k \\ f \\ \theta \\ s \\ \int \\ \widehat{t\int} \end{array} \right\} \_ \; \#$$

Schwa Epenthesis

$$0 \rightarrow \text{ə} \; / \; \left\{ \begin{matrix} \text{s} \\ \text{z} \\ \text{ʃ} \\ \text{ʒ} \\ \text{t͡ʃ} \\ \text{d͡ʒ} \end{matrix} \right\} \; \_ \; \left\{ \begin{matrix} \text{s} \\ \text{z} \\ \text{ʃ} \\ \text{ʒ} \\ \text{t͡ʃ} \\ \text{d͡ʒ} \end{matrix} \right\}$$

How Are They Ordered?

- Voicing assimilation

$$z \rightarrow s \; / \; \left\{ \begin{matrix} p \\ t \\ k \\ f \\ \theta \\ s \\ \int \\ \widehat{tʃ} \end{matrix} \right\} \_ \; \#$$

- Epenthesis

$$0 \rightarrow ə \; / \; \left\{ \begin{matrix} s \\ z \\ \int \\ ʒ \\ \widehat{tʃ} \\ \widehat{dʒ} \end{matrix} \right\} \_ \left\{ \begin{matrix} s \\ z \\ \int \\ ʒ \\ \widehat{tʃ} \\ \widehat{dʒ} \end{matrix} \right\}$$

Rule Interactions

feeding Rule A creates an environment where Rule B can apply

bleeding Rule A destroys and environment where Rule A would otherwise apply

counter-feeding Rule B would feed Rule A if their relative orders were reversed

counter-bleeding Rule B would bleed Rule A if their relative orders were reversed

Catalan Example I

MASC SG	FEM SG		MASC SG	FEM SG	
əkelʲ	əkelʲə	‘that’	mal	malə	‘bad’
siβil	siβilə	‘civil’	əskerp	əskerpə	‘shy’
ʃop	ʃopə	‘drenched’	sɛk	sɛkə	‘dry’
əspɛs	əspɛsə	‘thick’	gros	grosə	‘large’
baʃ	baʃə	‘short’	koʃ	koʃə	‘lame’
tot	totə	‘all’	brut	brutə	‘dirty’
pɔk	pɔkə	‘little’	prəsis	prəsizə	‘precise’
frənses	frənsezə	‘French’	gris	grizə	‘grey’
kəzat	kəzaðə	‘married’	bwit	bwiðə	‘empty’
rɔʧ	rɔʒə	‘red’	boʧ	boʒə	‘crazy’
orp	orβə	‘blind’	lʲark	lʲarɣə	‘long’
sek	seɣə	‘blind’	fəʃuk	fəʃuɣə	‘heavy’
grok	groɣə	‘yellow’	puruk	puruɣə	‘fearful’
kandit	kandiðə	‘candid’	frɛt	frɛðə	‘cold’

Catalan Example II

MASC SG	FEM SG		MASC SG	FEM SG	
səɣu	səɣurə	‘sure’	du	durə	‘hard’
səɣəðo	səɣəðorə	‘reaper’	kla	klarə	‘clear’
nu	nuə	‘nude’	kru	kruə	‘raw’
flɔɲʤu	flɔɲʤə	‘soft’	dropu	dropə	‘lazy’
əgzaktə	əgzaktə	‘exact’	əlβi	əlβinə	‘albino’
sa	sanə	‘healthy’	pla	planə	‘level’
bo	bonə	‘good’	sərɛ	sərɛnə	‘calm’
suβlim	suβlimə	‘sublime’	al	altə	‘tall’
fɔr	fɔrtə	‘strong’	kur	kurtə	‘short’
sor	sorðə	‘deaf’	bɛr	bɛrðə	‘green’
san	santə	‘saint’	kəlɛn	kəlɛntə	‘hot’
prufun	prufundə	‘deep’	fəkun	fəkundə	‘fertile’
dəsen	dəsentə	‘decent’	dulen	dulentə	‘bad’
əstuðian	əstuðiantə	‘student’	blaŋ	blaŋkə	‘white’