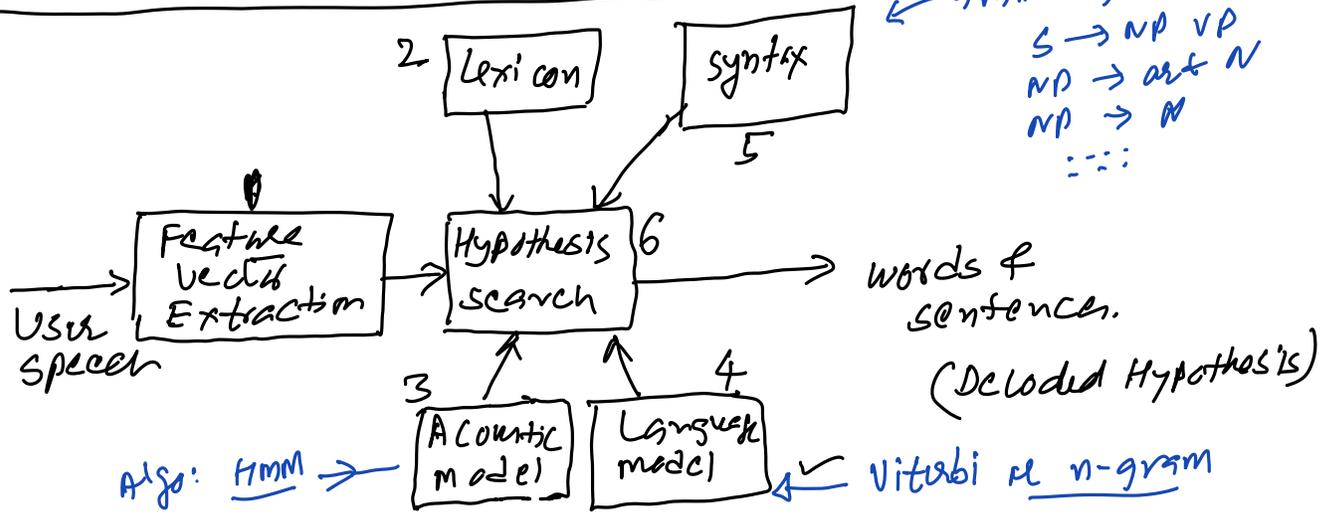


Architecture of Speech Recognition System:



Example of lexicon: book, cook, look, etc. selects the one that has maximum probability for the given feature vector, i.e. $\hat{w} = \max P(w_i | y_1^T)$ \rightarrow i.e. y_1, \dots, y_T
 the w_i for which prob. is maximum, we call that as \hat{w} .

This system works in two modes: ① Training & ② Testing

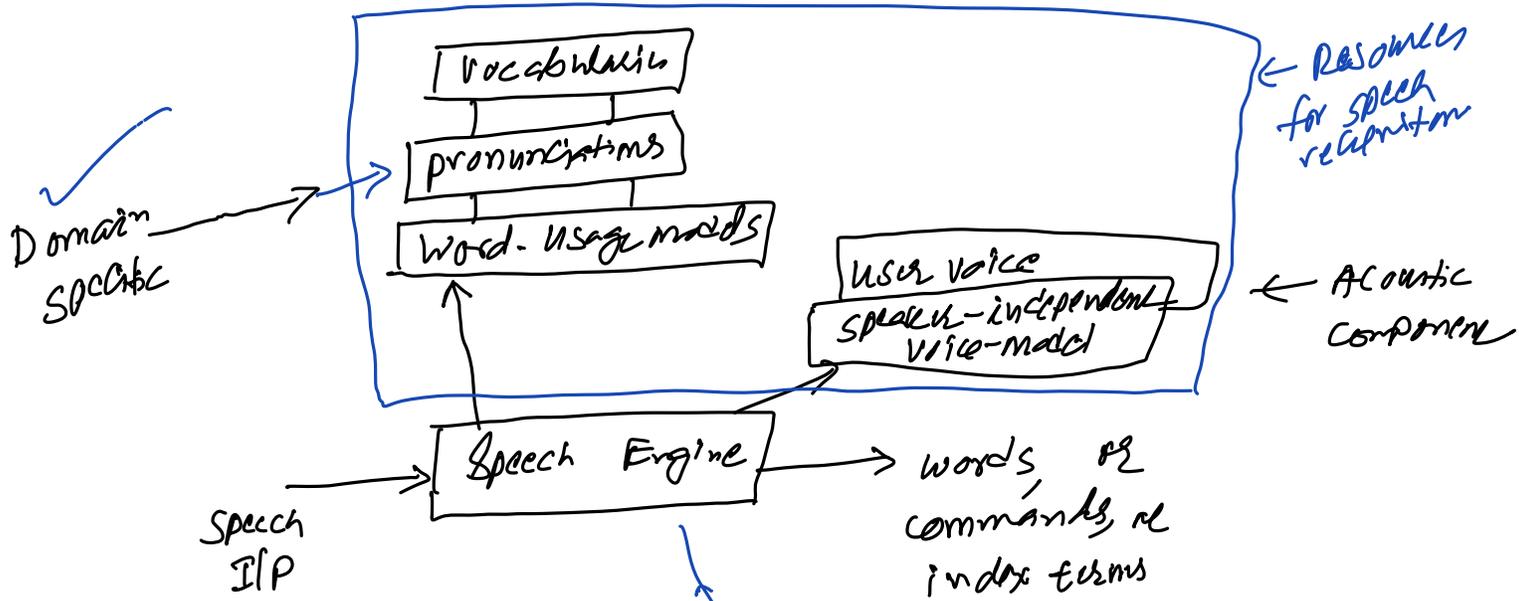
What actually is required in the database?

- resources for speech recognition
- words, and their pronunciations
 - all the words (vocabulary)
 - user voices (models) i.e. pronunciations in the domain
 - word-usage models

using these resources we do the job of speech recognition,
and o/p is produced ✓

↳ it is application dependent

- ✓ → speech-to-text, ∴ o/p is words sequence
- ✓ → operating ATM through user's voice,
∴ o/p is commands, i.e. sequence of steps
- ✓ → please sing song: "mere desh ki dharti"
↑ command ↳ act as index words for songs databases



✓ Represents applications (application oriented)

Comprises all the parts of Architecture which are not shown here.

The Architecture says about working of the system as whole

