Linguistics for CS

Lecture 4 – Recap seminar Phonetics and Morphology

Anca Dinu

NLP master programme

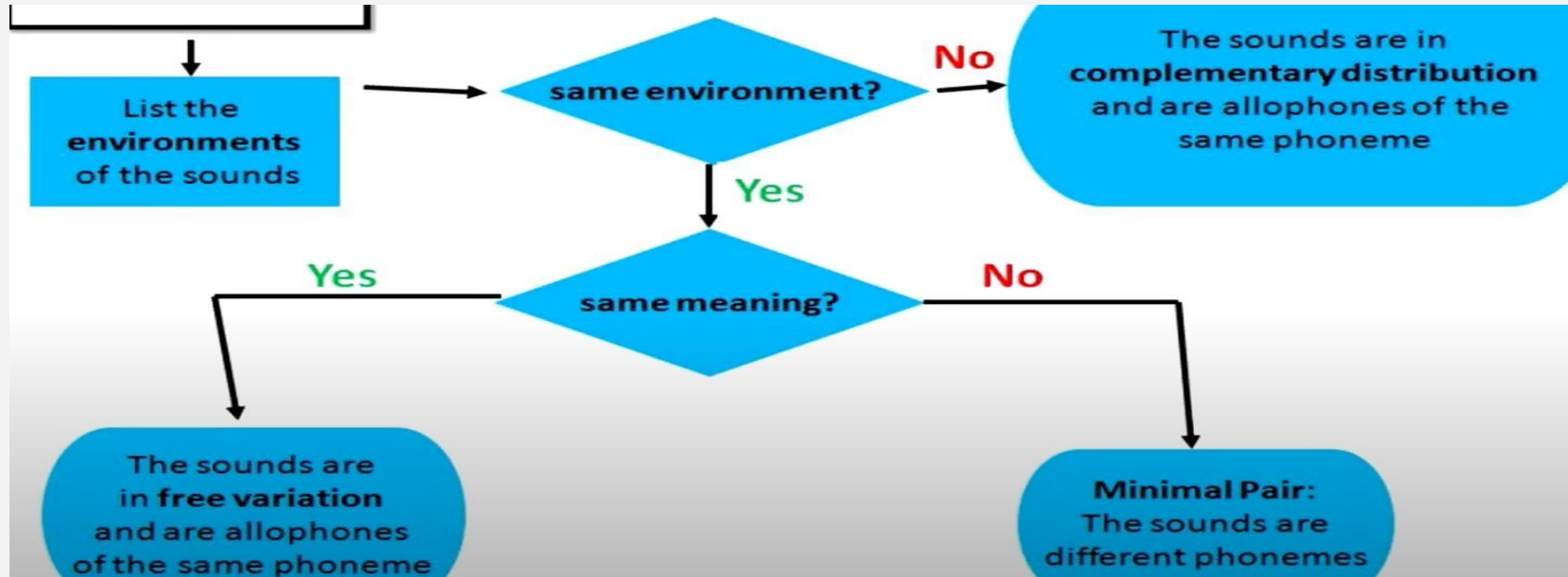**University of Bucharest, Novembre 2025**

01 Phonology: phonemes/ allophones, minimal pairs test

02 Morphology: root, stem, affix, Morphological analysis with FSAs and FSTs

Phonology - finding phonemes/allophones

Recap methodology: (slightly) different sounds from a language do they have:

Phonology: Exercise 1

- **Are the phones [ɹ] and [ɹ̥] allophones (in compelemtary distribution or free variation) or phonemes?**

(1) 'pray'	[pʰɹ̥eɪ]	'fresh'	[fɹ̥ɛʃ]
'gray'	[gɹeɪ]	'regain'	[ɹigeɪn]
'crab'	[kʰɹ̥æb]	'shriek'	[ʃɹ̥ik]
'par'	[pʰɑɹ]	'tar'	[tʰɑɹ]
'broker'	[bɹoʊkɹ̩]		

- **Step 1. Same environment? Look for minimal pairs (exactly the same words except for the phones of interest). If there are not minimal pairs, step 2.**

Phonology: Exercise 1

- **Step 2: list the environments**

[ɪ]   [ɪ̥]↓

[g_eɪ]   [pʰ_eɪ]
[ɑ_#]   [kʰ_æ]
[b_oʊ]   [f_ɛ]
[#_i]   [ʃ_i]

- **Step 3: find the rules (may be ML techniques can help):**

/ɪ/ → [ɪ̥] / after voiceless consonants;
/ɪ/ → [ɪ] / elsewhere.

