



VIETNAM NATIONAL UNIVERSITY – HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY

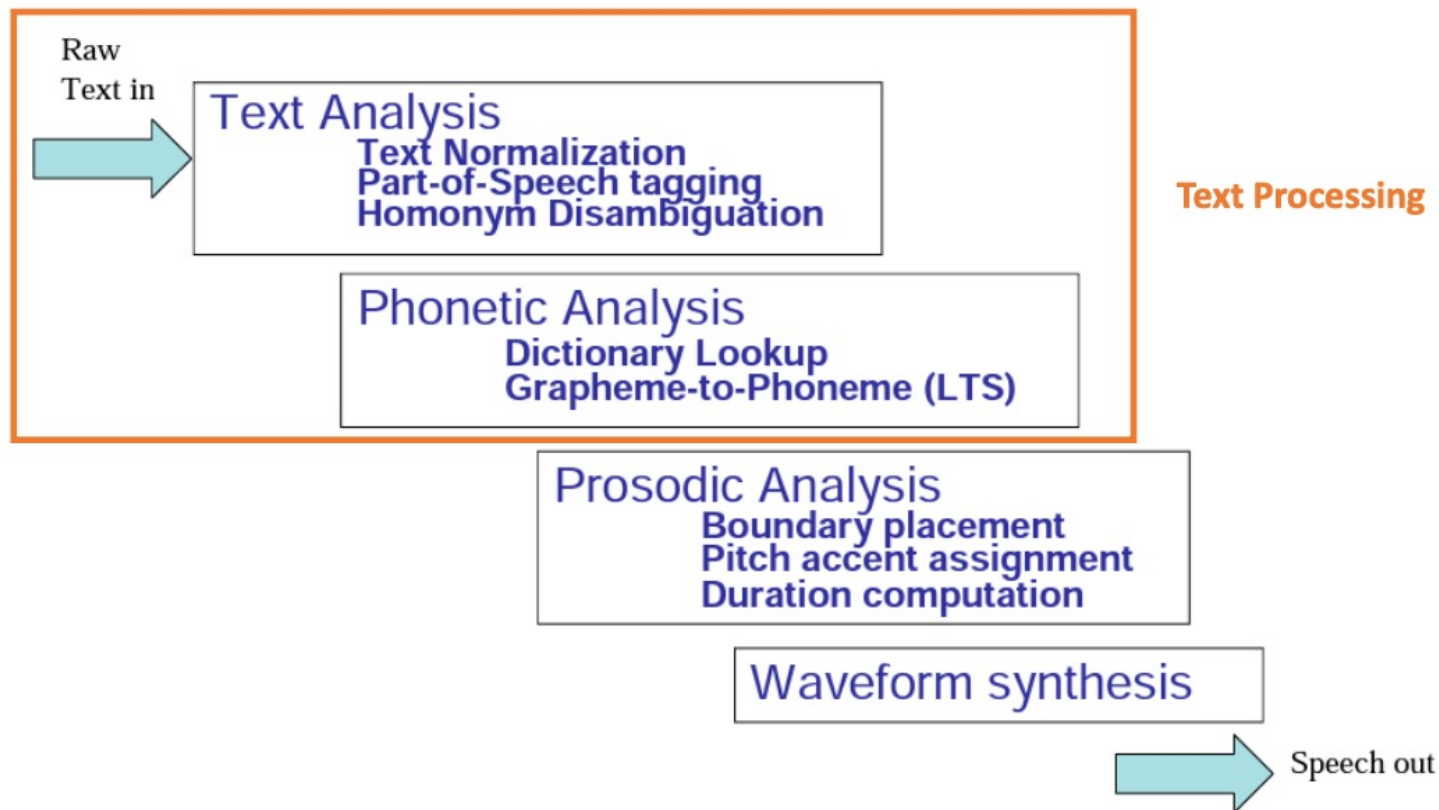
Chapter 4

SPEECH SYNTHESIS (cont.)

Lecturer: MSc. Luan Thanh Nguyen

Faculty of Information Science and Engineering

Text-to-Speech (TTS)



Text Analysis: Tokenization

❖ Tokenizing Words

- Splitting a text into separate words, or tokens, while preserving the meaning of the text.
- For English, whitespace is usually a good starting point.

