सत्यमेव जयते

Linguistic Data Consortium for
Indian Languages (LDC-IL)

LDC-IL

ROLE OF POS TAGGING IN TEXT TO SPEECH SYNTHESIS

AJU SAMUEL THOMAS
LDCIL, CIIL,
MYSORE
ajuthomas2008@gmail.com
prsamthomas@gmail.com

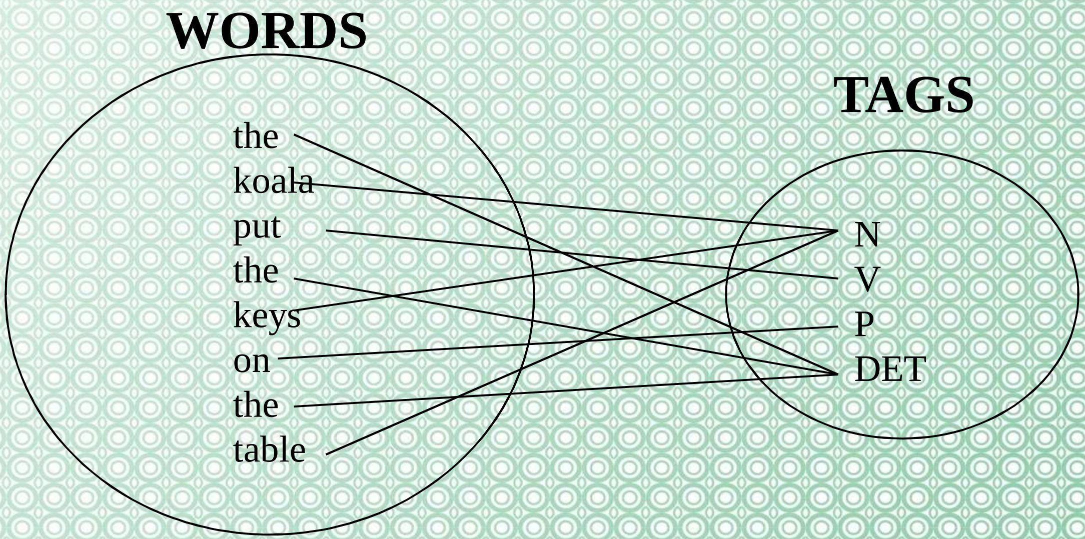
- INTRODUCTION

- POS Tagging is one of the essential parts in the processing of Natural Language.

- It is being used for many applications like developing Machine Translation, Information Retrieval etc.

- POS Tagging: Definition

The process of assigning a part-of-speech or lexical class marker to each word in a corpus



WORDS

the
koala
put
the
keys
on
the
table

TAGS

N
V
P
DET

- As mentioned earlier, POS Tagging is the inevitable part in Natural Language processing.

- It is also applicable in the development of a Text to Speech System.

- How? It will be explained shortly.

