Carnegie Mellon University
Language
Technologies
Institute

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

Computational Phonology

David R. Mortensen

October 6, 2021

Language Technologies Institute
Carnegie Mellon University

Introduction

Learning Objectives

At the end of this lecture, students will better understand:

- How to perform a morphophonological analysis
- How to evaluate such an analysis

Students will be familiar with the following concepts:

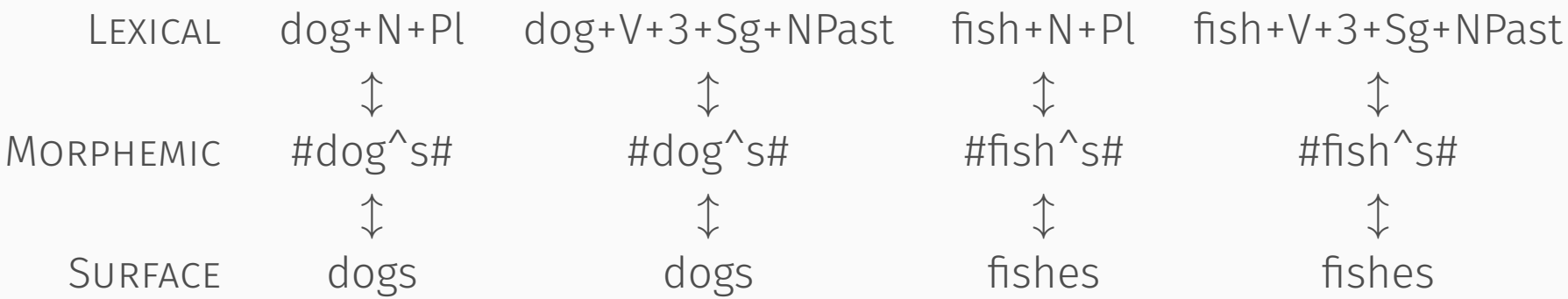
- The notion of morphotactics
- The application of FSTs to morphotactics and morphological analysis
- The Xerox model of morphological analysis
- Unsupervised morphological analysis

Students will be able to do the following things:

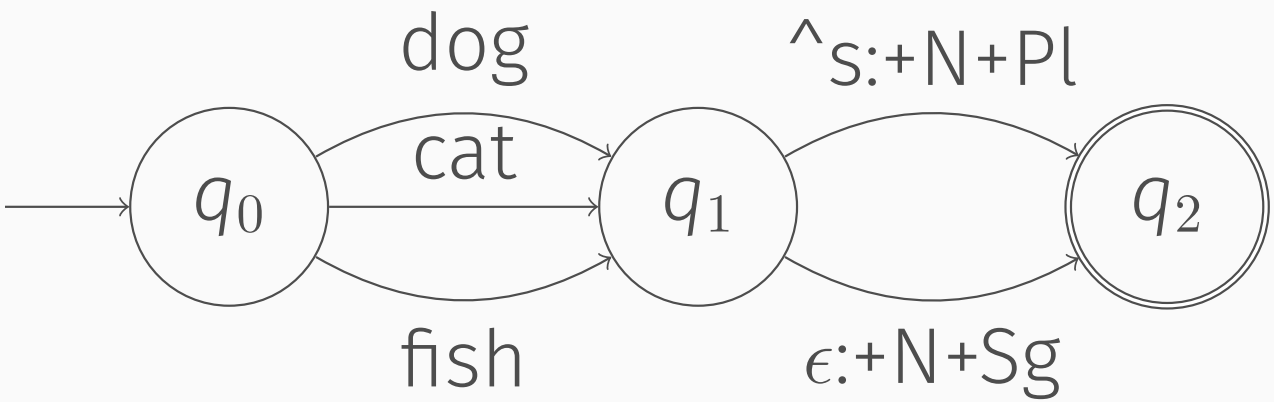
- Represent the morphotactics of a language using the LEXC formalism
- Integrate LEXC morphotactics with phonological rules written using XFST

