ters have an isolation form (how they look when there are no other Arabic characters around) and contextual forms (how they look when they connect with other characters to form a word). Ideally, the isolation form and the contextual forms of a letter have the same code point and smart font-rendering technology selects the appropriate form for each context. In practice, Unicode provides a set of "presentation forms" for each letter so that the combining forms can be rendered out of context.

Logical order versus visual order All scripts have a logical order. It corresponds to the order in which words written in the script are pronounced. Sometimes this corresponds to the visual order, whether the script is written left-to-right or right-to-left. But sometimes their are deviations between logical order and visual order. For example, in certain abugidas like the Thai script, diacritics representing vowels pronounced after consonant may be written left of the consonant, to the right of the consonant, above the consonant, or (in two components) on either side of the consonant.

Exercise: FST for Somali Morphophonology

Consider the following sets of data from Somali, a language of the Horn of Africa:

SG	SG.DEF	PL	GLOSS
daar	daarta	daaro	house
gees	geesta	geeso	side
laf	lafta	lafo	bone
lug	lugta	luɣo	leg
naag	naagta	naaɣo	woman
tib	tibta	tiβo	pestle
sab	sabta	saβo	outcast
bad	bada	baðo	sea
d͡ʒid	d͡ʒida	d͡ʒiðo	person
feeɖ	feeɖa	feezo	rib
ʕiir	ʕiirta	ʕiiro	buttermilk
ʔul	ʔuʃa	ʔulo	stick
bil	biʃa	bilo	month
meel	meeʃa	meelo	place
kaliil	kaliiʃa	kaliilo	summer
najl	najʃa	najlo	female lamb
sun	sunta	sumo	poison
laan	laanta	laamo	branch
sin	sinta	simo	hip
dan	danta	dano	affair
daan	daanta	daano	river bank

SG	SG.DEF	PL	GLOSS
saan	saanta	saano	hide
nirig	nirigta	nirgo	baby female camel
gaβaɖ	gaβaɖa	gabɖo	girl
hoɣol	hoɣoʃa	hoglo	downpour
baɣal	baɣaʃa	baglo	mule
waħar	waħarta	waħaro	female kid
irbad	irbada	irbaðo	needle
kefed	kefeda	kefeðo	pan
d͡ʒilin	d͡ʒilinta	d͡ʒilino	female dwarf
bohol	bohoʃa	boholo	hole
jirid	jirida	jirdo	trunk
ʔaajad	ʔaajada	ʔaajaðo	miracle
gaʕan	gaʕanta	gaʕmo	hand
ʔinan	ʔinanta	ʔinano	daughter

3SG.MASC	3SG.FEM	1PL.PAST	GLOSS
suɣaj	sugtaj	sugnaj	wait
kaβaj	kabtaj	kabnaj	fix
siðaj	sidaj	sidnaj	carry
dilaj	diʃaj	dillaj	kill
ganaj	gantaj	gannaj	aim
tumaj	tuntaj	tunnaj	hammer
argaj	aragtaj	aragnaj	see
gudbaj	guðubtaj	guðubnaj	cross a river
qoslaj	qosoʃaj	qosollaj	laugh
hadlaj	haðaʃaj	haðallaj	talk

Use Foma to construct an analysis of these data in terms of context-dependent rewrite rules. You should submit the following items:

1. A README file with the following items:

   (a) A list of the underlying representations you posit for each root

   (b) A list of the underlying representations for each suffix

   (c) Any special notes necessary to understand your implementation

2. A Foma script (`somali.xsft`) defining an FST that transduces between Somali underlying representations and surface representations (like those in the data tables)

3. Two lists of test-cases (1 test case per line):

(a) Underlying representations to be transduced into surface representations.

(b) Surface representations to be transduced into underlying representations.