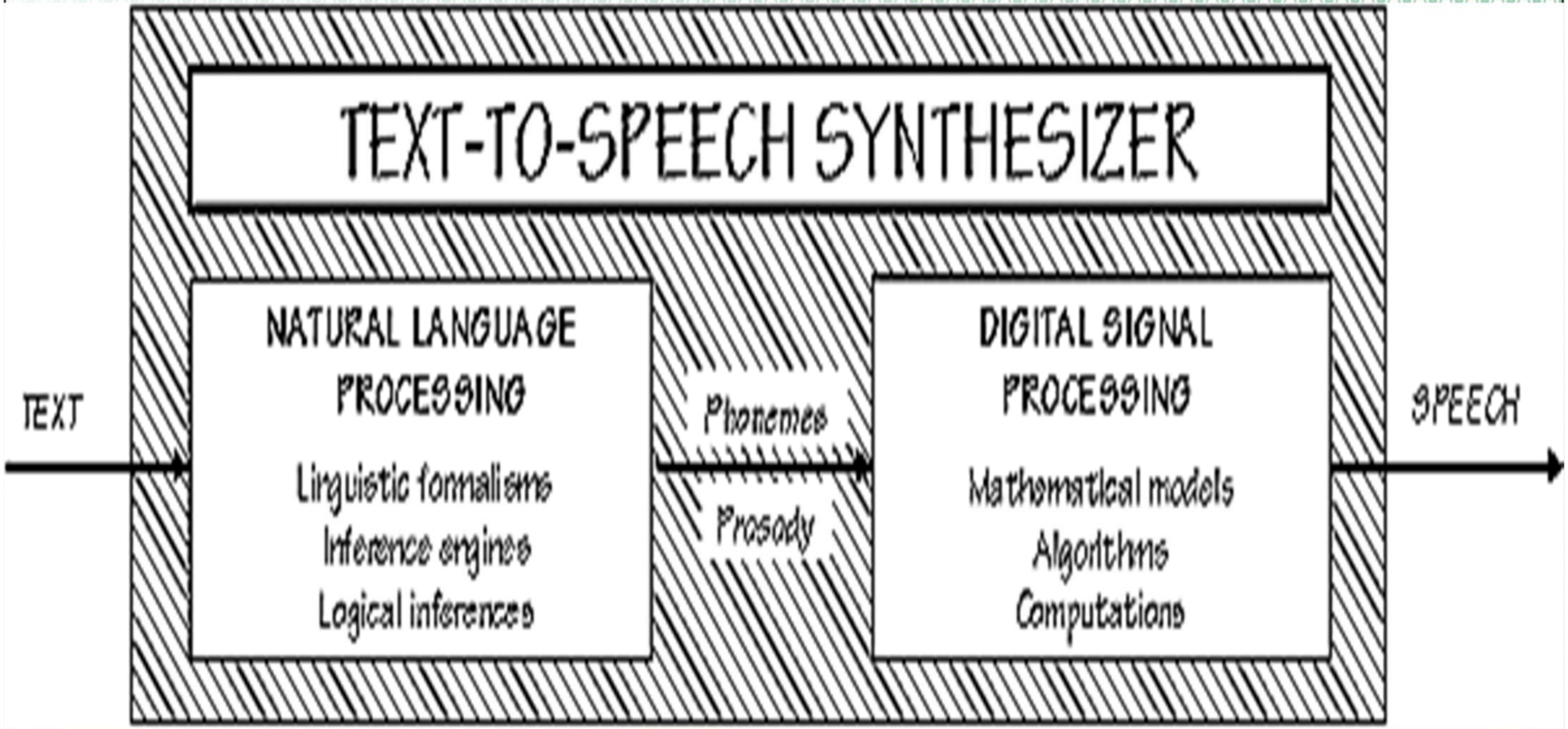
- What is a Text to Speech System ?

Automatic conversion of arbitrary or unrestricted natural language sentences from its text form into its spoken form.

A text-to-speech system must be

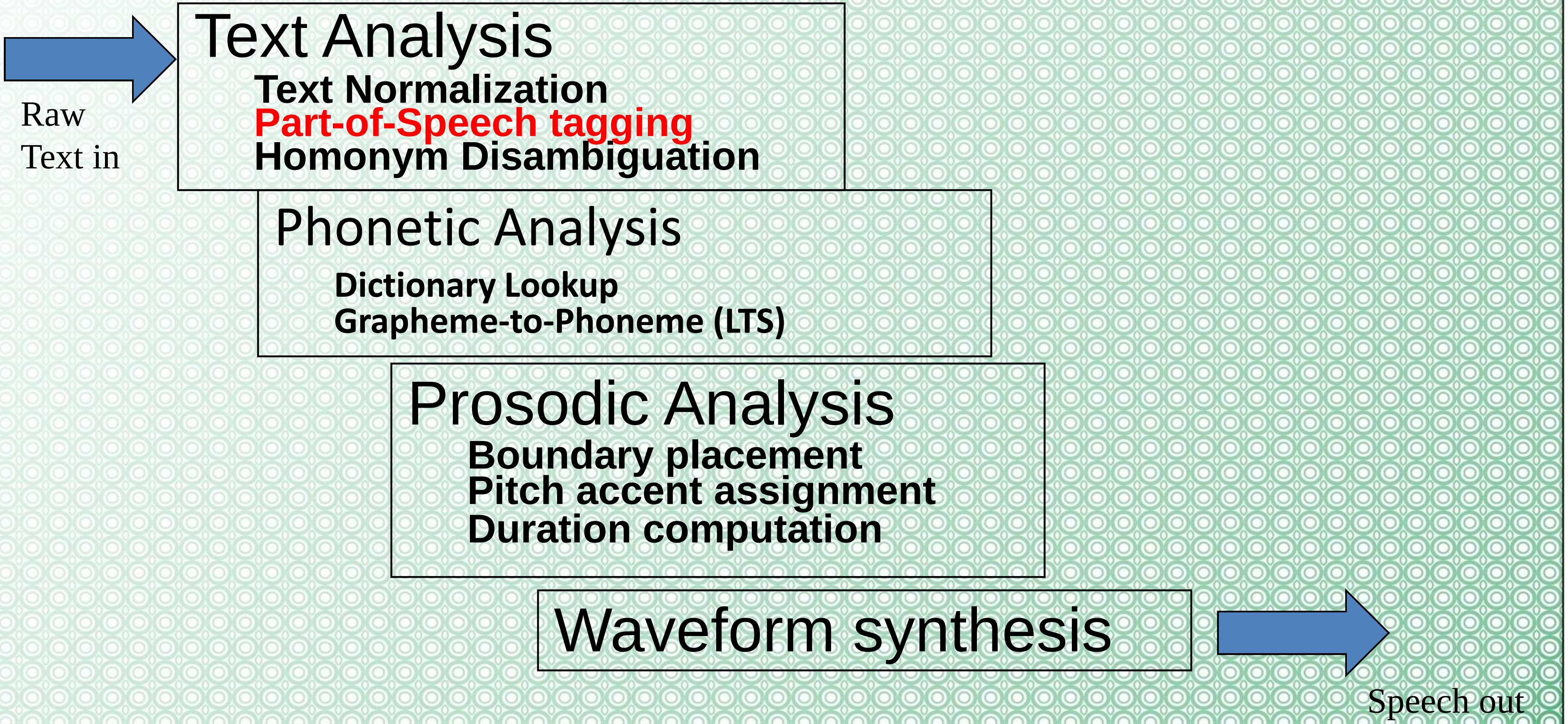
- Able to read **any** text
- Intelligible
- Natural sounding

