Carnegie Mellon University
Language
Technologies
Institute

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

Phonemes and Underlying Representations

David R. Mortensen

October 3, 2022

Language Technologies Institute
Carnegie Mellon University

Introduction

Learning Objectives

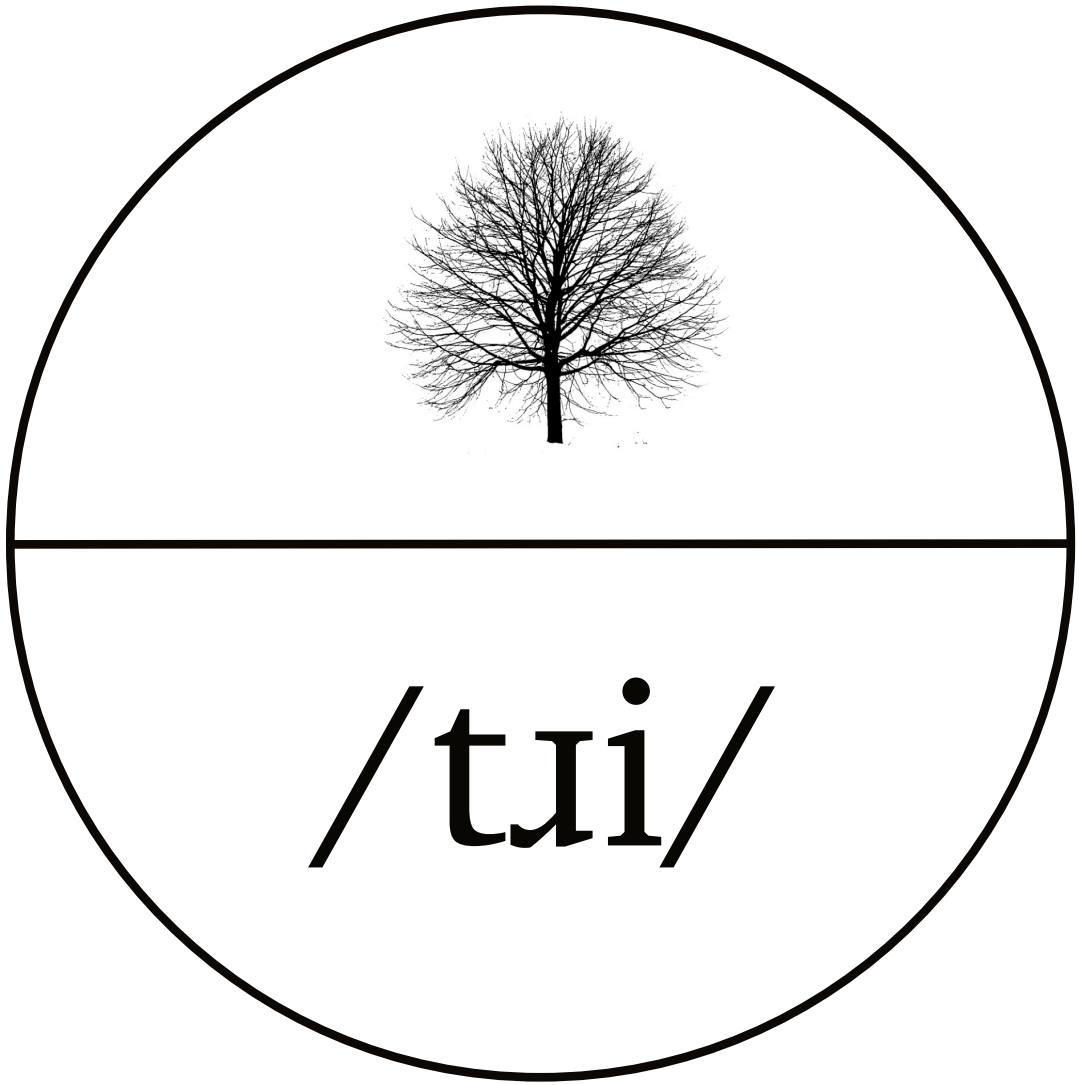
At the end of this lecture, students will understand the following concepts:

- The relationship between sign, construction, and morpheme
- The definition of “word” and why it’s problematic
- The difference between lemmas, lexemes, and listemes
- The difference between roots and affixes
- The difference between morphological form and morphological function
- The difference between derivation and inflection

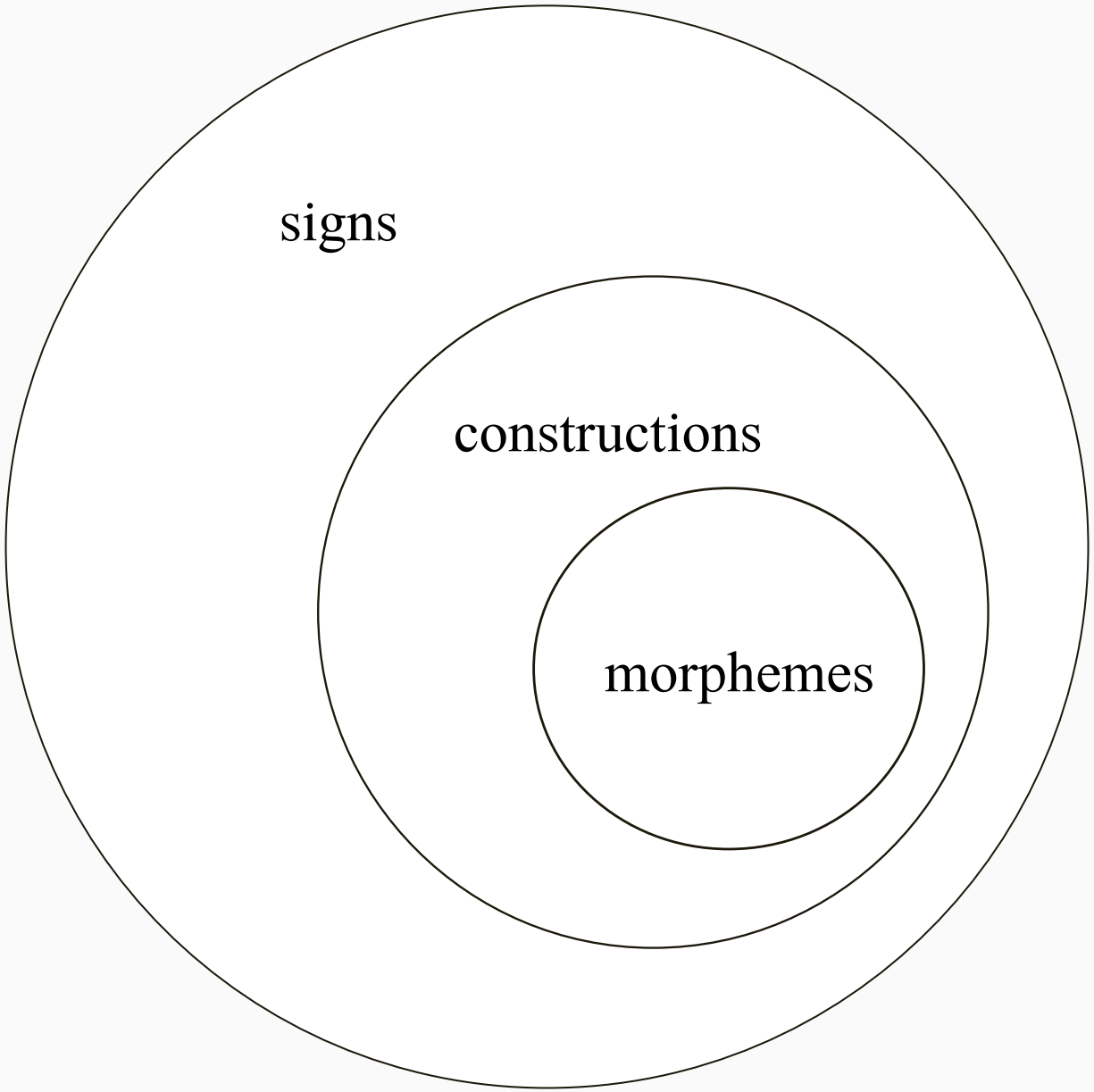
Students will be able to identify the following formal processes in morpghology:

- Prefixation
- Suffixation
- Circumfixation
- Infixation
- Transfixation (root-and-pattern morphology)
- Reduplication
- Internal change/apophony (umlaut, ablaut, etc.)

Signs



Signs, Constructions, and Morphemes



Words, Lexemes, and Listemes

- Listemes—things that must be memorized because they do not follow from any general rule
- Lexemes—"dictionary words". Sets of word forms (like {*sing*, *sang*, *sung*, *singing*}) which are in semantic contrast with other such sets (e.g. {*ring*, *rang*, *rung*, *ringing*} or {*run*, *ran*, *run*, *running*}).
- Words—"free" units of language; difficult to define universally.

If You Want to Start a Brawl in a Bar Full of Linguists...

...shout, “I have discovered a universal definition for ‘word’!”

It Depends on How You Divide Them

Text can be divided into sequences of words (tokenized, in NLP terms) according to various criteria:

- **Orthographically** (e.g., splitting on white space and punctuation)
- **Phonologically** (e.g., splitting text into tokens that satisfy certain conditions on pronunciation)
- **Syntactically** (e.g., splitting text into tokens that serve as units in the grammar)

And others.

Even for the same language, these do not always produce the same results. For example, the *'s* in *student's* is orthographically part of the preceding word and acts, from the standpoint of sound, like other suffixes (e.g., the *-s* in *books*. However, syntactically it is a separate unit.

Even English and Chinese Words Can Have Internal Structure

English and Chinese (among other languages) have lots of words with only one meaningful unit—one morpheme. But other words have many morphemes (meaningful parts):

reoperationalizations → re-operat(e)-(t)ion-al-iz(e)-ation-s

These words have **internal structure**

Chinese Words, too, Can Have Internal Structure

Even in Chinese, which is famous for one-morpheme words, words can have internal structure:

我	‘I’	我们	‘we’
你	‘you’	你们	‘ya’ll’
他	‘he/she’	他们	‘they’
同志	‘comrade’	同志们	‘comrades’

Morphology is the Study of the Internal Structure of Words

Morphology is the study of the internal structure of words

- How morphemes combine
- How morphemes function

Formal Operations

Roots and Suffixes are Kinds of Morphemes

Some morphemes give words their basic meaning. These are called ROOTS. Other morphemes are added to words to make new words, or to make new forms of existing words.

un-think-able; kitten-s

These are called AFFIXES.

Roots Provide Basic Meaning and Affixes Modify It

- The roots (in the first column) express the basic meaning
- Affixes add grammatical meaning (2nd column) or modify the semantic meaning (3rd column)

<root>	<root>ing	<root>er
run	running	runner
think	thinking	thinker
program	programming	programmer
kill	killing	killer

