FROM TEXT TO SPEECH
THE MITALK SYSTEM

Presented by
Najeeb Khan
2013-5-17

INTRODUCTION

INTRODUCTION

- We are concerned with describing a successful approach to the conversion of unrestricted English text to speech

INTRODUCTION

- We are concerned with describing a successful approach to the conversion of unrestricted English text to speech
- In order to get a view of the spectrum of speech synthesis approaches it is useful to consider them as the result of four different constraints which determine a design space for all possible speech output schemes

CONSTRAINTS ON SPEECH SYNTHESIS

- Task

CONSTRAINTS ON SPEECH SYNTHESIS

- Task
  - The application task determine the nature of the speech capability that must be provided

CONSTRAINTS ON SPEECH SYNTHESIS

- Task
  - The application task determine the nature of the speech capability that must be provided
  - Recorded speech can be used for fixed small number of utterances

CONSTRAINTS ON SPEECH SYNTHESIS

- Task
  - The application task determine the nature of the speech capability that must be provided
  - Recorded speech can be used for fixed small number of utterances
  - If the task is to simulate the human cognitive process of reading aloud, then an entirely different range of techniques is required

CONSTRAINTS ON SPEECH SYNTHESIS

- Human Vocal Apparatus

CONSTRAINTS ON SPEECH SYNTHESIS

- Human Vocal Apparatus
  - All systems must produce as output a speech waveform, but it is not an arbitrary signal

CONSTRAINTS ON SPEECH SYNTHESIS

- Human Vocal Apparatus
  - All systems must produce as output a speech waveform, but it is not an arbitrary signal
  - Efficient and insightful representation of the speech signal as the result of a signal source in the vocal tract exciting the vocal tract system function

CONSTRAINTS ON SPEECH SYNTHESIS

- Human Vocal Apparatus
  - All systems must produce as output a speech waveform, but it is not an arbitrary signal
  - Efficient and insightful representation of the speech signal as the result of a signal source in the vocal tract exciting the vocal tract system function
  - The human vocal tract is responsible for much of the co-articulatory smoothing or encoding that makes the relation between the underlying phonetic transcription and the speech waveform so difficult to characterize

CONSTRAINTS ON SPEECH SYNTHESIS

- Language Structure

CONSTRAINTS ON SPEECH SYNTHESIS

- Language Structure
  - Units and structures which linguists use to describe and explain language provide the base in terms of which the speech waveform can be characterized and constructed

CONSTRAINTS ON SPEECH SYNTHESIS

- Language Structure
  - Units and structures which linguists use to describe and explain language provide the base in terms of which the speech waveform can be characterized and constructed
  - The basic phonological laws, stress rules, morphological and syntactic structures and phono-tactic constraints all find their use in determining the speech output

CONSTRAINTS ON SPEECH SYNTHESIS

- Technology

CONSTRAINTS ON SPEECH SYNTHESIS

- Technology
  - Our ability to model and construct speech output devices is strongly conditioned by the current and past technology

CONSTRAINTS ON SPEECH SYNTHESIS

- Technology
  - Our ability to model and construct speech output devices is strongly conditioned by the current and past technology
  - Speech science has profited greatly from variety of technologies including x-rays, motion pictures, the sonograph modern filter and sampled data theory and the modern computer

SYNTHESIS TECHNIQUES

- Waveform Coding

SYNTHESIS TECHNIQUES

- Waveform Coding
  - The simplest strategy would be to simply record the required speech

SYNTHESIS TECHNIQUES

- Waveform Coding
  - The simplest strategy would be to simply record the required speech
  - The speech in general will be of high quality

SYNTHESIS TECHNIQUES

- Waveform Coding
  - The simplest strategy would be to simply record the required speech
  - The speech in general will be of high quality
  - Since memory size is a major limitation of these schemes, efforts have been made to cut down the number of bits per sample used for digital storage while still retaining good quality of speech

SYNTHESIS TECHNIQUES

- Waveform Coding
  - The simplest strategy would be to simply record the required speech
  - The speech in general will be of high quality
  - Since memory size is a major limitation of these schemes, efforts have been made to cut down the number of bits per sample used for digital storage while still retaining good quality of speech
  - If messages must be concatenated, then it is extremely difficult to produce good quality speech because of problems at boundaries

SYNTHESIS TECHNIQUES

- Parametric Representation

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by
    - Summing up weighted waveforms

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by
    - Summing up weighted waveforms
    - Resonances that are excited by sound sources

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by
    - Summing up weighted waveforms
    - Resonances that are excited by sound sources
    - Vocal tract shape can be simulated along with appropriate acoustic excitation

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by
    - Summing up weighted waveforms
    - Resonances that are excited by sound sources
    - Vocal tract shape can be simulated along with appropriate acoustic excitation
  - The parametric representation represents an abstraction on the speech waveform to a level where attributes that contribute to speech quality (F_i B F_0 A_i) can be manipulated

SYNTHESIS TECHNIQUES

- Parametric Representation
  - The speech waveform can be formed by
    - Summing up weighted waveforms
    - Resonances that are excited by sound sources
    - Vocal tract shape can be simulated along with appropriate acoustic excitation
  - The parametric representation represents an abstraction on the speech waveform to a level where attributes that contribute to speech quality (F_i B F_0 A_i) can be manipulated
  - Parametric Representation uses the a knowledge of the human production of speech but little (if any) use is made of the linguistic structure of the language

SYNTHESIS TECHNIQUES

- Synthesis by Rule

SYNTHESIS TECHNIQUES

- Synthesis by Rule
  - The synthesis process has been abstracted even further to a set of rules which seek to compute the needed parameters for the speech production model from an input phonetic description

SYNTHESIS TECHNIQUES

- Synthesis by Rule
  - The synthesis process has been abstracted even further to a set of rules which seek to compute the needed parameters for the speech production model from an input phonetic description
  - Synthesis by rule techniques can utilize a very low bit rate message description (<100bits/sec)

SYNTHESIS TECHNIQUES

- Synthesis by Rule
  - The synthesis process has been abstracted even further to a set of rules which seek to compute the needed parameters for the speech production model from an input phonetic description
  - Synthesis by rule techniques can utilize a very low bit rate message description (<100bits/sec)
  - The synthesis by rule is well suited to the needs of converting unrestricted text to speech

TEXT TO SPEECH CONVERSION

- The synthesis by rule procedure is needed for TTS conversion

TEXT TO SPEECH CONVERSION

- The synthesis by rule procedure is needed for TTS conversion
- The first step is to analyze the text to obtain a phonetic transcription

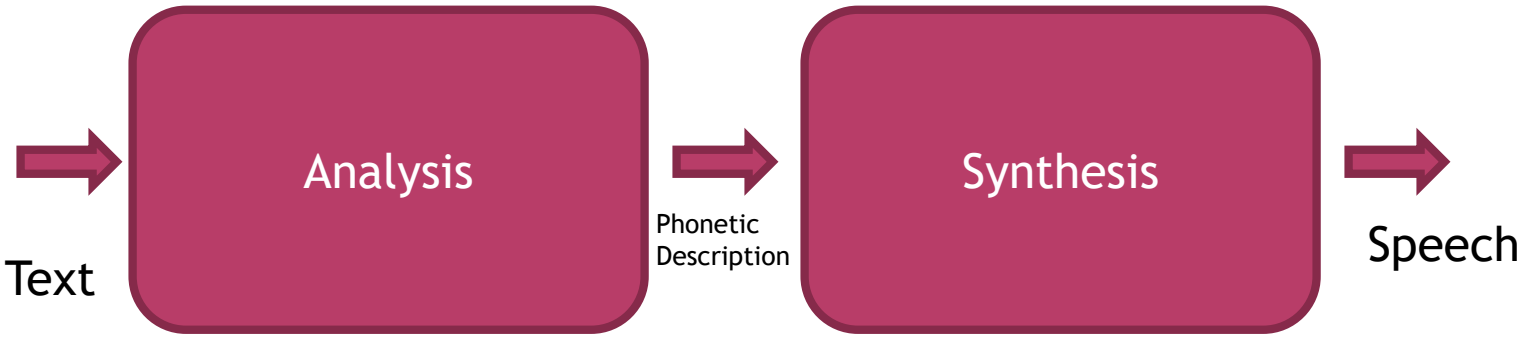
TEXT TO SPEECH CONVERSION

- The synthesis by rule procedure is needed for TTS conversion
- The first step is to analyze the text to obtain a phonetic transcription
- The phonetic transcription is subjected to a synthesis procedure to yield the output speech waveform