❖ Examples:

- ❑ I love my school so so so much!
 - Tokens: ['I', 'love', 'my', 'school', 'so', 'so', 'so', 'much!']
- ❑ I get up at 5 AM every morning.
 - Tokens: ['I', 'get', 'up', 'at', '5', 'AM', 'every', 'morning.']

Text Analysis: Tokenization

❖ Subword Tokenization

- Tokens are smaller than words. Subwords can be arbitrary substrings.

"This is tokenizing."

Character Level

[T] [h] [i] [s] [i] [s] [t] [o] [k] [e] [n] [i] [z] [i] [n] [g] [.]

Word Level

[This] [is] [tokenizing] [.]

Subword Level

[This] [is] [token] [izing] [.]

Text Analysis: Tokenization

❖ Subword Tokenization

➤ Byte-pair Encoding (BPE)

aaabdaaabc

aaabdaaabc Replace aa with Z

ZabdZabac Replace ab with Y

ZabdZabac Replace ab with Y

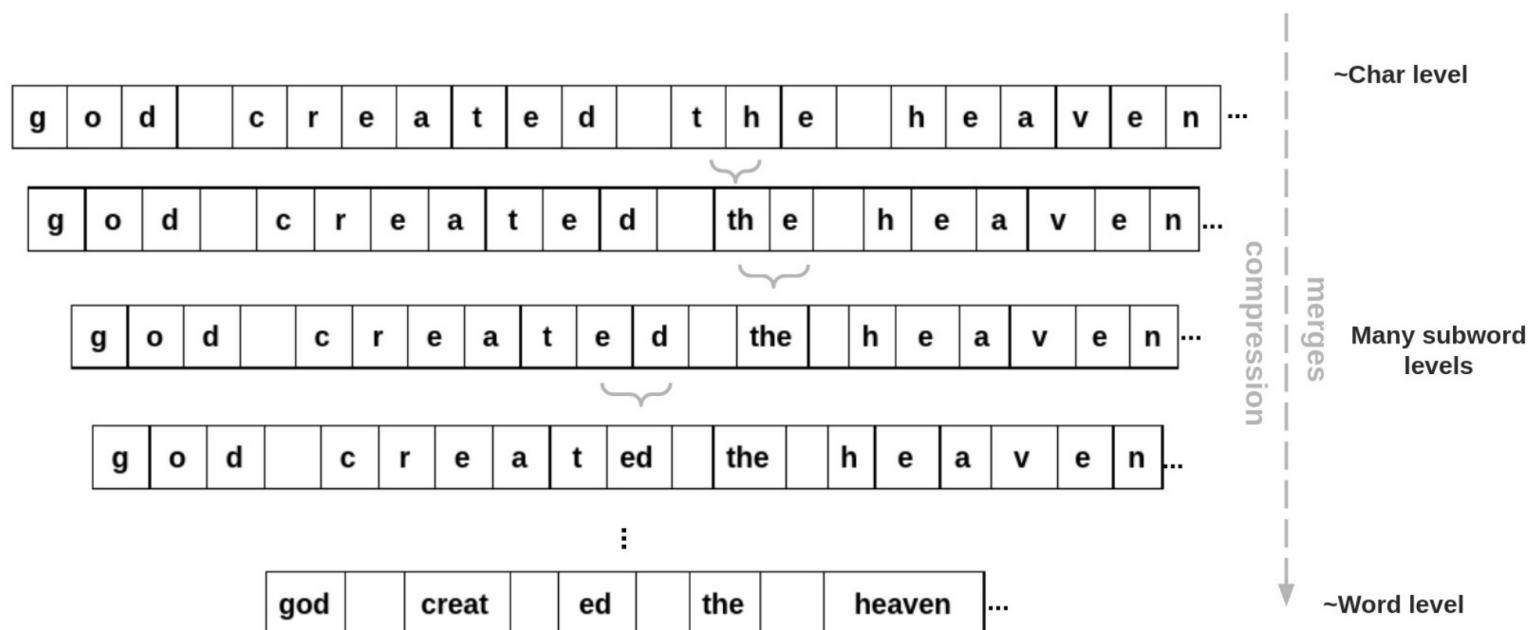
ZYdZYac

...