- **Step 4: conclusion: they are allophones in complementary distribution**

Phonology: Exercise 2

- **Are the phones [s] and [ʃ] in Korean allophones (in compelemtary distribution or free variation) or phonemes?**
- **Data:**

[son]	'hand'	[som]	'cotton'
[ʃilsu]	'mistake'	[sosəl]	'novel'
[isa]	'relocating'	[sal]	'skin'
[ʃinho]	'signal'	[maʃida]	'drink'
[ʃihap]	'match'	[sɛk]	'color'
[ʃipsam]	'13'	[oʃip]	'50'

- **Step 1: Look for minimal pairs**

Phonology: Exercise 2

- **Step 2: list the environments**

- **Step 3: find the rules (may be ML techniques can help):**

$$/s/ \rightarrow [ʃ] \ / \ \underline{\quad} \ [i]$$

- **Step 4: conclusion: they are allophones in complementary distribution**

Phonology: Exercise 3

- **Consider the following phones:: [t], [tʰ], [ř] [ʔ], and [t=]. What is their phonemic status?**
- **Data:**

1) [tʰiym] ' team'
2) [tʰowr] ' tore'
3) [stiym] ' steam'
4) [stowr] ' store'
5) [bařl] ' bottle'
6) [kɪřiy] ' kitty'
7) [bæřr] ' batter'
8) [karʔn] ' carton'
9) [bəʔn] ' button'
10) [eyt=] ' eight'
11) [nayt=] ' night'

Phonology: Exercise 3

- **Step 2: list the environments**

[t]	[tʰ]	[ř]	[ʔ]	[t˺]
3) s___	1) #___i	5) a___l	8) r___n	10) ey___#
4) s___	2) #___o	6) ɪ___iy	9) ə___n	11) ay___#
		7) æ___r		

- **Step 3: find the rules (maybe ML techniques can help):**

- **/t/=**
  [t] s___
  [tʰ] #___
  [ř] V___V
  [ʔ] ___N
  [t˺] ___#

- **Step 4: conclusion: they are allophones in complementary distribution**

Phonology: Exercise 4

- **Consider the following data in Italian.**
- Find the minimal pairs. What phonemes do these minimal pairs demonstrate?
- Are the phones [n] and [ŋ] allophones or phonemes?

tinta 'dye'   nero 'black'
tiŋgo 'I dye'   byaŋko 'white'
tenda 'tent'   dʒente 'people'
teŋgo 'I keep'   aŋke 'also'
dansa 'dance'   fuŋgo 'mushroom'
faŋgo 'mud'   sapone 'soap'

Morphology: recap inflectional vs. derivational morpheme

- An **inflectional morpheme** is a morpheme that
  - varies the word by adding grammatical contrast,
  - preserving the meaning
  - keeping the word class the same.
- verbal inflection: conjugation
- nominal inflection: declension

- Examples:
  - number: scaun - **scaun+e**
  - person: merg - **merg +i**
  - gender lup - lup + **oaică**
  - tense merg - merg +**eam**
  - diminutives: suc - suc + **uleț**
  - others: mood, voice, aspect, case

Morphology: recap inflectional vs. derivational morpheme

- A **derivational morpheme** is a morpheme that:
  - creates a new word from an existing word,
  - altering the meaning and
  - changing the word class (usually).

