

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

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.

1

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. $/l/ \rightarrow [r] / V_V$ (between vowels)
 - 2. $/l/ \rightarrow [l]$ / elsewhere
- American English
 - 1. $/t/ \rightarrow [t^h]$ / at the beginning of stressed syllables
 - 2. $/t/ \rightarrow [\dot{t}] / _#$ (word-finally)
 - 3. $/t/ \rightarrow [t] / elsewhere$

Algorithm for Phonemic Analysis

```
1: function Phonemicize(C: string of phones, n: integer)
```

- 2: $G \leftarrow \operatorname{ngrams}(C, k)$
- 3: $P \leftarrow \emptyset$
- 4: for each $g \in G$ do
- 5: $l \leftarrow g[:(k-1)/2]$
- $6: r \leftarrow g[-(k-1)/2:]$
- 7: $p \leftarrow g[(k-1)/2]$
- 8: $P[p] \leftarrow P[p] \cup \{\langle l, r \rangle\}$

▶ Map from string to tuple

⊳ index starts at zero

Different Forms of the Same Morpheme

Singular	Phonemic	Plural	Phonemic
dog	/dag/	dogs	/dag-z/
cat	/kæt/	cats	/kæt-s/
horse	/bs.ch/	horses	/se-such/

Different Forms of the Same Morpheme

Infinitive	Phonemic	3sg	Phonemic
take	/tejk/	takes	/tejk-s/
give	/giv/	gives	/giv-z/
watch	/watʃ/	watches	/watJ-əz/

Maori I

hopu hopukia hopukaŋa 'catch' aru arumia arumaŋa 'follow' tohu tohuŋia tohuŋaŋa 'point out' maatu maaturia maaturaŋa 'know'	Active	Passive	Gerund	Gloss
tohu tohunia tohunana 'point out'	hopu	hopukia	hopukaŋa	'catch'
3 3 1	aru	arumia	arumaŋa	'follow'
maatu maaturia maaturaŋa 'know'	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.

6

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'

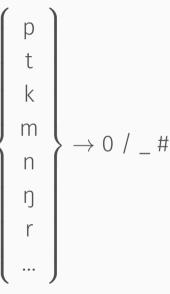
Phonemic Representations Versus Underlying Representations

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.

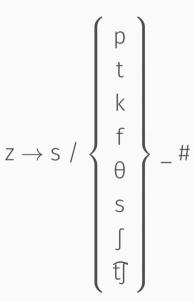
Phonemic Representations	Underlying Representations			
Phonemes Normalization over phones Each phoneme always has the same representation	Underlying segments Normalization over allomorphs Each morpheme always has the same representation			

9

Maori Apocope



Voicing Assimilation



Schwa Epenthesis

$$0 \to \partial / \begin{cases} s \\ z \\ f \\ 3 \\ ff \\ d3 \end{cases} - \begin{cases} s \\ z \\ f \\ 3 \\ ff \\ d3 \end{cases}$$

How Are They Ordered?

· Voicing assimilation

$$z \to s / \left\{ \begin{array}{c} p \\ t \\ k \\ f \\ \theta \\ s \\ f \\ f \end{array} \right\} - \#$$

· Epenthesis

$$0 \to \partial / \left\{ \begin{array}{c} s \\ z \\ \int \\ 3 \\ \text{ff} \\ \text{d3} \end{array} \right\} - \left\{ \begin{array}{c} s \\ z \\ \int \\ 3 \\ \text{ff} \\ \text{d3} \end{array} \right\}$$

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

FEM SG		MASC SG	FEM SG	
əkel ^j ə	'that'	mal	malə	'bad'
siβilə	'civil'	əskerp	əskerpə	'shy'
∫opə	'drenched'	sεk	sεkə	'dry'
əspɛsə	'thick'	gros	grosə	'large'
ba∫ə	'short'	ko∫	ko∫ə	'lame'
totə	ʻall'	brut	brutə	'dirty'
pokə	'little'	prəsis	prəsizə	'precise'
frənsezə	'French'	gris	grizə	'grey'
kəzaðə	'married'	bwit	bwiðə	'empty'
rɔʒə	'red'	botJ	boʒə	'crazy'
orβə	'blind'	l ^j ark	l ^j arɣə	'long'
seyə	'blind'	fə∫uk	fə∫uɣə	'heavy'
groyə	'yellow'	puruk	puruyə	'fearful'
kandiðə	'candid'	fret	frɛðə	'cold'
	akelia siβila Jopa aspesa ba∫a tota poka franseza kazaða roga orβa seya groya	əkeliə 'that' sißilə 'civil' ∫opə 'drenched' əspɛsə 'thick' ba∫ə 'short' totə 'all' pɔkə 'little' frənsezə 'French' kəzaðə 'married' rɔʒə 'red' orβə 'blind' groyə 'yellow'	əkeliə'that'malsiβilə'civil'əskerp∫opə'drenched'sɛkəspɛsə'thick'grosba∫ə'short'ko∫totə'all'brutpɔkə'little'prəsisfrənsezə'French'griskəzaðə'married'bwitrɔʒə'red'botʃorβə'blind'l'arkseɣə'blind'fəʃukgroɣə'yellow'puruk	əkel'ə'that'malmaləsiβilə'civil'əskerpəskerpə∫opə'drenched'sɛksɛkəəspɛsə'thick'grosgrosəbaʃə'short'koʃkoʃətotə'all'brutbrutəpɔkə'little'prəsisprəsizəfrənsezə'French'grisgrizəkəzaðə'married'bwitbwiðərɔʒə'red'botʃ]boʒəorβə'blind'l'arkl'aryəseyə'blind'fəʃukfəʃuyəgroyə'yellow'purukpuruyə

Catalan Example II

MASC SG	FEM SG		MASC SG	FEM SG	
səyu	səyurə	'sure'	du	durə	'hard'
səyəðo	səyəðorə	'reaper'	kla	klarə	'clear'
nu	nuə	'nude'	kru	kruə	'raw'
flɔɲd͡ʒu	flɔɲd͡ʒə	'soft'	dropu	dropə	'lazy'
əgzaktə	əgzaktə	'exact'	əlβi	əlβinə	'albino'
sa	sanə	'healthy'	pla	planə	'level'
bo	bonə	'good'	səre	sərenə	'calm'
suβlim	suβlimə	'sublime'	al	altə	'tall'
for	fortə	'strong'	kur	kurtə	'short'
sor	sorðə	'deaf'	ber	bɛrðə	'green'
san	santə	'saint'	kəlεn	kəlεntə	'hot'
prufun	prufundə	'deep'	fəkun	fəkundə	'fertile'
dəsen	dəsentə	'decent'	dulen	dulentə	'bad'
əstuðian	əstuðiantə	'student'	blaŋ	blaŋkə	'white'