Finite State Transducers for Morphological Analysis

The Xerox Model of Morphological Analysis



A Simple English Example



An Inuit Example

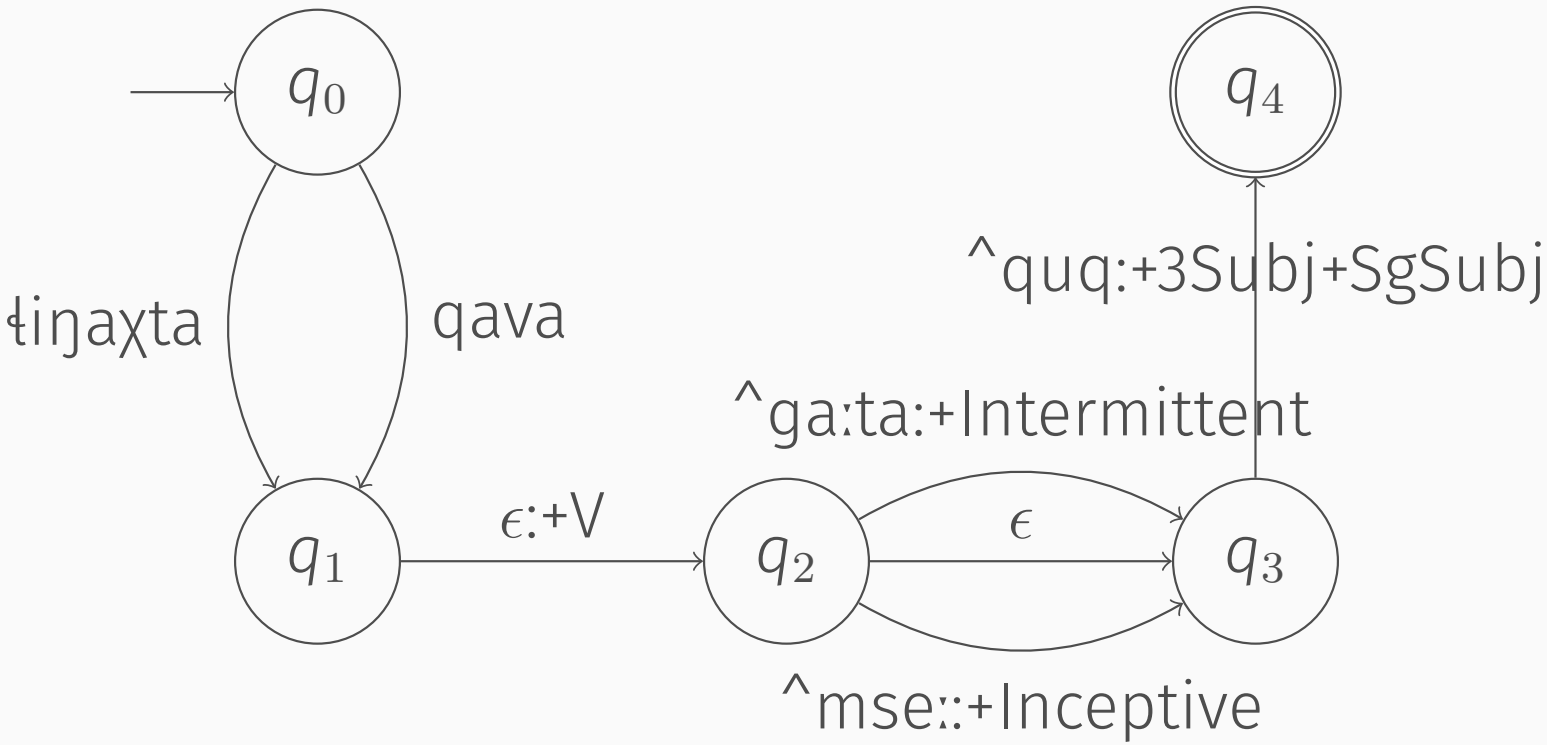
Inuit Data

Inuit	Gloss
aglukata:quq	‘she begins to work’
agluka:quq	‘she works with an intermittent stoppage’
aglunani:ʁaquq	‘she stops working’
aglufaqara:quq	‘she rarely works’
agluvɨʁa:quq	‘she works with difficulty’
aqujgaquq	‘she wanders about’
aqujviluxtaquq	‘she walks back and forth’
iglɨχtɨpixtaquq	‘she walks a lot’
iglɨχtɨkʃa:ga:quq	‘she walks very slowly’
iglɨχtɨkjo:ʁaquq	‘she scarcely drags herself along’
qɨɬpɨχta:quq	‘she makes holes in something’
qɨɬpɨχquʁa:quq	‘she makes holes in various places’
qavaχɬɨqja:quq	‘she sleeps fitfully’
qavaruga:quq	‘she sleeps soundly’
qavamse:quq	‘she dozes’
ku:jma:quq	‘she is swimming habitually towards …’
ku:jmaʁo:ʁaquq	‘she swims habitually’
ɬiŋaχtaquq	‘she rings’
ɬiŋaχtaga:taquq	‘she rings intermittently’