Text Analysis: Tokenization

❖ Subword Tokenization

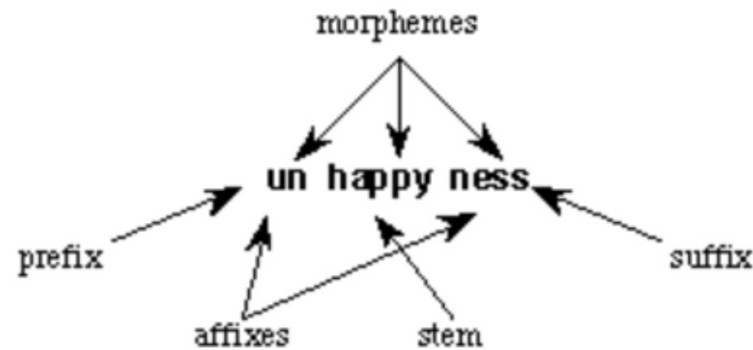
➤ Byte-pair Encoding (BPE)



Text Analysis: Tokenization

❖ Tokenization in Languages without Spaces

- Use Morphological Analyses – The study of the structure and formation of words



Text Analysis: Normalization

❖ Text Normalization

- Transform the text into a pronounceable state
- Detect non-standard words: rewrite as standard words

Which tokens here need normalization?

Example Text from BBC news:

Staff on the Caledonian Sleeper will hold two 24-hour strikes. One from 11:59 on Sunday 31 October and one on Thursday 11 November.

Unite Scotland has also said about 1,000 workers across the Stagecoach Group had backed strike action at the end of October which would affect COP26 bus travel in central Scotland.

Text Analysis: Normalization

❖ Text Normalization

- Transform the text into a pronounceable state
- Detect non-standard words: rewrite as standard words

Which tokens here need normalization?

Example Text from BBC news:

Staff on the Caledonian Sleeper will hold two 24-hour strikes. One from 11:59 on Sunday 31 October and one on Thursday 11 November.

Unite Scotland has also said about 1,000 workers across the Stagecoach Group had backed strike action at the end of October which would affect COP26 bus travel in central Scotland.

Text Analysis: Normalization

❖ Text Normalization

- Transform the text into a pronounceable state
- Detect non-standard words: rewrite as standard words

Text from BBC news:

Which tokens here need normalization?

Staff on the Caledonian Sleeper will hold two **twenty four hour** strikes. One from **eleven fifty nine** on Sunday **thirty first of** October and one on Thursday **eleventh of** November.

Unite Scotland has also said about **one thousand** workers across the Stagecoach Group had backed strike action at the end of October which would affect **COP twenty six** bus travel in central Scotland.

Text Analysis: Normalization

❖ Non-standard Words

- Numbers: 1,280, 1.50, 2024
 - Non-letter symbols: \$, %, &
 - Acronyms: UIT, ISE, DS, IT
 - Abbreviations: Dr, St, kHz
- ❖ Pronunciations don't always correspond to the ordering of the symbols: £1.50 → One pound fifty
 - ❖ Different pronunciation for different types of numbers: e.g. years versus measurements: the year 2022 v 2022 cm
 - ❖ Different rules for different locales : e.g. fifteen hundred (and) twenty
 - ❖ Acronyms may be pronounced as separate letters or as “words”

Text Analysis: Disambiguation

❖ Word Sense Ambiguity

➤ Words that are written the same way can have different pronunciations

- She **records** her **records** in the office
Verb *Noun*

- My **bass** guitar has a picture of a sea **bass** on it
Noun *Noun*

➤ We can disambiguate some cases by their Part Of Speech (POS)

Text Analysis: Disambiguation

❖ Punctuation Ambiguity

- TTS systems (still!) often work on a sentence-by-sentence basis
- So sentence segmentation is required
- But a full stop “.” doesn’t always represent the end of a sentence and sentences don’t always end with full stops:
- ❑ Eg: “The lecturer lives in Fife Dr. Richmond is actually a linguist by training”

Text Analysis: Disambiguation

❖ How to Resolve Ambiguity?