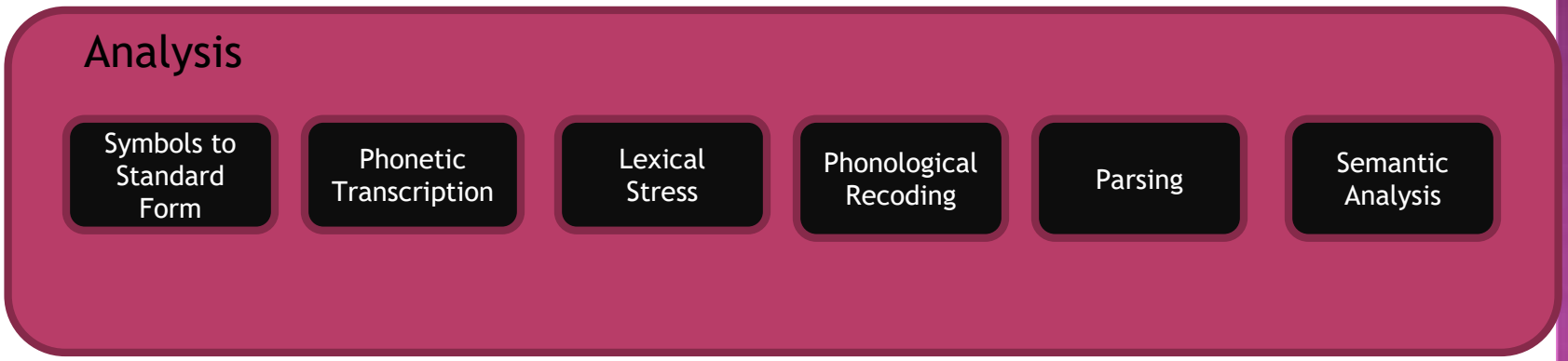
TEXT TO SPEECH CONVERSION

- The synthesis by rule procedure is needed for TTS conversion
- The first step is to analyze the text to obtain a phonetic transcription
- The phonetic transcription is subjected to a synthesis procedure to yield the output speech waveform
- The analysis of the text is heavily linguistic in nature, involving a determination of the underlying phonemic, syllabic, morphemic, syntactic form of the message plus whatever semantic information can be obtained