- Examples:
  - Nominalization: cerceta +**re**, trada + **re**
  - Adjectivization: adora + **bil**, fenomen + **al**

Morphology: recap root, stem, affix, lemma

- **Root** is the central morpheme, which contributes the meaning of the word, common to a set of derived or inflected forms.
- **Stem** is the root of word together with the derivational morphemes, to which one can add inflectional morphemes, like:

  - undone, **do** is the **root**, **un** is a derivational morpheme, **undo** is the stem and **ne** is an inflectional morpheme.
- **Lemma**- citation of a word as it appear in a dictionary

Morphology: Exercises 4.1-4.6 from the book The linguistic Structure of Modern English (Brinton and Brinton)

- https://benjamins.com/sites/z.156/exercise/c4q1

Morphological analysis with FSAs

- **Material from the book** *An Introduction to Natural Language Processing, Computational* Linguistics, and Speech Recognition, by Daniel Jurafsky and James H. Martin
- **Derivational data: adjectives become comparatives, adverbs, and opposites:**
  1. big, bigger, biggest,
  2. cool, cooler, coolest, coolly
  3. happy, happier, happiest, happily, unhappy, unhappier,unhappiest, unhappily
  4. clear, clearer, clearest, clearly, unclear, unclearly
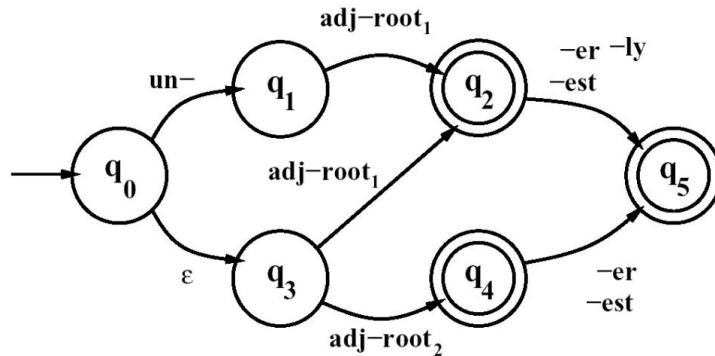
- **Rules: un+**ε/root +er/est/ly

Morphological analysis with FSAs

FSA 1:
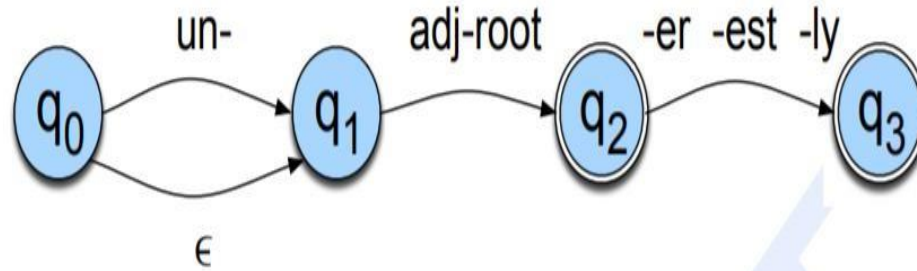
Morphological analysis with FSAs

- The FSA#1 recognizes all the listed adjectives, and ungrammatical forms like *unbig, redly,* and *realest.*
- Thus #1 is revised to become #2.
- The complexity is expected from English derivation.

An FSA for a fragment of English adjective

Morphological analysis with FSTs

Exercise 1:English plural with FST
Data:

reg-noun	irreg-pl-noun	irreg-sg-noun	plural
fox	geese	goose	-s
cat	sheep	sheep	
aardvark	mice	mouse	

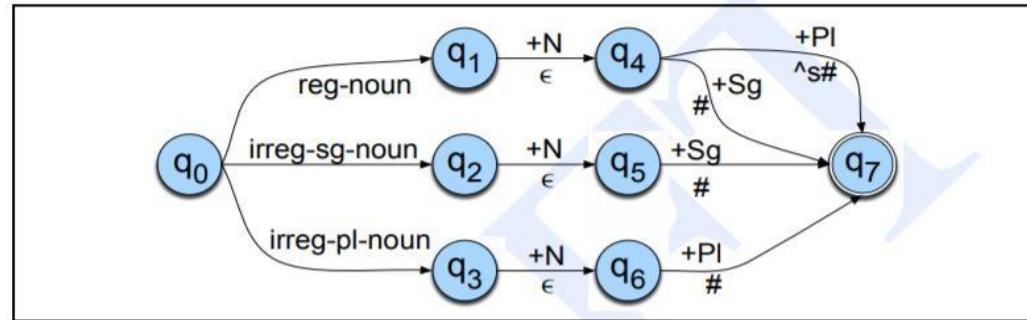| English	
Input	Morphologically Parsed Output
cats	cat +N +PL
cat	cat +N +SG
cities	city +N +Pl
geese	goose +N +Pl
goose	goose +N +Sg
goose	goose +V
gooses	goose +V +1P +Sg
merging	merge +V +PresPart
caught	catch +V +PastPart
caught	catch +V +Past

