

Big Homework 2

Human Language for AI

Name: Aadit Deshpande (AndrewID: aaditd)

1. Underlying Representations (URs) of Roots, Suffixes

Root Words (Somali Nouns)

S. No.	UR of Root Word	Gloss
1	daar	house
2	gees	side
3	laf	bone
4	lug	leg
5	naag	woman
6	tib	pestle
7	sab	outcast
8	bad	sea
9	ḍʒid	person
10	feed	rib
11	ʃiir	buttermilk
12	ʔul	stick
13	bil	month
14	meel	place
15	kaliil	summer

16	najl	female lamb
17	sum	poison
18	laam	branch
19	sim	hip
20	dan	affair
21	<u>daan</u>	river bank
22	<u>saan</u>	hide
23	nirg	baby female camel
24	gabq	girl
25	hogl	downpour
26	bagl	mule
27	wahar	female kid
28	irbad	needle
29	kefed	pan
30	ḍĩlin	female dwarf
31	bohol	hole
32	jird	trunk
33	?aajad	miracle
34	gaʕam	hand
35	?inan	daughter

Root Words (Somali Verbs)

S. No.	UR of Root Word	Gloss
1	sug	wait

2	kab	fix
3	sid	carry
4	dil	kill
5	gan	aim
6	tum	hammer
7	arg	see
8	gudb	cross a river
9	qosl	laugh
10	hadl	talk

Suffixes

Glossing Abbreviation	Suffix Underlying Form
SG	0 (No suffix)
SG.DEF	-ta
PL	-o
3SG.MASC	-aj
3SG.FEM	-taj
1PL.PAST	-naj

Rules

I discovered 6 rules while determining the URs of the Somali verbs and nouns. I named and ordered them as follows:-

1. Vowel Insertion

Any cluster of two consecutive consonants (with the exception of 'j' as the first consonant) in the root word have the same vowel inserted between them as the one that precedes this cluster.

(This applies only if the two consonants are word-final or before certain suffixes)

$0 \rightarrow V' / V' C_ C [\#, taj\#, naj\#, ta\#]$

2. M-N Form Change

'm' changes to 'n' word-finally and before certain suffixes.

$m \rightarrow n / _ [\#, ta\#, taj\#, naj\#]$

3. Spirantization

Voiced stops become voiced fricatives when they occur between vowels.

$[g, b, d, d] \rightarrow [\gamma, \beta, \delta, z] / V_ V$

4. DT Deletion

't' is deleted if it occurs after the voiced stops 'd' or 'q'

$ta \rightarrow a / [d, d] _$

5. LNAJ Conversion

When 'l' is followed by the sequence 'naj', the 'n' assimilates and becomes an 'l'

$lnaj \rightarrow llaj / _ \#$

6. Sha-ification

The sequence 'lt' becomes 'ʃ' when it occurs before suffixes starting with 'a'

$lt \rightarrow ʃ / _ [a\#, aj\#]$

Note:

- 0 is used for empty string here (In the Vowel Insertion Rule)
- C stands for Consonant
- V stands for Vowel

My Logic for creating and Ordering the rules

- **Rule Creation**

- Vowel Insertion - I initially hypothesized Vowel Deletion while Looking at examples like '*jirid, jirida, jirdo*', (thinking the underlying root is '*jirid*', where the 'i' gets deleted). However, looking at the syllable structure of Somali (Usually CVC), it made more sense for the underlying form to be '*jird*' as consonant clusters were very rarely seen otherwise in the given surface forms (consonants and vowels usually alternate).
- M-N Form Change - From the surface forms '*dan, danta, dano*', and '*sin, sinta, simo*', I concluded that the root word for 'sin' should be 'sim', where the 'm' changes to an 'n' before the suffixes (*ta, taj, naj*) or word-finally.
- Spirantization - From examples like '*tiβo, tib, tibta*' and '*naag, naagta, naayo*', clearly voiced stops were becoming voiced fricatives, when in between vowels.
- DT Deletion - This was apparent, due to the suffix for SG.DEF being 'ta' in almost every case, except when preceded by 'd' or 'q'
- LNAJ Conversion - This was again an exception, seen only in the case of the 1PL.PAST Somali verbs, where the 'lnaj' becomes 'llaj'
- Sha-ification - From examples like '*qoslaj, qosofaj, qosollaj*', and '*dilaj, difaj, dillaj*', the sequence 'lt' would become 'f' each time it appeared in the UR.