Inuit Data Segmented

Inuit	Gloss
aglu-kata:-quq	‘she begins to work’
aglu-ka:-quq	‘she works with an intermittent stoppage’
aglu-nani:ʁa-quq	‘she stops working’
aglu-faqara:-quq	‘she rarely works’
aglu-ʋiʁa:-quq	‘she works with difficulty’
aquj-ga-quq	‘she wanders about’
aquj-viluxta-quq	‘she walks back and forth’
igɬiχti-pixta-quq	‘she walks a lot’
igɬiχti-kʃa:ga:-quq	‘she walks very slowly’
igɬiχti-kjo:ʁa-quq	‘she scarcely drags herself along’
qɨɬpɨχ-ta:-quq	‘she makes holes in something’
qɨɬpɨχ-quʁa:-quq	‘she makes holes in various places’
qava-χɬɨqja:-quq	‘she sleeps fitfully’
qava-ruga:-quq	‘she sleeps soundly’
qava-mse:-quq	‘she dozes’
ku:jm-a:-quq	‘she is swimming habitually towards …’
ku:jm-aʁo:ʁa-quq	‘she swims habitually’
ɬiŋaχta-quq	‘she rings’
ɬiŋaχta-ga:ta-quq	‘she rings intermittently’

Inuit Morphotactic FST (partial)



Multicharacter Symbols

```
 Multichar_Symbols
+Inceptive +Intermittent1 +Cessive +Rarely +WithDifficulty1
+Aimless +Oscillating
+Frequentive +Slowly +WithDifficulty2
+InSomething +VariousLoc
+Intermittent2 +Soundly +Episodic
+Directive +Habitual
+Intermittent3
+3Subj +SgSubj
```

Transitions from Start State (Root)

```
LEXICON Root
aglu Derivation ;
aquj Derivation ;
iglɨχtɨ Derivation ;
qɨɬpɨχ Derivation ;
qava Derivation ;
ku:jm Derivation ;
ɬiŋaχta Derivation ;
Inflection ;
```

Derivation

```
LEXICON Derivation
+Inceptive:kataː Inflection ;
+Intermittent1:kaː Inflection ;
+Cessive:naniːʁa Inflection ;
+Rarely:faqaraː Inflection ;
+WithDifficulty1:vɨʁaː Inflection ;
...
+Intermittent3:gaːta Inflection ;
```

Inflection

```
LEXICON Inflection
+3Subj+SgSubj:quq # ;
```

Putting it Together

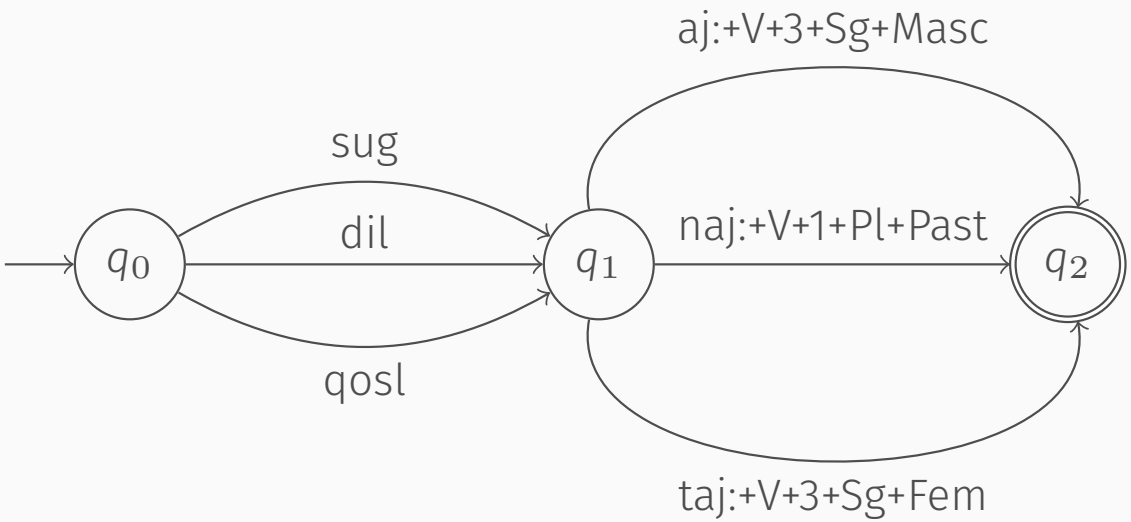
```
read lexc < inuit.lexc
define Verbs;
read regex Verbs .o. SpellingRules;
```

Somali Example

Somali Verbs

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

Somali FST



Other Models of Computational Morphology

Unsupervised Models

- Morfessor, Morfessor Flatcat, etc.
- Detect patterns in the occurrence of letters/segments to find morphemes (or morpheme-like subword units)
- Information-theoretic basis, as in unsupervised word segmentation
- Just provide segmentations, typically, not full analyses

Seq2seq Models of Morphological Reinflection

- Currently very successful for some morphological tasks (e.g. reinflection)
- Treat morphological reinflection and analysis as instances of the general sequence-to-sequence transduction problem
- Supervised—requires training data (though in modest quantities)
- Cross-lingual transfer can help