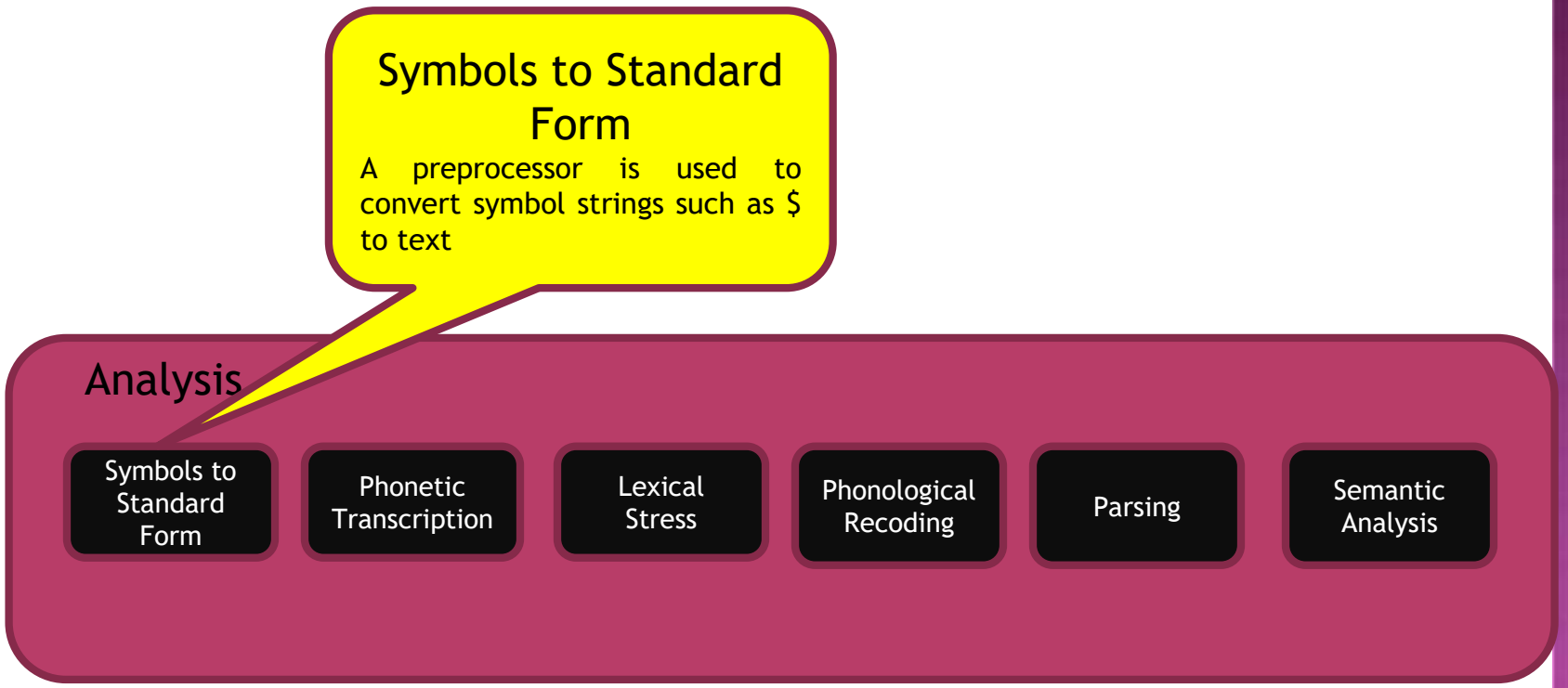
FUNCTIONAL OUTLINE OF MITALK



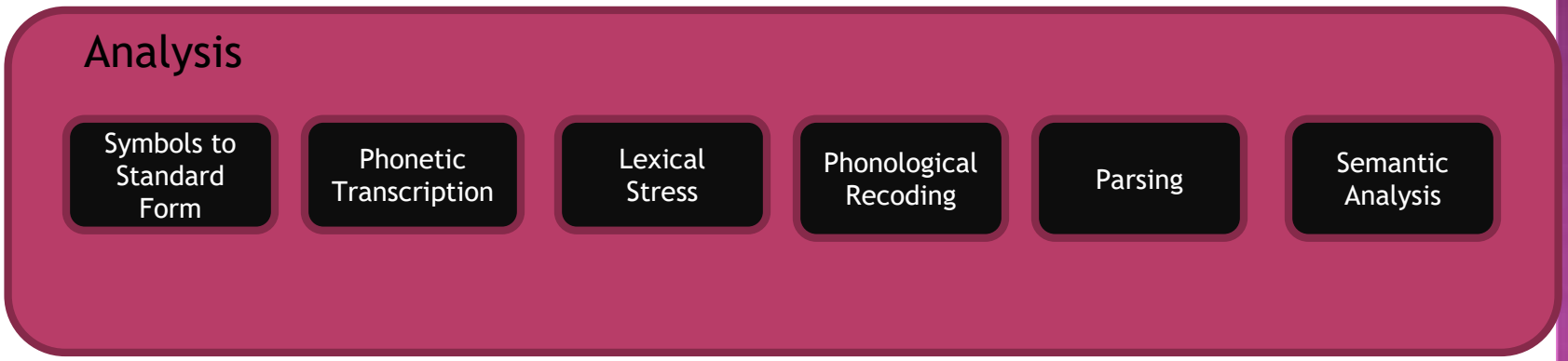
FUNCTIONAL OUTLINE OF MITALK



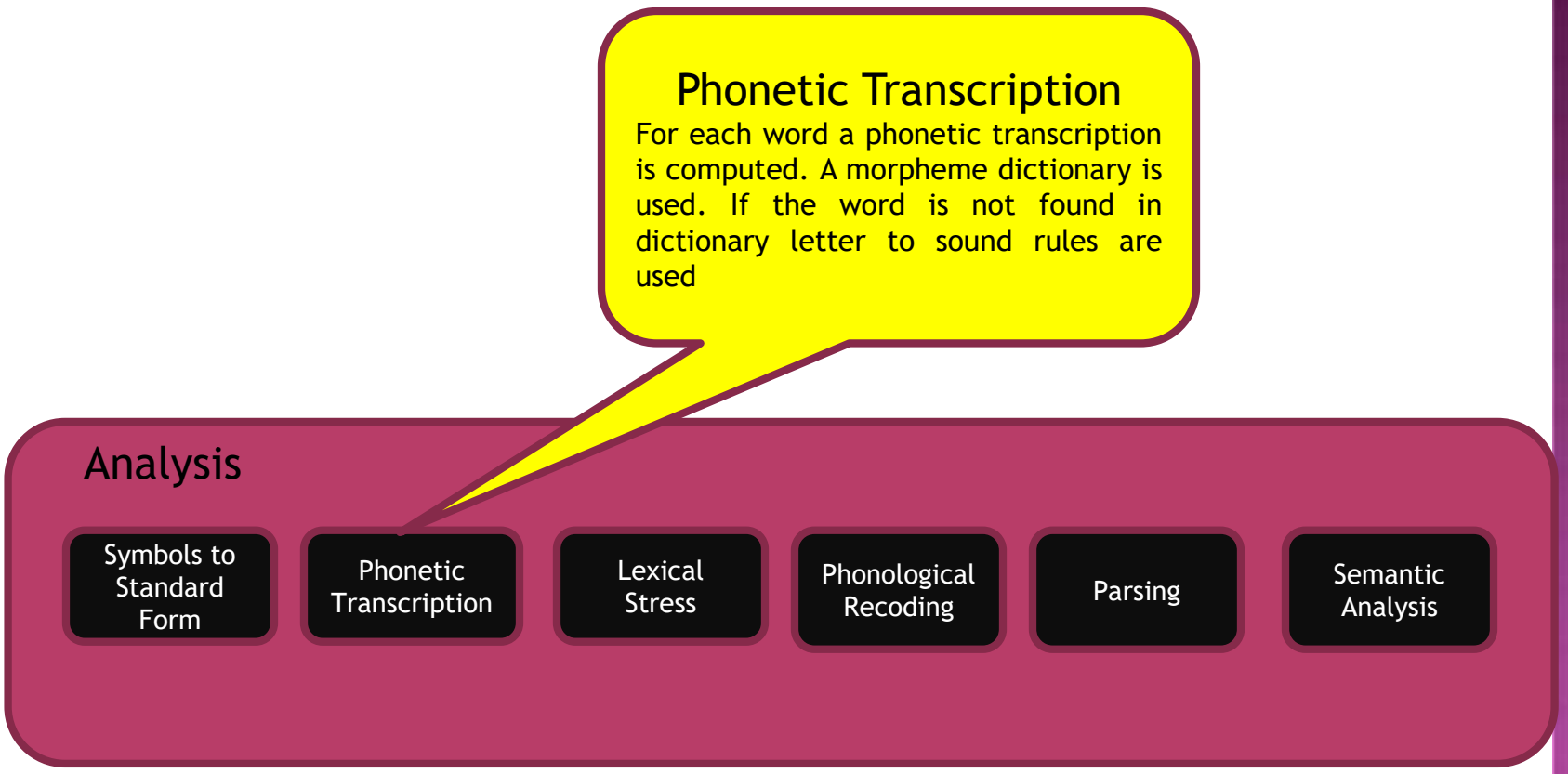
FUNCTIONAL OUTLINE OF MITALK



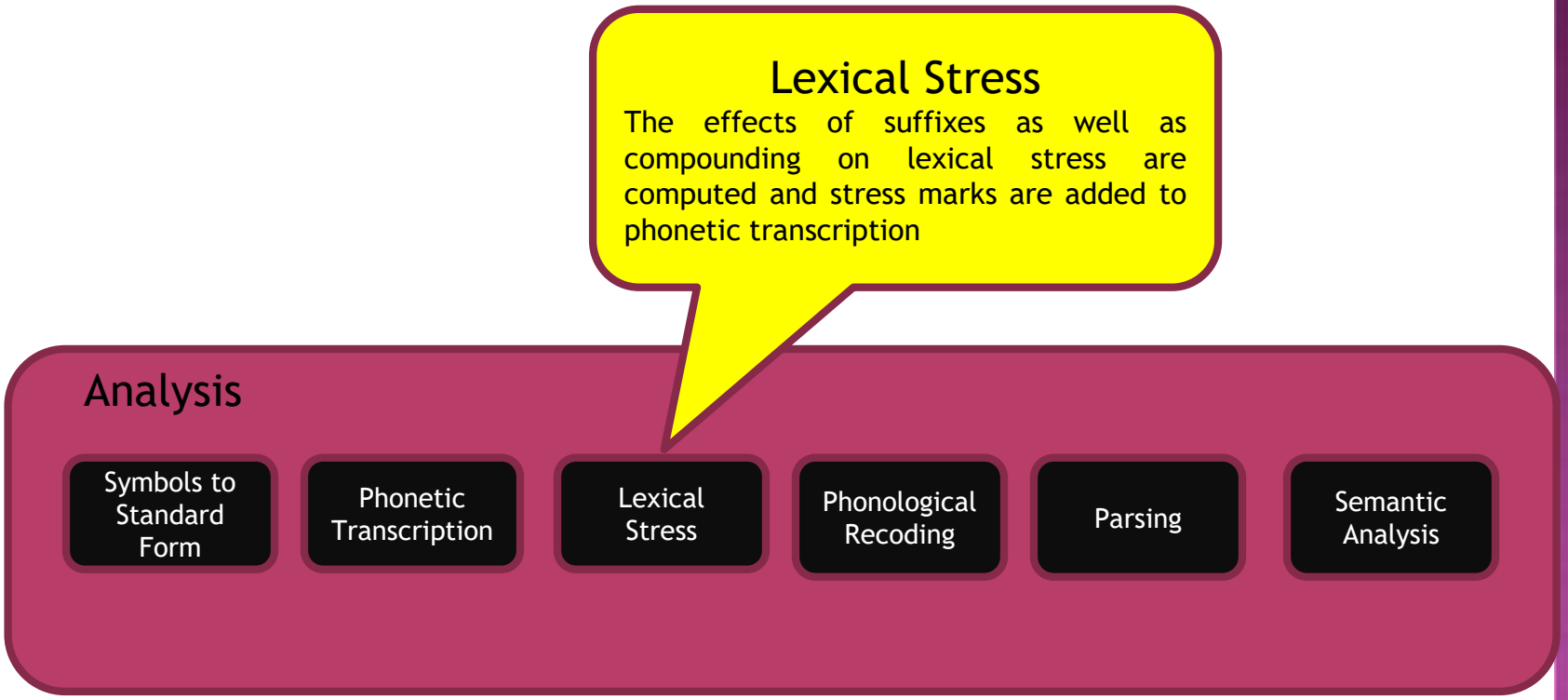
FUNCTIONAL OUTLINE OF MITALK



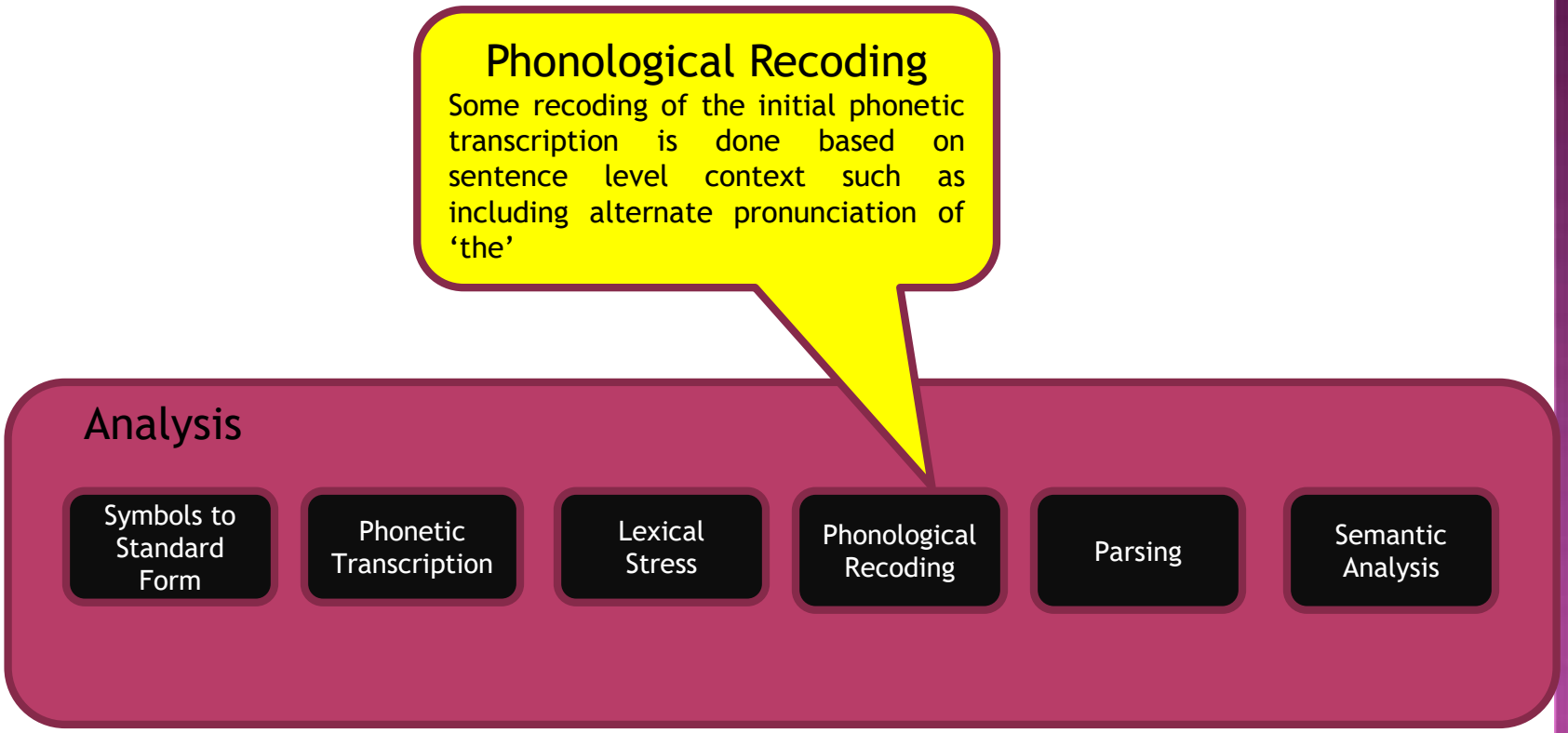
FUNCTIONAL OUTLINE OF MITALK



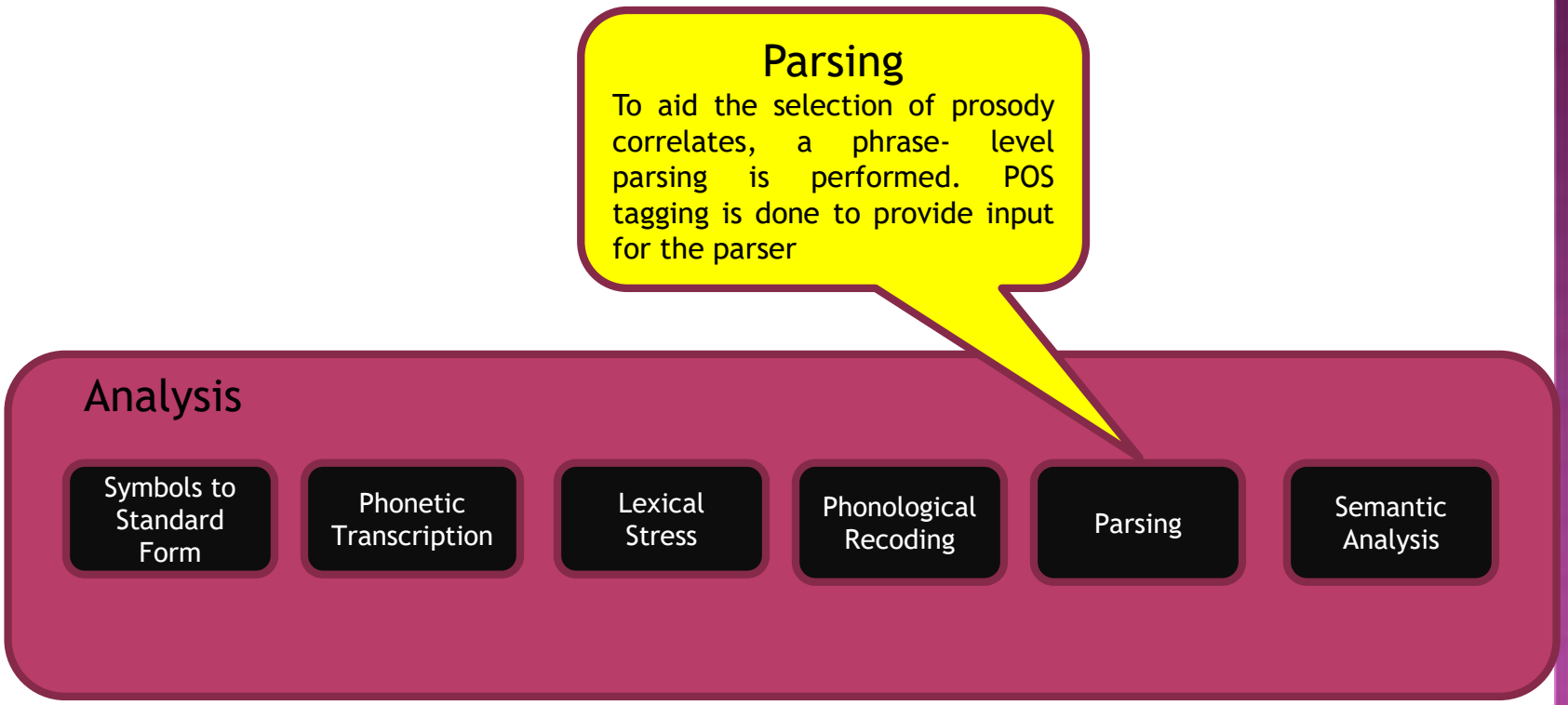
FUNCTIONAL OUTLINE OF MITALK



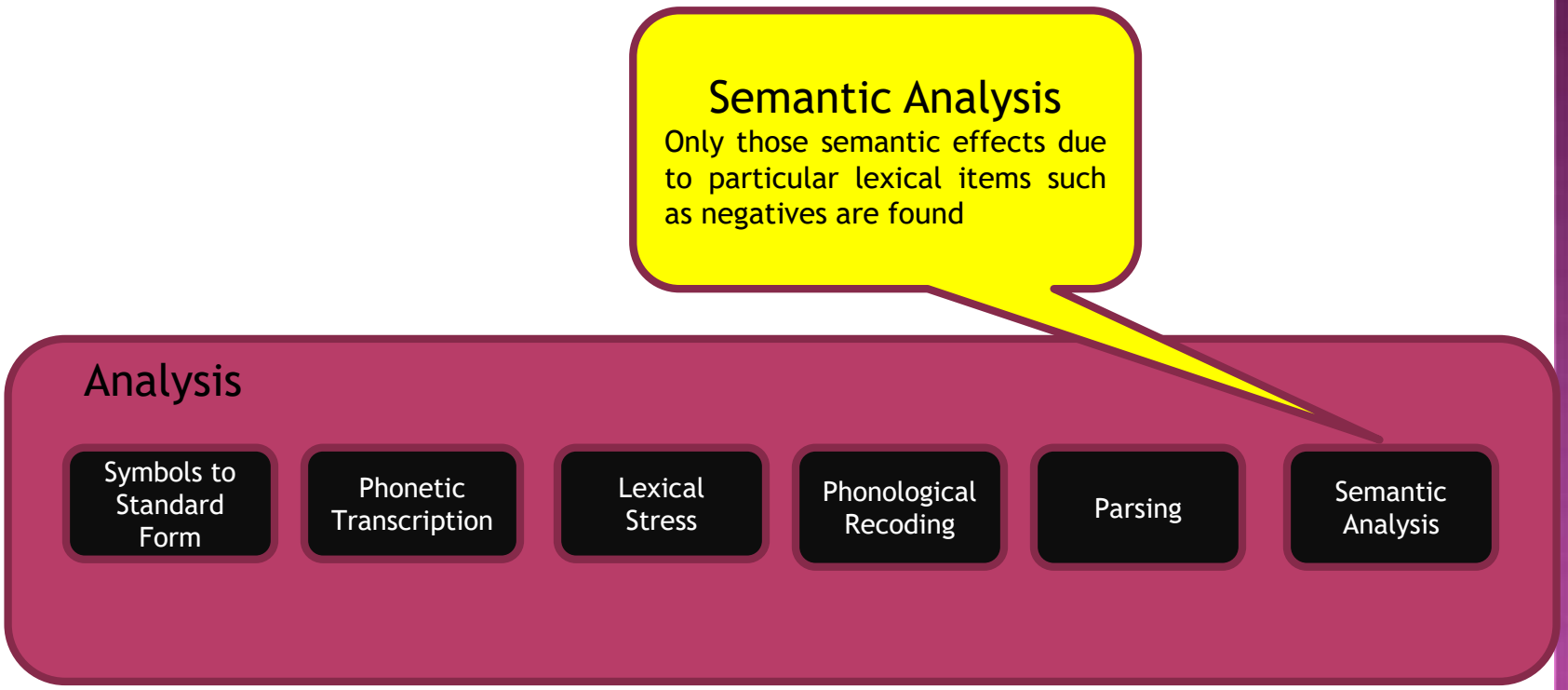
FUNCTIONAL OUTLINE OF MITALK



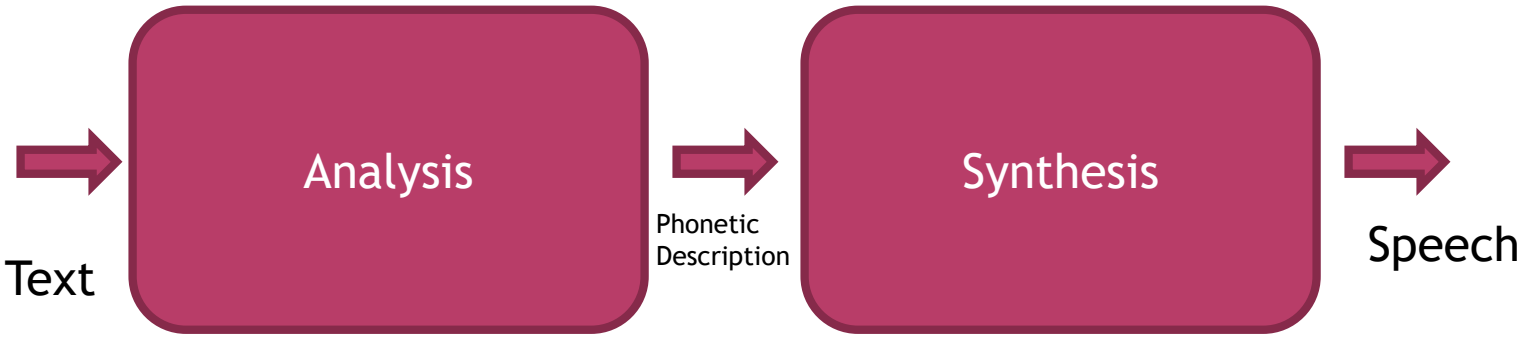
FUNCTIONAL OUTLINE OF MITALK



FUNCTIONAL OUTLINE OF MITALK

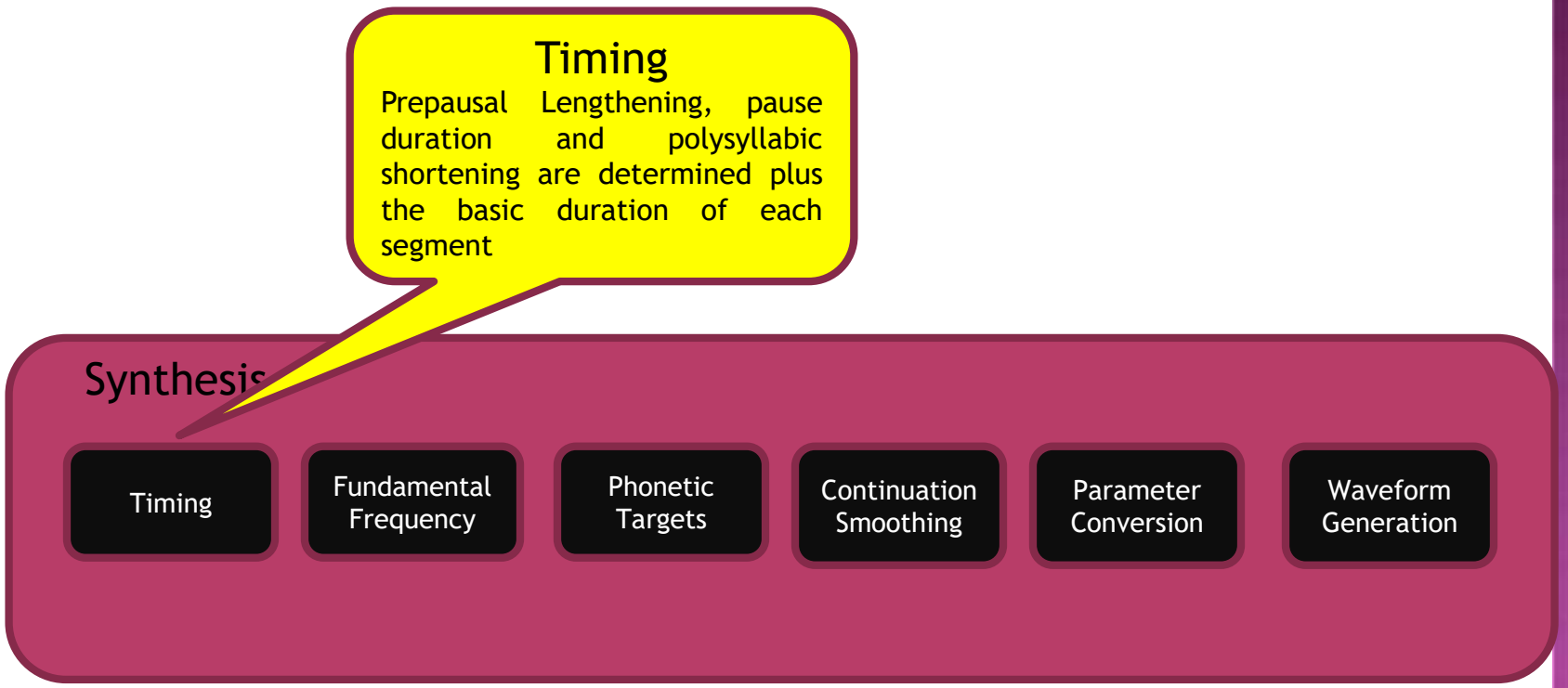


FUNCTIONAL OUTLINE OF MITALK

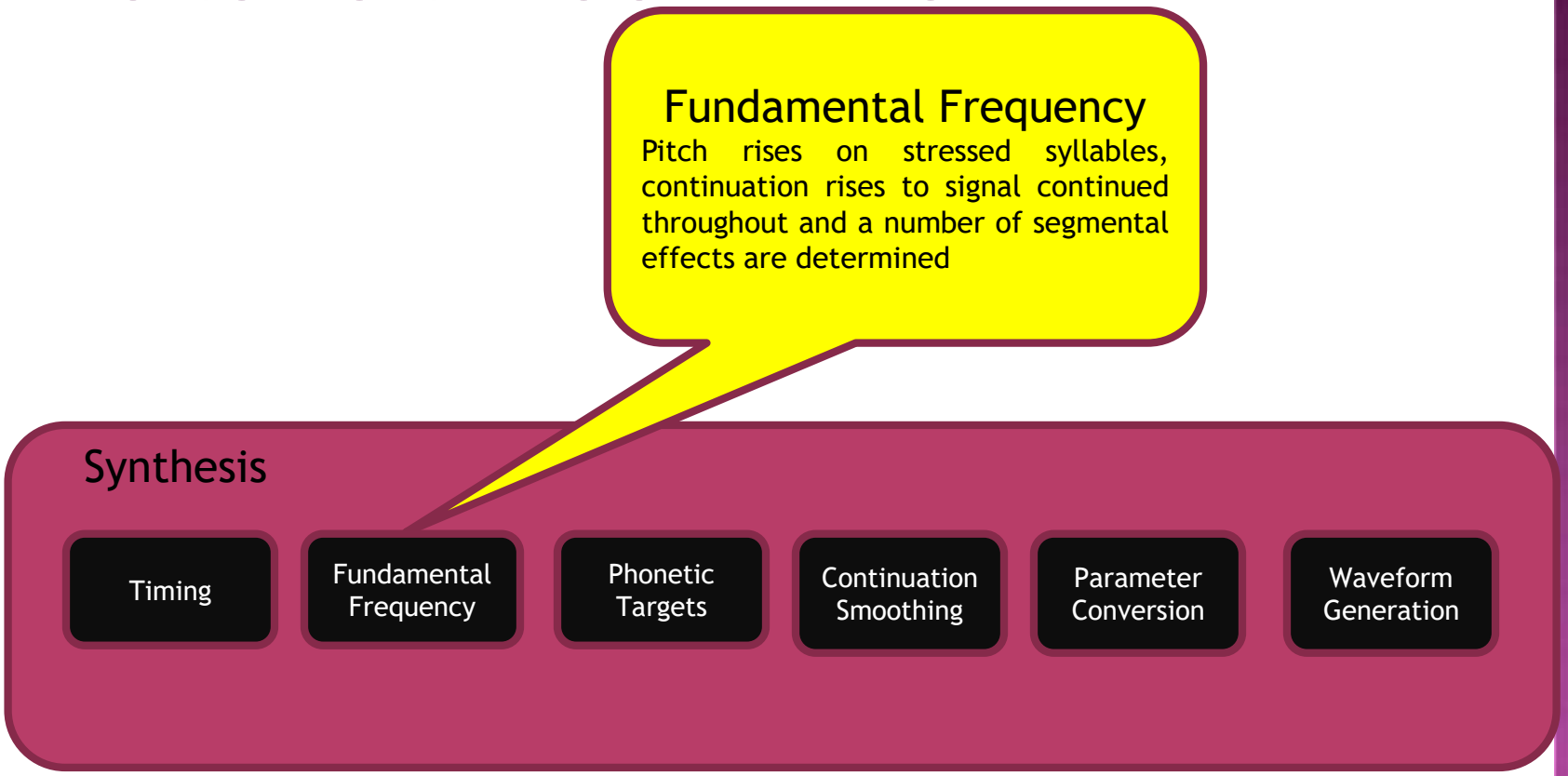


FUNCTIONAL OUTLINE OF MITALK

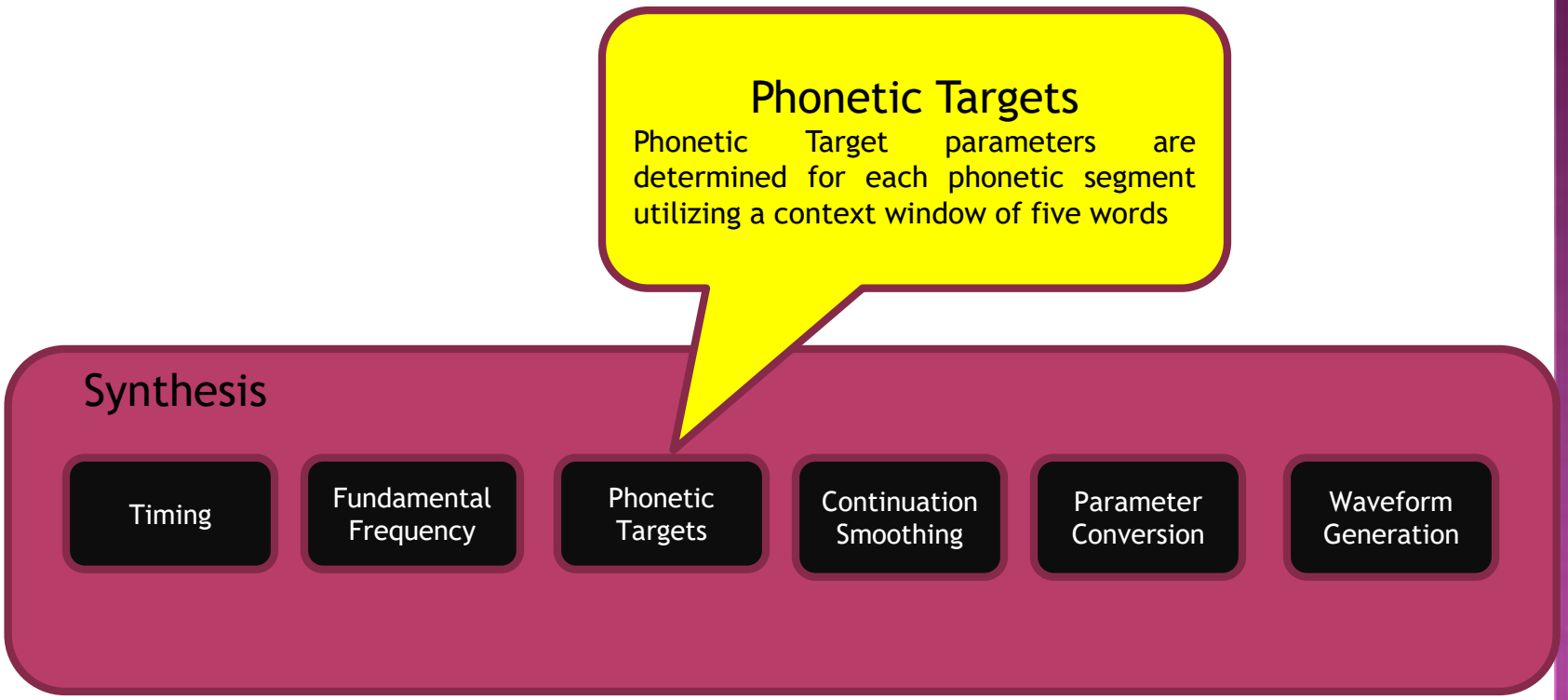
FUNCTIONAL OUTLINE OF MITALK



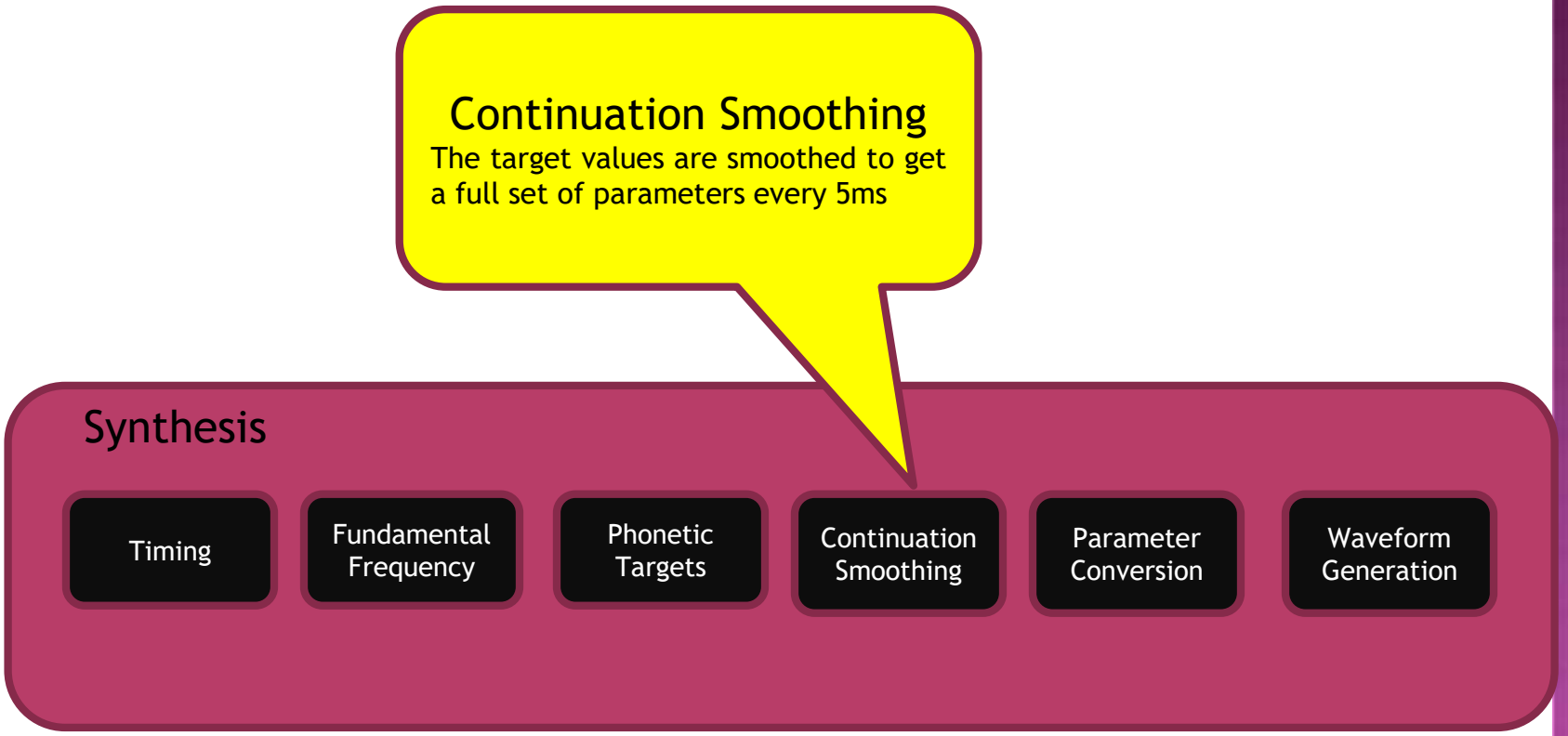
FUNCTIONAL OUTLINE OF MITALK



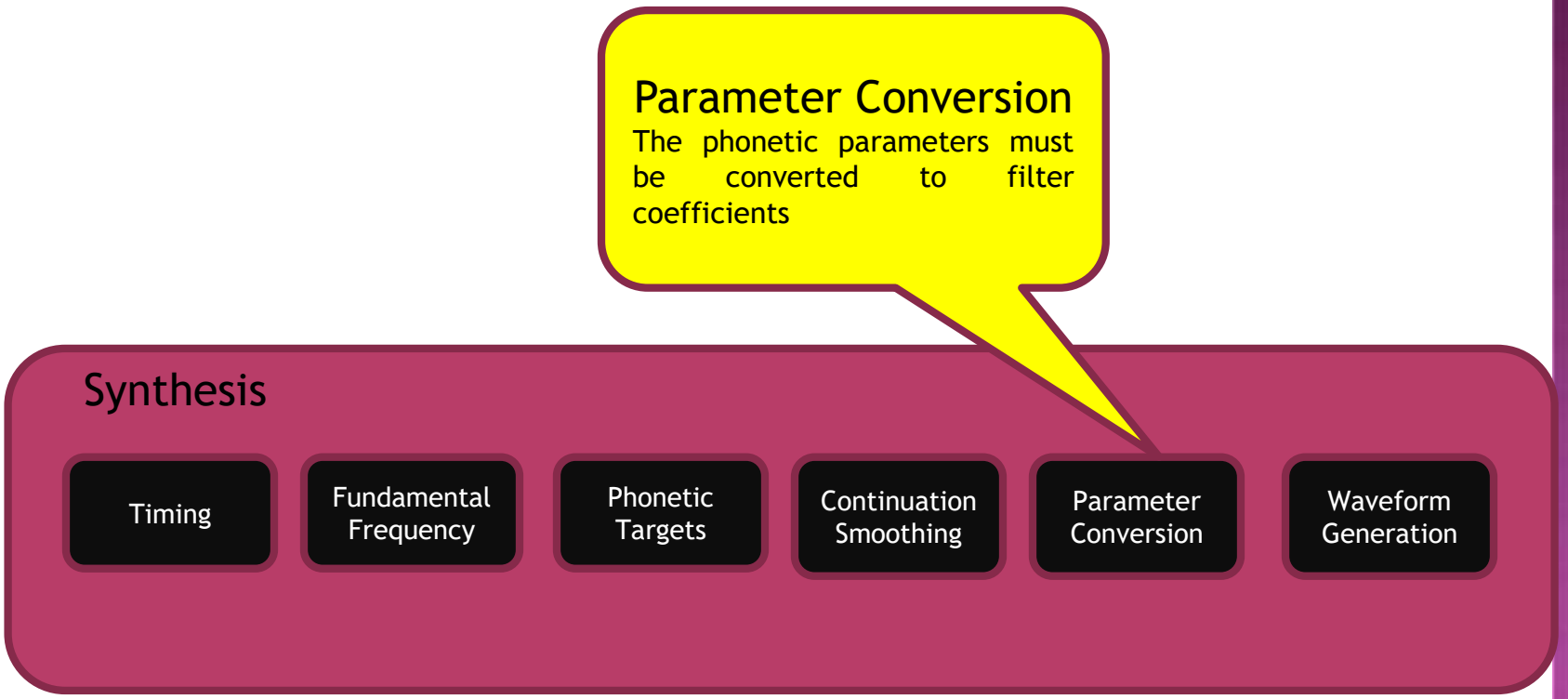
FUNCTIONAL OUTLINE OF MITALK



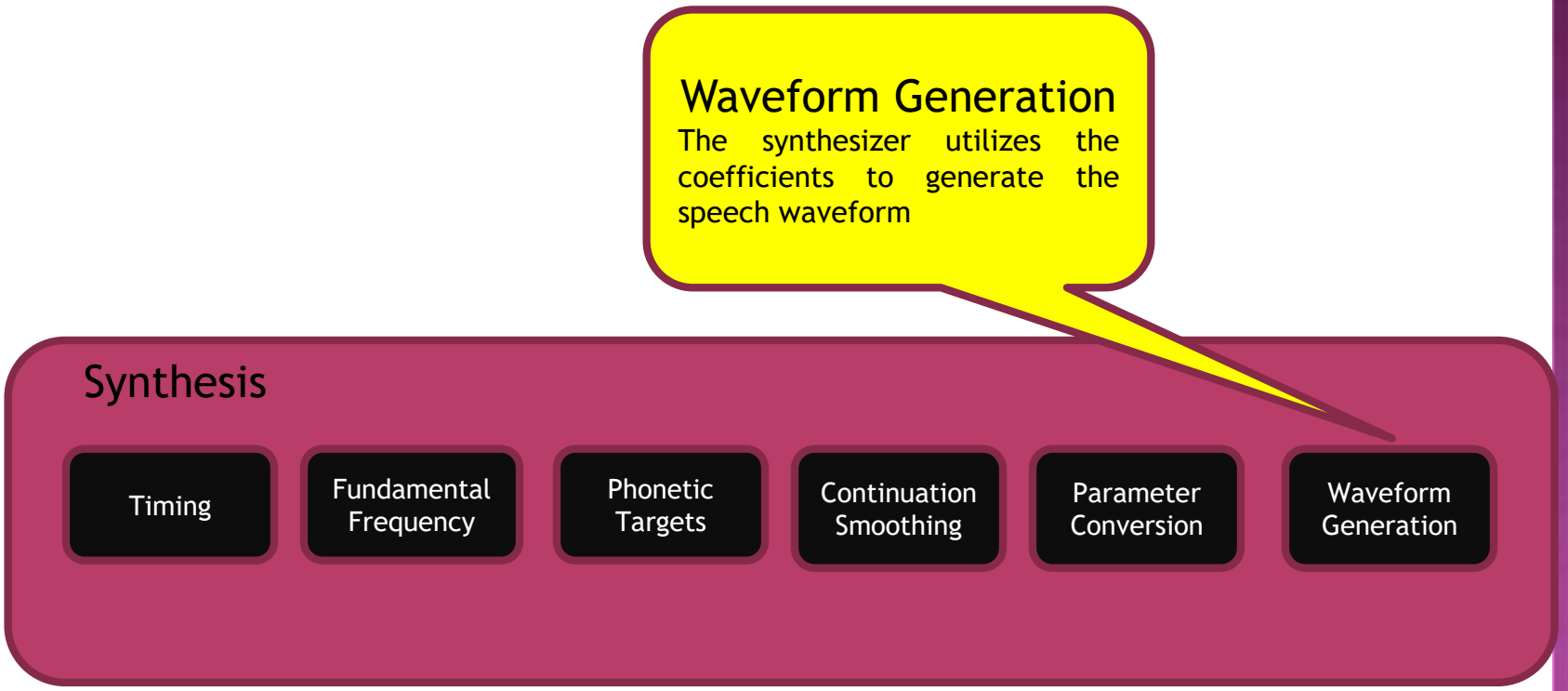
FUNCTIONAL OUTLINE OF MITALK



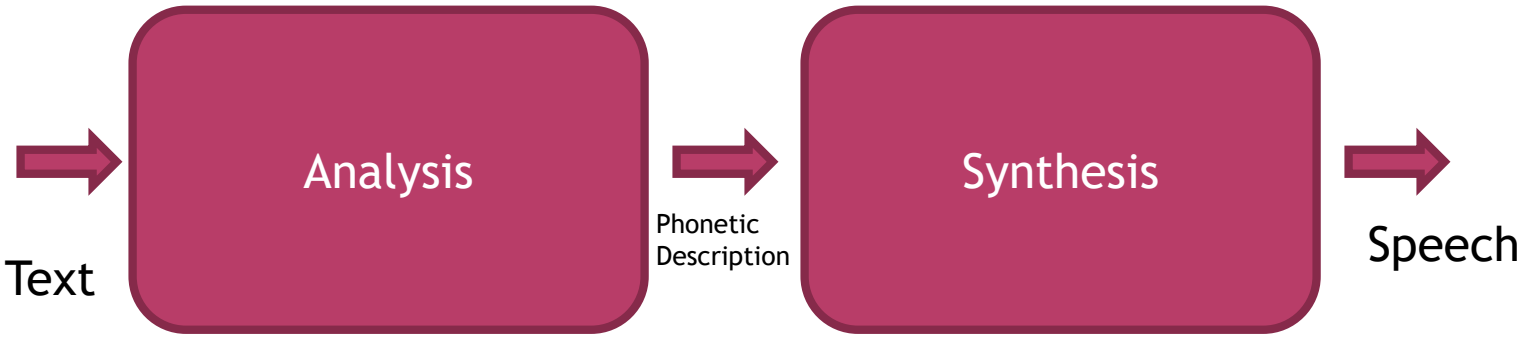
FUNCTIONAL OUTLINE OF MITALK



FUNCTIONAL OUTLINE OF MITALK



FUNCTIONAL OUTLINE OF MITALK



TEXT PREPROCESSING

- Unrestricted text may contain a wide variety of symbols, abbreviations and conventions.