- How does TTS work?



- **The NLP block**: converts the input text into a sequence of sound units with a set of specifications
- **Speech inventory**: a database of sound units
- **DSP block**: `*appropriate'* sound units from the speech inventory are selected and then concatenated to produce output speech (concatenative synthesis)

- TTS Architecture



Raw
Text in

Text Analysis

Text Normalization
Part-of-Speech tagging
Homonym Disambiguation

Phonetic Analysis

Dictionary Lookup
Grapheme-to-Phoneme (LTS)

Prosodic Analysis

Boundary placement
Pitch accent assignment
Duration computation

Waveform synthesis

Speech out

THE TEXT PROCESSING ASPECT OF SPEECH SYNTHESIS.

Text processing breaks the input of Text to Speech Synthesis into units suitable for further processing. such as

Expanding abbreviations,

Part-of-speech (POS) tagging

Letter-to-sound rules.

- **HOMOGRAPH DISAMBUGATION IN POS TAGGING FOR SPEECH SYTHESIS**
- Every language does have homographs ie: words having same spelling, but having different pronunciations.
- POS tagging can better solve the problem of ambiguity here by giving the right tag according to the context in which a respective word occur.
- Eg : In English, there are some homographs whose pronunciation change according to the category.

REcord	**reCORD**
INsult	**inSULT**

A morphological analysis can determine the part-of-speech (POS) information for many of the words, but some will have multiple possible POS categories.

Without POS information, pronunciation might be ambiguous e.g. “lives”

POS will also be used to predict the prosody.

The lexicon entries have three parts

1. Head word
2. POS
3. Phonemes

The **POS** is sometimes necessary to distinguish homographs.

HEAD	**POS**	**PHONEMES**
LIVES	NOUN	l a i v z
LIVES	VERB	l i v z

- POS Tagging and Phrasing

POS tagging also useful for CONTENT/FUNCTION distinction, which is useful for phrasing which helps the Text to Speech synthesis process a lot.

- **Suprasegmental** aspects of Speech Synthesis.

Prosody versus POS Tagging

Prosody may reflect various features of the speaker or the utterance:

The emotional state of the speaker.

The form of the utterance (statement, question, or command)

The presence of irony or sarcasm; emphasis, etc

- In terms of acoustics, the prosody of languages involve variation in syllable length, loudness, pitch, and the formant frequencies of speech sounds.

- Generally, Orthographic conventions to mark for prosody include punctuation (commas, exclamation marks, question marks, scare quotes and ellipses)

- In a corpus, punctuation, commas, exclamation marks, question marks, scare quotes and ellipses etc occur and it is quite natural.
- Even a quote can generate a meaning sometimes and that quote must be given importance while doing POS tagging.
- It will help the process of text to speech synthesis

- **Stress verses POS Tagging**
- **Stress** is the relative emphasis that may be given to certain syllables in a word,
- Or to certain words in a phrase or sentence.
- The stress placed on syllables within words is called **word stress** or **lexical stress**.
- The stress placed on words within sentences is called **sentence stress**.

- In POS tagging the along with giving tags to each word, it is better to mark stress occurs in each word in a sentence. Sentence stress can also be marked.
- Though it is a tedious and time consuming job, it will cater to the needs of text to speech synthesis.

