

# Arabic TTS (status & problems)

O. Al Dakkak & N. Ghneim

# Specifications of Arabic

## Generalities

- Arabic is a Semitic language.
- Written Arabic has 28 letters plus “hamza” which has different forms.
- Spoken Arabic has 38 phones. These phones are composed of 28 consonants and 10 vowels.

# Specifications of Arabic

## Generalities

- Some of these vowels (long ones) are written while the short ones are usually omitted. Arabic speakers can easily guess them.
- Some consonants are also omitted from the written words (Shadda, Tanween) Ex: درّس،  
كتابٌ

# Specifications of Arabic

## Morphology:

- Words may be formed of original parts called roots of the verbs, from which one can construct stems using regular forms (subject, object, tool,...) Ex: مكتب، مكتوب، كاتب، كتب , or may be stand-alone nouns Ex: بحر.
- According to the type of the word (verb, noun, preposition,...), it can have several prefixes and suffixes.

# Specifications of Arabic

## Syntax:

- According to the role of the word in the sentence (verb, object, subject, adverb,...) the word either changes its suffixes and/or the vowel at its end. This in turn, play a crucial role on the semantic of the phrase in which the word exists.

# Specifications of Arabic

## Syntax:

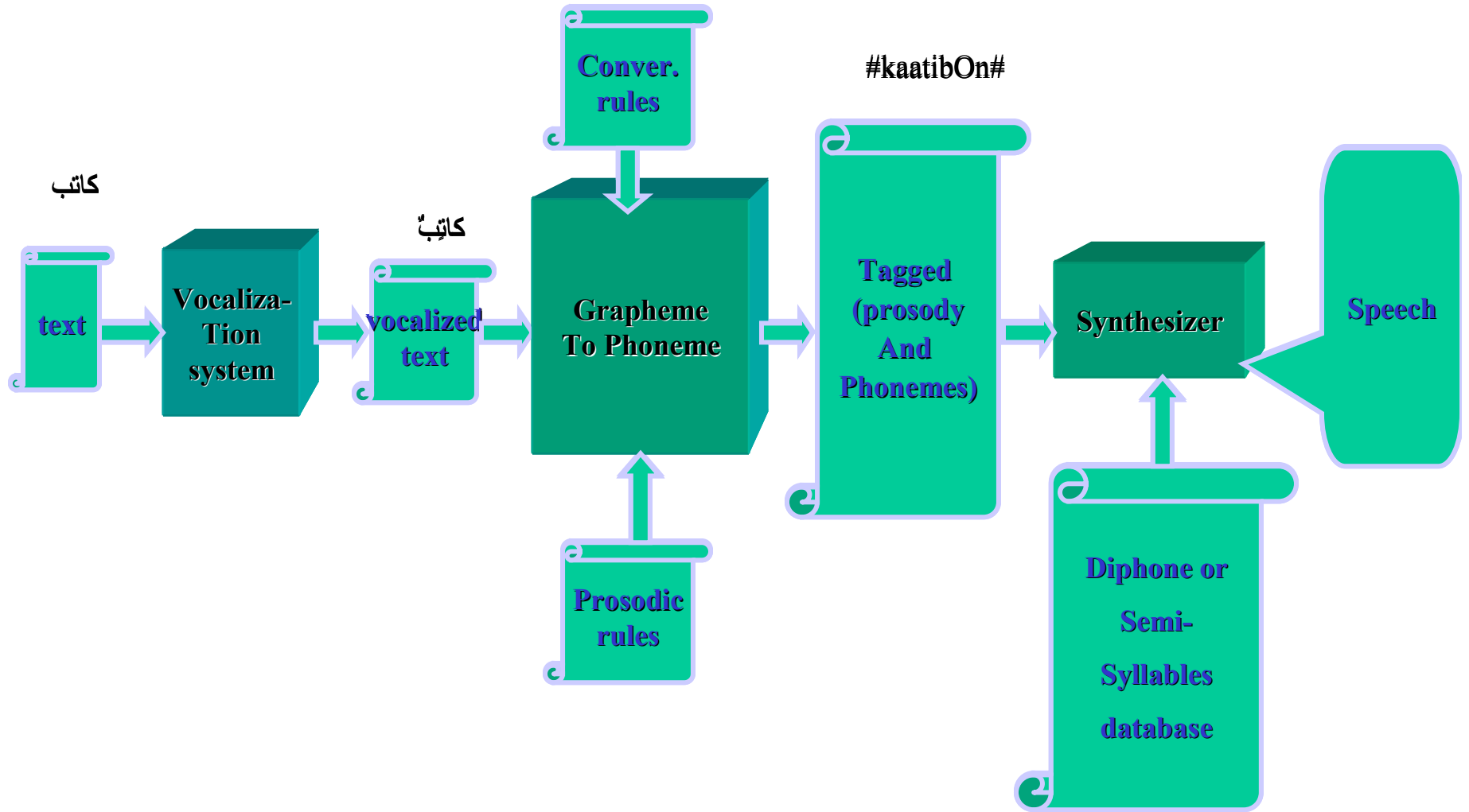
- Sentences can be either verbal (begin by a verb), or nominal (begins by a noun or a preposition).
- A whole phrase can play the role of one word

# Specifications of Arabic

## Semantics:

- As the short vowels are usually omitted; different words with different meanings can have the same written form.
- Sometimes, the same word with the same short vowels can have different meanings according to the context Ex: عين

# Arabic Text-to-Speech System





# HIAST ATTS

- Text preprocessing:
  - If not vocalized Apply vocalization module
  - Apply graphemes-to-phonemes conversion
  - For numbers, we need Part-Of-Speech of the concerned object [gender (m/f), syntactic position (mansub, marfuC or majrou; specific to Arabic), definitive or not, has Tanween or not (specific to Arabic)]

# HIAST ATTS

- Text preprocessing (Vocalization System):  
This system is based on unsupervised machine method composed of four steps:
  - Parsing
  - Morphological Analysis
  - Part of Speech tagging
  - Application of heuristic linguistic rules

For more details see the joint paper “Computational methods to vocalize Arabic Texts” a 1<sup>st</sup> version of the work

# HIAST ATTS

- Prosody Generation
  - (based on the size of each phrase, and the punctuation mark)
  - Generation of F0 contours.
  - Generation of duration for each phoneme.

# HIAST ATTS

- **Waveform Production**
  - based on a diphone database from MBROLA.  
Work in progress for the construction of our own semi-syllable database.
  - The user can choose to listen different voices from the synthesizer (man, woman, child..) and choose the volume of the speech

# HIAST ATTS

- Emotion Inclusion
  - Rules have been extracted and formalized to modify prosody parameters in view of synthesizing different emotions (sadness, joy, anger, surprise, fear).
  - The type of emotion is chosen manually by the user. An automatic choice needs syntactic and semantic analysis, which is not available for the moment.

For more details see joint paper on “Emotion Inclusion in an Arabic Text-to-Speech” presented in EUSIPCO2005

# Points for SSML

- Including tags for the type of speaker and the volume. (already exist)
- Including tags for the type of emotion.
- Incorporation of the vocalization module.