TEXT PREPROCESSING

- Unrestricted text may contain a wide variety of symbols, abbreviations and conventions.
- In order to convert text to speech it is necessary to find an appropriate expression in words for such symbols

TEXT PREPROCESSING

- Unrestricted text may contain a wide variety of symbols, abbreviations and conventions.
- In order to convert text to speech it is necessary to find an appropriate expression in words for such symbols
- FORMAT module of the MITalk system performs the conversion of unrestricted text to a sequence of words and punctuation recognizable by the latter modules

TEXT PREPROCESSING

- Input

TEXT PREPROCESSING

- Input
  - FORMAT accepts the original unrestricted English text to be analyzed

TEXT PREPROCESSING

- Input
  - FORMAT accepts the original unrestricted English text to be analyzed
  - Any character that is not recognized by FORMAT causes a warning message and is treated as a space

TEXT PREPROCESSING

- Input
  - FORMAT accepts the original unrestricted English text to be analyzed
  - Any character that is not recognized by FORMAT causes a warning message and is treated as a space
  - Words are allowed 40 characters each and the maximum number of words per sentence is 200

TEXT PREPROCESSING

- Output

TEXT PREPROCESSING

- Output
  - The output of FORMAT is a sequence of words and punctuation marks

TEXT PREPROCESSING

- Output
  - The output of FORMAT is a sequence of words and punctuation marks
  - FORMAT scans each input from left to right and converts each recognized construct (word, number, symbol etc.) into an appropriate word or sequence of words

TEXT PREPROCESSING

- Output

```
Mr. Jones gets 35.3%.
  FORMAT:  MISTER
  FORMAT:  JONES
  FORMAT:  GETS
  FORMAT:  THIRTY
  FORMAT:  FIVE
  FORMAT:  POINT
  FORMAT:  THREE
  FORMAT:  PERCENT
  FORMAT:   .
  FORMAT:   .
```

TEXT PREPROCESSING

- Formatting Operations

TEXT PREPROCESSING

- Formatting Operations
  - An additional pause is inserted after each sentence longer than 5 words

TEXT PREPROCESSING

- Formatting Operations
  - An additional pause is inserted after each sentence longer than 5 words
  - Not all periods denote the end of a sentence

TEXT PREPROCESSING

- Formatting Operations
  - An additional pause is inserted after each sentence longer than 5 words
  - Not all periods denote the end of a sentence
  - If a period ends an abbreviation, then it is taken as an end of sentence marker if it is at the end of a line and if it is followed by whitespace and a capitalized word