- **Rule Ordering**

- From the Somali Noun list, it was clear that Vowel Insertion should be applied **before** Spirantization
 - This is evidenced by the coexistence of the surface forms 'gaβaq' and 'gabqo'
- It was also clear the DT Deletion must apply **after** Spirantization
 - This is evidenced by the two surface forms derived as follows
 - sidaj -> siđaj
 - sidtaj -> sidaj
- The ordering for the final 3 rules (Sha-ification, LNAJ Conversion, and DT Deletion) doesn't matter much, since they're never applied successively at the same location in a word.
- Additionally, I ordered the M-N Form Change after the Vowel Insertion and before Spirantization, in case it affects another rule like LNAJ Conversion.
- Hence, the Rule ordering is
 - 1. Vowel Insertion
 - 2. M-N Form Change
 - 3. Spirantization

- 4. DT Deletion
- 5. LNAJ Conversion
- 6. Sha-ification

2. Underlying Representations for Each Word

Somali Nouns

Both columns are presented in the order “SG, SG.DEF, PL”

S. No.	UR of Root Word	Surface Forms
1	daar, daarta, daaro	daar, daarta, daaro
2	gees, geesta, geeso	gees, geesta, geeso
3	laf, lafta, lafo	laf, lafta, lafo
4	lug, lugta, lugo	lug, lugta, lugo
5	naag, naagta, naago	naag, naagta, naago
6	tib, tibta, tibo	tib, tibta, tibo
7	sab, sabta, sabo	sab, sabta, sabo
8	bad, badta, bado	bad, bada, bado
9	ḍid, ḍidta, ḍido	ḍid, ḍidta, ḍido
10	feed, feedta, feedo	feed, feedta, feedo
11	siir, siirta, siiro	siir, siirta, siiro
12	ʔul, ʔulta, ʔulo	ʔul, ʔulta, ʔulo
13	bil, bilta, bilo	bil, bilta, bilo
14	meel, meelta, meelo	meel, meelta, meelo
15	kaliil, kaliilta, kaliilo	kaliil, kaliilta, kaliilo
16	najl, najlta, najlo	najl, najlta, najlo
17	sum, sumta, sumo	sum, sumta, sumo

18	laam, laamta, laamo	laan, laanta, laamo
19	sim, simta, simo	sin, sinta, simo
20	dan, danta, dano	dan, danta, dano
21	daan, daanta, daano	daan, daanta, daano
22	saan, saanta, saanto	saan, saanta, saanto
23	nirg, nirgta, nirgo	nirig, nirigta, nirgo
24	gabq, gabqta, gabqo	gaḃaq, gaḃaqta, gabqo
25	hogl, hoglta, hoglo	hoɣol, hoɣofa, hoglo
26	bagl, baglta, baglo	baɣal, baɣafa, baglo
27	wahar, waharta, waharo	wahar, waharta, waharo
28	irbad, irbadta, irbado	irbad, irbada, irbaḏo
29	kefed, kefedta, kefedo	kefed, kefeda, kefeḏo
30	ḍĩlin, ḍĩlinta, ḍĩlino	ḍĩlin, ḍĩlinta, ḍĩlino
31	bohol, boholta, boholo	bohol, bohoɣa, boholo
32	jird, jirdta, jirdo	jirid, jirida, jirdo
33	?aajad, ?aajadta, ?aajado	?aajad, ?aajada, ?aajaḏo
34	gaʕm, gaʕmta, gaʕmo	gaʕan, gaʕanta, gaʕmo
35	?inan, ?inanta, ?inano	?inan, ?inanta, ?inano

Somali Verbs

Both columns are presented in the order “3SG.MASC, 3SG.FEM, 1PL.PAST”

S. No.	URs	Surface Forms
1	sugaj, sugtaj, sugnaj	suyaj, sugtaj, sugnaj
2	kabaj, kabtaj, kabnaj	kaḃaj, kabtaj, kabnaj

3	sidaj, sidtaj, sidnaj	siĉaj, sidaj, sidnaj
4	dilaj, diltaj, dilnaj	dilaj, di/aj, dillaj
5	ganaj, gantaj, gannaj	ganaj, gantaj, gannaj
6	tumaj, tumtaj, tumnaj	tumaj, tuntaj, tumnaj
7	argaj, argtaj, argnaj	argaj, aragtaj, aragnaj
8	gudbaj, gudbtaj, gudbnaj	gudbaj, guĉubtaj, guĉubnaj
9	qoslaj, qosltaj, qoslnaj	qoslaj, qoso/aj, qosollaj
10	hadlaj, hadltaj, hadlnaj	hadlaj, hada/aj, hadallaj

3. - 7. FOMA Programming Tasks

The zip file contains

- This pdf
- The foma script with the Phonological Rules for Somali - “somali.xfst”
- The list of 135 underlying forms is named “underlying-forms.txt”
- The output of my script is named “output-surface-forms.txt”
- I have verified that the output matches the correct surface forms exactly (all 135 words)