- Hand-written rules
 - ❑ Eg: York St. versus St. Lucas
 - St. → [Capitalized Word] / Street / _
 - St. → _ / Saint / [Capitalized word]

Phonetic Analysis

❖ Mapping written words to pronunciations (phonemes)

- Lookup phonetic transcript in a pronunciation dictionary (i.e., lexicon)
- Use hand-written letter-to-sound rules (grapheme-to-phoneme) => Requires expert knowledge
- Learn rules from the data - requires machine learning methods
 - ❑ Eg: Decision Trees, state-of-the-art neural network models

Phonetic Analysis

❖ Pronunciation Dictionary

- The pronunciation of a word can be represented as a sequence of phones
- The standard phonetic representation for transcribing the world's languages is the International Phonetic Alphabet (IPA)



tomato

/tə'meɪ.tou/

Phonetic Analysis

❖ Pronunciation Dictionary

- The pronunciation of a word can be represented as a sequence of phones
- The standard phonetic representation for transcribing the world's languages is the International Phonetic Alphabet (IPA)

ɪ READ	ɪ SIT	ʊ BOOK	u: TOO	ɪə HERE	eɪ DAY	John & Sarah Free Materials 1996	
e MEN	ə AMERICA	ɜ: WORD	ɔ: SORT	ʊə TOUR	ɔɪ BOY		
æ CAT	ʌ BUT	ɑ: PART	ɒ NOT	eə WEAR	aɪ MY		
p PIG	b BED	t TIME	d DO	tʃ CHURCH	dʒ JUDGE	k KILO	g GO
f FIVE	v VERY	θ THINK	ð THE	s SIX	z ZOO	ʃ SHORT	ʒ CASUAL
m MILK	n NO	ŋ SING	h HELLO	l LIVE	r READ	w WINDOW	j YES

Phonetic Analysis

❖ Pronunciation Dictionary

- Multiple Pronunciation: Same writing may have different pronunciation (i.e., due to different accent)



tomato

/tə'meɪ.tou/ /tə'maɪ.tou/

Phonetic Analysis

❖ Pronunciation Dictionary

- Letter-to-sound rules (grapheme-to-phoneme)
 - ❑ **Grapheme:** a letter or a group of letters that represent a single phoneme
 - ❑ **Phoneme:** the smallest unit of sound that can distinguish one word from another in a particular language
- ❑ Eg: When a child says the sound /t/ this is a phoneme, but when they write the letter 't' this is a grapheme

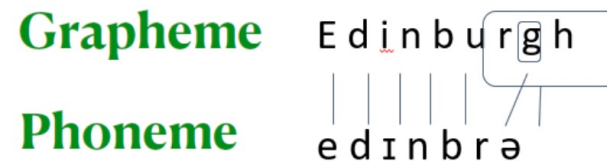
Grapheme t o m a t o

Phoneme /t ə' m eɪ. t ou/

Phonetic Analysis

❖ Pronunciation Dictionary

- Letter-to-sound rules (grapheme-to-phoneme)
 - ❑ **Grapheme:** a letter or a group of letters that represent a single phoneme
 - ❑ **Phoneme:** the smallest unit of sound that can distinguish one word from another in a particular language
- ❑ Eg: The mapping between letters used for writing (graphemes) and pronunciation (phonemes) isn't always straightforward (especially for English)



Phonetic Analysis

❖ Pronunciation Dictionary

- Letter-to-sound rules (grapheme-to-phoneme)
 - ❑ **Grapheme:** a letter or a group of letters that represent a single phoneme
 - ❑ **Phoneme:** the smallest unit of sound that can distinguish one word from another in a particular language
- ❑ Eg: From words to utterances

Grapheme	I live in Edinburgh
Phoneme	aɪ lɪv ɪn ɛdɪnbərə

Phonetic Analysis

❖ Pronunciation Dictionary

➤ Different Phoneme-set

- [Edinburgh](#)
- [IPA: edɪmbɹə](#)
- [ARPABET: EH1 D AH0 N B R AX0](#)
- [Unilex-edi: e d i m b r @](#)
- [Unilex-edi full: \(\(\(e\)1\) \(\(d i m\)0\) \(\(b r @\)0\)\)](#)

Phonetic Analysis

❖ Pronunciation Dictionary

➤ Different Phoneme-set

- [Edinburgh](#)
- [IPA: edɪmbɹə](#)
- [ARPABET: EH1 D AH0 N B R AX0](#)
- [Unilex-edi: e d i m b r @](#)
- [Unilex-edi full: \(\(\(e\)1\) \(\(d i m\)0\) \(\(b r @\)0\)\)](#)

Practice at Class

❖ Vietnamese Pronunciation Dictionary:

- ❑ Vowels, relevant IPA, and examples
- ❑ Consonants, relevant IPA, and examples