TEXT PREPROCESSING

- Formatting Operations
  - An additional pause is inserted after each sentence longer than 5 words
  - Not all periods denote the end of a sentence
  - If a period ends an abbreviation, then it is taken as an end of sentence marker if it is at the end of a line and if it is followed by whitespace and a capitalized word
  - A period inside a numeric string is considered as decimal point

TEXT PREPROCESSING

- Formatting Operations
  - An additional pause is inserted after each sentence longer than 5 words
  - Not all periods denote the end of a sentence
  - If a period ends an abbreviation, then it is taken as an end of sentence marker if it is at the end of a line and if it is followed by whitespace and a capitalized word
  - A period inside a numeric string is considered as decimal point

TEXT PREPROCESSING

- Formatting Operations

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols
    - A word is an alphabetic string delimited by a punctuation or whitespace character

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols
    - A word is an alphabetic string delimited by a punctuation or whitespace character
    - If a word is followed by a period, then FORMAT looks in a table of abbreviations to see if a translation is specified for that word

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols

Ms	→	MIZ
Mr	→	MISTER
Mrs	→	MIZZES
Dr	→	DOCTOR
Num	→	NUMBER
Jan	→	JANUARY
Feb	→	FEBRUARY
Mar	→	MARCH
Apr	→	APRIL
Aug	→	AUGUST
Sept	→	SEPTEMBER
Oct	→	OCTOBER
Nov	→	NOVEMBER
Dec	→	DECEMBER
etc	→	ET CETERA
Jr	→	JUNIOR
Prof	→	PROFESSOR

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols
    - A word is an alphabetic string delimited by a punctuation or whitespace character
    - If a word is followed by a period, then FORMAT looks in a table of abbreviations to see if a translation is specified for that word

TEXT PREPROCESSING

- Formatting Operations
  - Words Abbreviations and Special Symbols
    - A word is an alphabetic string delimited by a punctuation or whitespace character
    - If a word is followed by a period, then FORMAT looks in a table of abbreviations to see if a translation is specified for that word
    - A word that is in capital letters or which contains digits as well as letters is considered to be a symbol and is translated by pronouncing each character

TEXT PREPROCESSING

- Formatting Operations

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks
    - The apostrophe is included in the word if it appears after the last letter in the word and the last letter is an s