- Tone versus POS Tagging
- Tone is usually used in language
- The tone here used is not at all related with the tone in the tonal languages.
- In communication, people use tones as part of expressing a concept and its usual.
- Rising Tone and Falling Tone are common in communication.

- While doing POS tagging, it is better to give markers for rising and falling tone in the text.
- It will give naturalness to the speech in the process of converting the respective text into equivalent speech.
- Also the appropriate tone will occur in the output of Synthesis.
- Those who are doing tagging in their respective languages are supposed to know the tone of a given text according to the context since they are well versed in their mother tongue.
- He/She can give the markers of rising and falling tone accordingly.

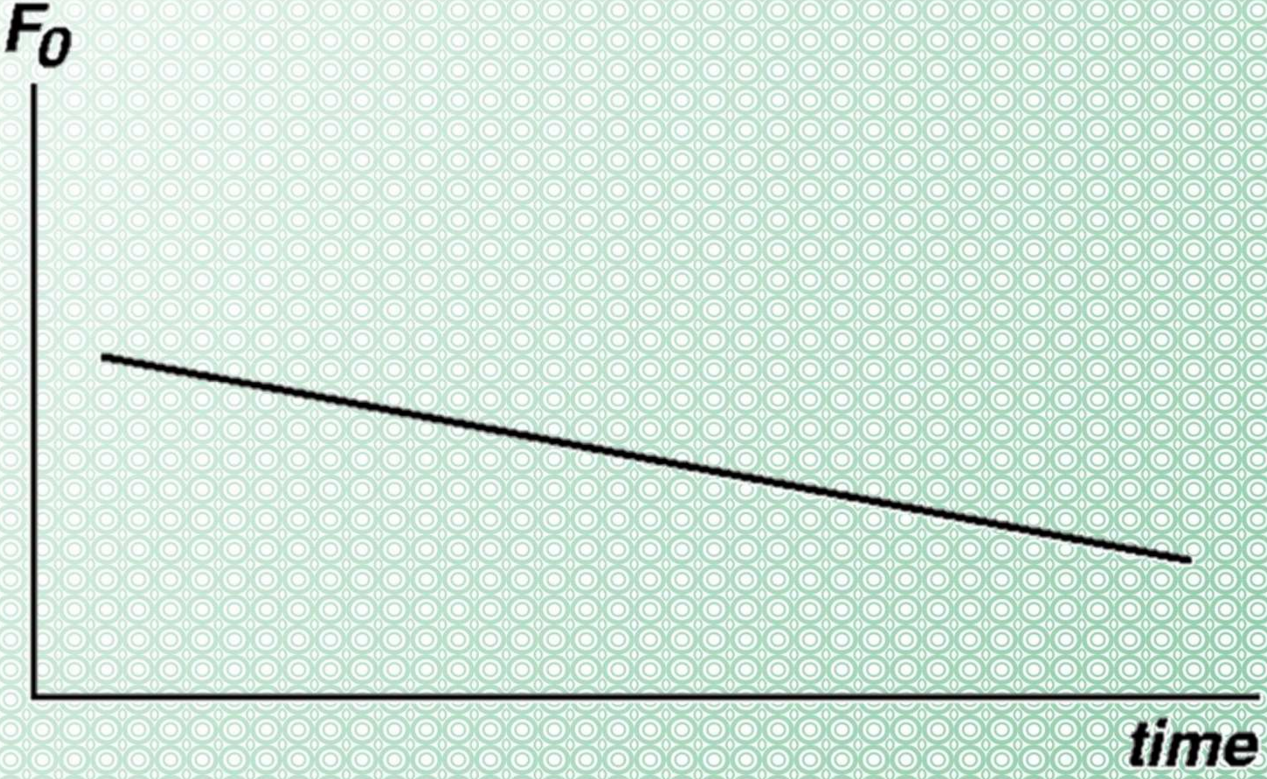
- TONE IN ACOUSTIC FORMAT

The **tone** of an utterance has two components:

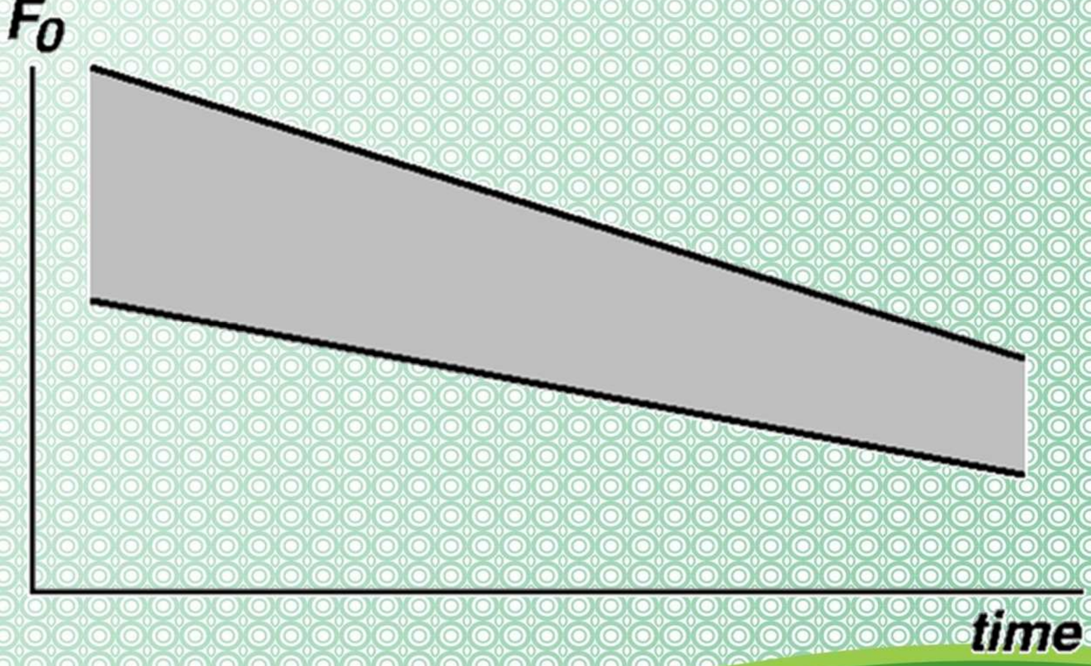
- The global pitch contour shape

- Localised pitch accents

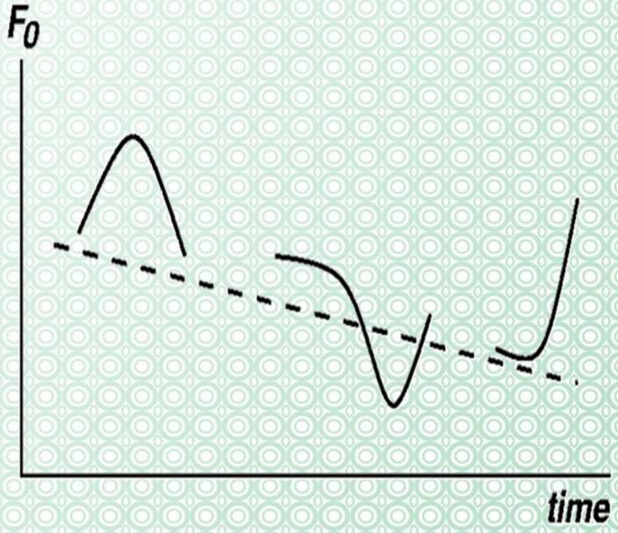
Most utterances show an overall downward trend in f0 called **declination**. We run out of breath, so air flow and pressure decrease and the vocal folds vibrate more slowly.

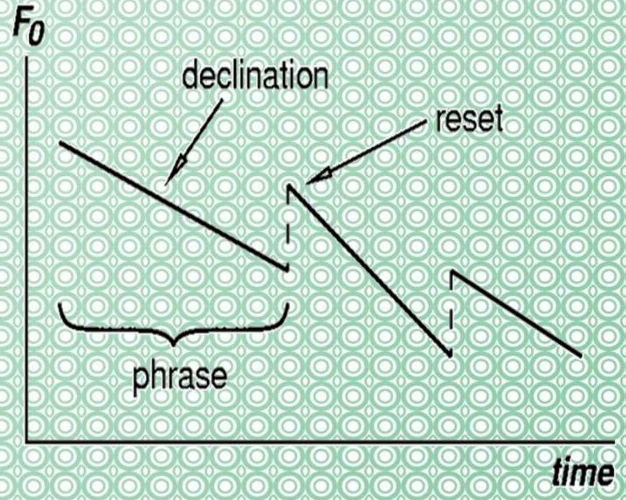

F_0
time

- Tone: Base line and Top line
- Not only does the mean value of f0 decrease with time, the range does too.



- **Tone: Pitch accents**





- So while doing POS tagging, if one gives some markers for rising and falling tone, it will be very helpful in generating naturalness as it is, in the process of converting a text into speech.

THANK YOU

MAY GOD BLESS YOU