Morphological analysis with FSTs

-

Figure 3.13 A schematic transducer for English nominal number inflection T_{num}. The symbols above each arc represent elements of the morphological parse in the lexical tape; the symbols below each arc represent the surface tape (or the intermediate tape, to be described later), using the morpheme-boundary symbol ^ and word-boundary marker #. The labels on the arcs leaving q_0 are schematic, and need to be expanded by individual words in the lexicon.

Morphological analysis with FSTs

- **Example of a FST for a larger fragment of English**
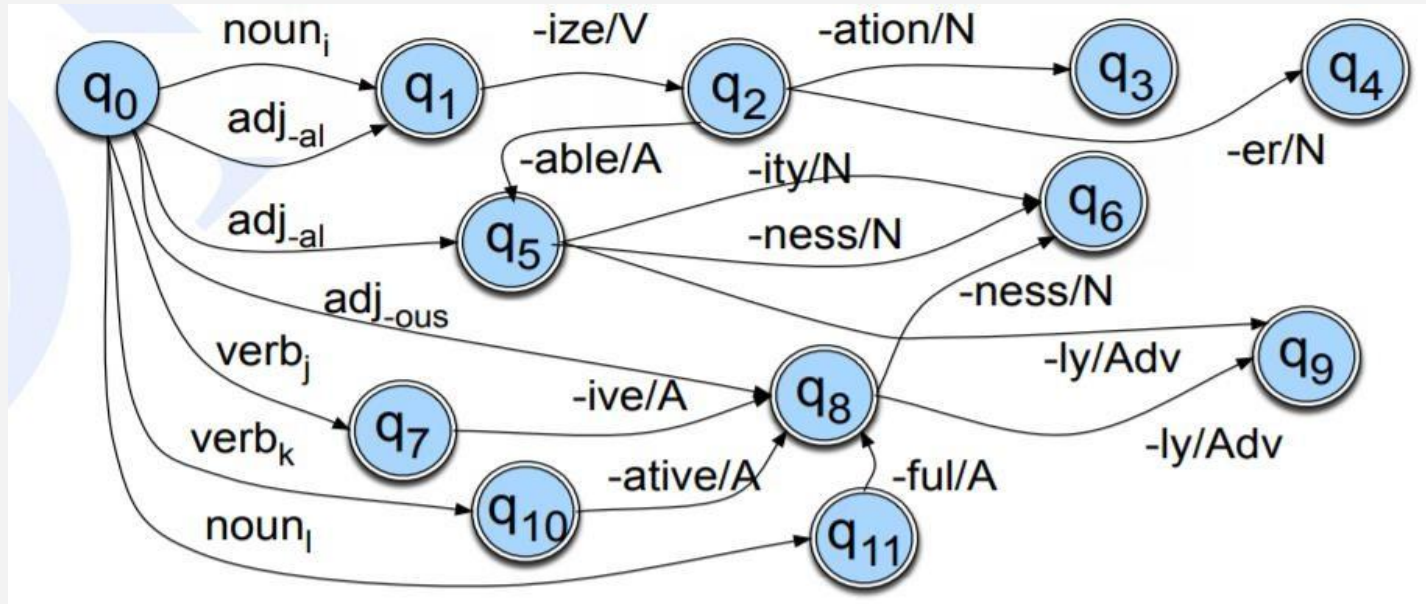
Morphological analysis with FSTs

- **Tutorial for implementation:**
- **https://fomafst.github.io/morphtut.html**

THANKS

You should have some questions now!

anca.dinu@lls.unibuc.ro
ancaddinu@gmail.com
+0785641041