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks
    - The apostrophe is included in the word if it appears after the last letter in the word and the last letter is an s
    - An apostrophe in any other position is considered as single quotation mark and is output as a punctuation

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks
    - The apostrophe is included in the word if it appears after the last letter in the word and the last letter is an s
    - An apostrophe in any other position is considered as single quotation mark and is output as a punctuation
  - Hyphens and Dashes

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks
    - The apostrophe is included in the word if it appears after the last letter in the word and the last letter is an s
    - An apostrophe in any other position is considered as single quotation mark and is output as a punctuation
  - Hyphens and Dashes
    - If a dash character is embedded between two words it is considered to be a hyphen separating compound word elements

ting hyphen

TEXT PREPROCESSING

- Formatting Operations
  - Apostrophe and Single Quotation Marks
    - The apostrophe is included in the word if it appears after the last letter in the word and the last letter is an s
    - An apostrophe in any other position is considered as single quotation mark and is output as a punctuation
  - Hyphens and Dashes
    - If a dash character is embedded between two words it is considered to be a hyphen separating compound word elements
    - If a dash appear at the end of the last word on a line hen it is considered as the word split-

      ting hyphen

TEXT PREPROCESSING

- Formatting Operations

TEXT PREPROCESSING

- Formatting Operations
  - Numerals

TEXT PREPROCESSING

- Formatting Operations
  - Numerals

- 715 → SEVEN HUNDRED FIFTEEN
- 71.50 → SEVENTY ONE POINT FIVE OH
- 159,106 → ONE HUNDRED FIFTY NINE THOUSAND ONE HUNDRED SIX

TEXT PREPROCESSING

- Formatting Operations
  - Numerals

- 715 → SEVEN HUNDRED FIFTEEN
- 71.50 → SEVENTY ONE POINT FIVE OH
- 159,106 → ONE HUNDRED FIFTY NINE THOUSAND ONE HUNDRED SIX

  - Years and Comma less Numbers

TEXT PREPROCESSING

- Formatting Operations
  - Numerals

- 715 → SEVEN HUNDRED FIFTEEN
- 71.50 → SEVENTY ONE POINT FIVE OH
- 159,106 → ONE HUNDRED FIFTY NINE THOUSAND ONE HUNDRED SIX

  - Years and Comma less Numbers

- 0159 → OH ONE FIVE NINE
- 1590 → FIFTEEN NINETY
- 7150 → SEVEN ONE FIVE OH
- 1906 → NINETEEN OH SIX
- 1800 → EIGHTEEN HUNDRED

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- In the analysis phase text characters are converted to a narrow phonetic transcription consisting of phonetic symbols and prosodic markers

MORPHOLOGICAL ANALYSIS

- In the analysis phase text characters are converted to a narrow phonetic transcription consisting of phonetic symbols and prosodic markers
- To convert words to phonetic transcript there can be two approaches

MORPHOLOGICAL ANALYSIS

- In the analysis phase text characters are converted to a narrow phonetic transcription consisting of phonetic symbols and prosodic markers
- To convert words to phonetic transcript there can be two approaches
  - Use a complete word dictionary: this is not feasible because of the new words added every day and the size of the lexicon is too large

MORPHOLOGICAL ANALYSIS

- In the analysis phase text characters are converted to a narrow phonetic transcription consisting of phonetic symbols and prosodic markers
- To convert words to phonetic transcript there can be two approaches
  - Use a complete word dictionary: this is not feasible because of the new words added every day and the size of the lexicon is too large
  - Letter to Sound rules that would convert the input letter strings to phonetic segments through some sort of phonetic and transformation process

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- It has been difficult to provide a high degree of accuracy from these rule sets leading to increases in the size of exceptions dictionary

MORPHOLOGICAL ANALYSIS

- It has been difficult to provide a high degree of accuracy from these rule sets leading to increases in the size of exceptions dictionary
- This problem arise in part due to the fact that there is an internal structure in words that must be realized to derive the correct pronunciation

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- The minimum syntactic unit of a language is the morpheme and it has an important role in the determination of pronunciations

MORPHOLOGICAL ANALYSIS

- The minimum syntactic unit of a language is the morpheme and it has an important role in the determination of pronunciations
- When morphemes are represented by the letter strings called morphs they can be effectively used as the basis for determining word pronunciation

MORPHOLOGICAL ANALYSIS

- The minimum syntactic unit of a language is the morpheme and it has an important role in the determination of pronunciations
- When morphemes are represented by the letter strings called morphs they can be effectively used as the basis for determining word pronunciation
- Morphs consists of prefixes, roots and suffixes

MORPHOLOGICAL ANALYSIS

- The minimum syntactic unit of a language is the morpheme and it has an important role in the determination of pronunciations
- When morphemes are represented by the letter strings called morphs they can be effectively used as the basis for determining word pronunciation
- Morphs consists of prefixes, roots and suffixes
- ‘Snowplows’, ‘antidisestablishmentarianism’

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- Morphs are often ingredients of newly coined compound words but new morphs are rarely formed

MORPHOLOGICAL ANALYSIS

- Morphs are often ingredients of newly coined compound words but new morphs are rarely formed
- For this reason they are good candidates for lexical entries, provided a means can be found to analyze words into their morphs

MORPHOLOGICAL ANALYSIS

- Morphs are often ingredients of newly coined compound words but new morphs are rarely formed
- For this reason they are good candidates for lexical entries, provided a means can be found to analyze words into their morphs
- An effective morph lexicon can have less than 10,000 entries

MORPHOLOGICAL ANALYSIS

- Morphs are often ingredients of newly coined compound words but new morphs are rarely formed
- For this reason they are good candidates for lexical entries, provided a means can be found to analyze words into their morphs
- An effective morph lexicon can have less than 10,000 entries
- When morphs are joined together they often change pronunciation depending on the nature of the morphs involved

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- ‘Dogs’ and ‘cats’ : can be described using a morphophonemic rule

MORPHOLOGICAL ANALYSIS

- ‘Dogs’ and ‘cats’ : can be described using a morphophonemic rule
- ‘th’ in ‘this’ and ‘hothouse’ , ‘sch’ in ‘school’ and ‘discharge’ can be pronounced correctly using morpheme decomposition

MORPHOLOGICAL ANALYSIS

- ‘Dogs’ and ‘cats’ : can be described using a morphophonemic rule
- ‘th’ in ‘this’ and ‘hothouse’ , ‘sch’ in ‘school’ and ‘discharge’ can be pronounced correctly using morpheme decomposition
- Morphs leads to an efficient and productive lexicon and provides for important pronunciation effects

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- Morpheme is not a segment of the word and it has no position in a word, it is an abstract unit arising from linguistic distributional analysis

MORPHOLOGICAL ANALYSIS

- Morpheme is not a segment of the word and it has no position in a word, it is an abstract unit arising from linguistic distributional analysis
- ‘Walked’ has two morphs ‘walk’ and ‘ed’

MORPHOLOGICAL ANALYSIS

- Morpheme is not a segment of the word and it has no position in a word, it is an abstract unit arising from linguistic distributional analysis
- ‘Walked’ has two morphs ‘walk’ and ‘ed’
- ‘went’ has two morphemes ‘go’ and ‘PAST’ but direct segmentation to morphs is not possible

MORPHOLOGICAL ANALYSIS

- Morpheme is not a segment of the word and it has no position in a word, it is an abstract unit arising from linguistic distributional analysis
- 'Walked' has two morphs 'walk' and 'ed'
- 'went' has two morphemes 'go' and 'PAST' but direct segmentation to morphs is not possible
- A special morph type STRONG is defined in MITalk to indicate the two underlying morphemes

MORPHOLOGICAL ANALYSIS

MORPHOLOGICAL ANALYSIS

- Input

MORPHOLOGICAL ANALYSIS

- Input
  - The morphological analysis is provided by the DECOMP module in MITalk

MORPHOLOGICAL ANALYSIS

- Input
  - The morphological analysis is provided by the DECOMP module in MITalk
  - The output of the FORMAT is the input to the DECOMP

MORPHOLOGICAL ANALYSIS

- Input
  - The morphological analysis is provided by the DECOMP module in MITalk
  - The output of the FORMAT is the input to the DECOMP
  - Each record in the data stream contains either a word or a punctuation mark

MORPHOLOGICAL ANALYSIS

- Input
  - The morphological analysis is provided by the DECOMP module in MITalk
  - The output of the FORMAT is the input to the DECOMP
  - Each record in the data stream contains either a word or a punctuation mark
  - DECOMP also accesses a compiled binary format morph lexicon

MORPHOLOGICAL ANALYSIS

- Output

MORPHOLOGICAL ANALYSIS

- Output
  - The output data stream consists of decomposed word records

MORPHOLOGICAL ANALYSIS

- Output
  - The output data stream consists of decomposed word records
  - The following information is contained in each record

MORPHOLOGICAL ANALYSIS

- Output
  - The output data stream consists of decomposed word records
  - The following information is contained in each record

1. Word spelling
2. Word part of speech (possibly more than one)
3. For each part of speech, an optional list of part-of-speech *features*
4. The series of morphs obtained by decomposition
5. For each morph, the following information:
   a. Morph spelling
   b. Morph type
   c. One or two homographs
   d. For each homograph, a pronunciation and part(s) of speech