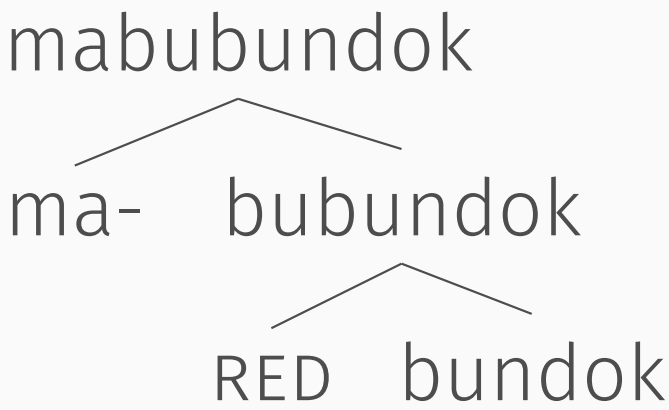
There are other formal processes

- Infixation
- Transfixation
- Reduplication
- Apophony

Tagalog Adjectives Show Multiple Formal Processes

Stem	Singular	Plural	Gloss
laki	malaki	malalaki	‘big’
ganda	maganda	magaganda	‘beautiful’
bundok	mabundok	mabubundok	‘mountainous’

Tagalog Adjectives Can Have a Tree Structure



Tagalog Verbs Also Feature Multiple Formal Processes

Stem	Perfective	Contemplative	Imperfective	Gloss
tapos	tinapos	tatapusin	tinatapos	‘finish’
kain	kumain	kakain	kumakain	‘eat’
sulat	sumulat	susulat	sumusulat	‘write’
hanap	humanap	hahanap	humahanap	‘seek’

German Verbs Show Circumfixation as Well as Suffixation

	Present	Perfect	Preterit
1SG	mache	gemacht	machte
2SG	machst	gemacht	machtest
3SG	macht	gemacht	machte
1PL	machen	gemacht	machten
2PL	macht	gemacht	machtet
3PL	machen	gemacht	machten

Arabic Morphology Uses Transfixes

	Perfect		Imperfect		Participle	
	Active	Passive	Active	Passive	Active	Passive
I	katab	kutib	ktub	ktab	kaatib	ktuub
II	kattab	kuttib	kattib	kattab	kattib	kattab
III	kaatab	kuutib	kaatib	kaatab	kaatib	kaatab
IV	ʔaktab	ʔuktib	ktib	ktab	ktib	ktab
V	takattab	tukuttib	takattab	takattab	takattib	takattab
VI	takaatab	tukuutib	takaatab	takaatab	takaatib	takaatab
VII	nkatab	nkutib	nkatib	nkatab	nkatib	nkatab
VIII	ktatab	ktutib	ktatib	ktatab	ktatib	ktatab
IX	ktab(a)b	ktab(i)b	ktab(i)b			
X	staktab	stuktib	staktib	staktab	staktib	staktab

Mandarin Personal Pronouns Feature Affixation

我	*wo*	1SG	我们	*women*	1PL
你	*ni*	2SG	你们	*nimen*	2PL
他	*ta*	3SG	他们	*tamen*	3PL

Compounding Is Much More Important than Affixation in Mandarin

客厅	‘living room’	沙发	‘sofa’	‘living room sofa’
眼	‘eye’	药	‘medicine’	‘eye medicine’
马	‘horse’	房	‘house’	‘manger’
雨	‘rain’	帽	‘hat’	‘rain hat’

Internal Change/Apophony in English

- Ablaut affects verbs
  - *sing* : *sang* : *sung*
  - *begin* : *began* : *begun*
  - *bleed* : *bled* : *bled*
- Umlaut affects nouns
  - *foot* $\rightarrow$ *feet*
  - *tooth* $\rightarrow$ *teeth*
  - *goose* $\rightarrow$ *geese*

Morphological Functions

Derivational morphology creates new lexemes.

It always changes meaning, part of speech, or both.

Examples of Derivation

(1) a. Pittsburgh-er

b. re-surface

c. black-ish

d. dov-ish

e. acquitt-al

f. efficient-ly

Inflectional morphology does not create new lexemes.

It adds information based on the syntactic context in which a word occurs.

Some Kinds of Inflection

- Case
- Number
- Clusivity
- Grammatical gender
- TAM
  - Tense
  - Aspect
  - Modality
  - Evidentiality
- Formality
- Voice (?)

Examples of Inflection

(2) dog-s

(3) a. walk-s

b. walk-ed

c. walk-ing

Questions?