Significance of Syntactic Information in TTS

Arulmozhi P, A G Ramakrishnan
IISc, Bangalore.

About TTS

Conversion of sequence of symbols conveying linguistic information to acoustic wave form.

The accuracy of a TTS depends on the naturalness and intelligibility.

MILE TTS : http://mile.ee.iisc.ernet.in:8080/tts_demo/

Two major modules:
NLP (Natural Language Processing) Module
DSP (Digital Signal Processing) Module

NLP Module

- Takes care of the phonetic transcription of the text.
- Takes care of :
  - Text normalization
    - Normalization of non-standard words
      - mi.mI => milli meter
    - Number Expansion
  - Grapheme to Phoneme Conversion
    - Character-Phone mapping
      - அ – a, க - k
    - G2P Rules
      - Ka-ga conversion and similar.
    - Foreign Word lexicon

Why POS tagger

- POS tagger is made as a part of NLP module.
- The natural sentence read by a native speaker is intelligible because,
  - It has appropriate pronunciation
  - Right intonation
  - Right stress
  - Pauses at the right place

- The appropriate pronunciation is obtained by selecting the best unit in the phoneme context.
- The pauses at the right place has to be automatically identified by the usage of the type of word in the context.
- POS tagging is the first step to identify the right place to pause.
- Using POS tags, the phrases and clause boundary can be identified which is also useful in identifying pauses.

Continued..

- In general case, POS is needed for identifying the right pronunciation of certain words.
  - English : Project, Lead
- In Tamil, there is no such pronunciation differences.
- Intonation pattern and pitch contour can also be found out with syntactic information.
- Syntactic phrases must be contained in phonological phrases.
- Taking POS as a base, the phonological phrase can be found out

Nature of the Language : Tamil

- Dravidian Language
- Agglutinative
  - vantucenRAn
- Morphologically rich
  - varukiRAnE
  - vA+kiR+An+E
- Partially free word order
  - Aciriyar nanRAka paTitta mANavanukku paricae kotuttAr
  - nanRAka paTitta mANavanukku Aciriyar paricae kotuttAr
  - Aciriyar paricae nanRAka paTitta mANavanukku kotuttAr

Tuning of Tagset for TTS

A Tagset represents all the parts of speech in a language

Can vary according to the use

For English : Penn Tree bank, C5, C7, CLAWS

For Tamil : AU-KBC, CIIL, IIIT, Amrita, open source tagsets.

Tags for each word in a sentence is decided according to the tagset used.

- Two level of tags
  - Main tags : Identifies the main category of the word
  - Sub Tags : Identifies the category of inflections
- Monadic tags similar to English are not used
- For TTS purpose we do not need very detailed tags like other NLP activities. But only the main tags will not give sufficient information.

Main Tags

No	Representation	Name
1	NN	Noun
2	VB	Verb
3	PR	Pronoun
4	ADJ	Adjective
5	ADV	Adverb
6	AJP	Adjectival Participle
7	VBP	Verbal Participle
8	Q	Quantifier
9	QW	Question Word
10	CNJ	Conjunction
11	PP	Postposition
12	DET	Determiner
13	COMP	Complimentizer
14	EMPH	Emphatic
15	SYM	Symbol

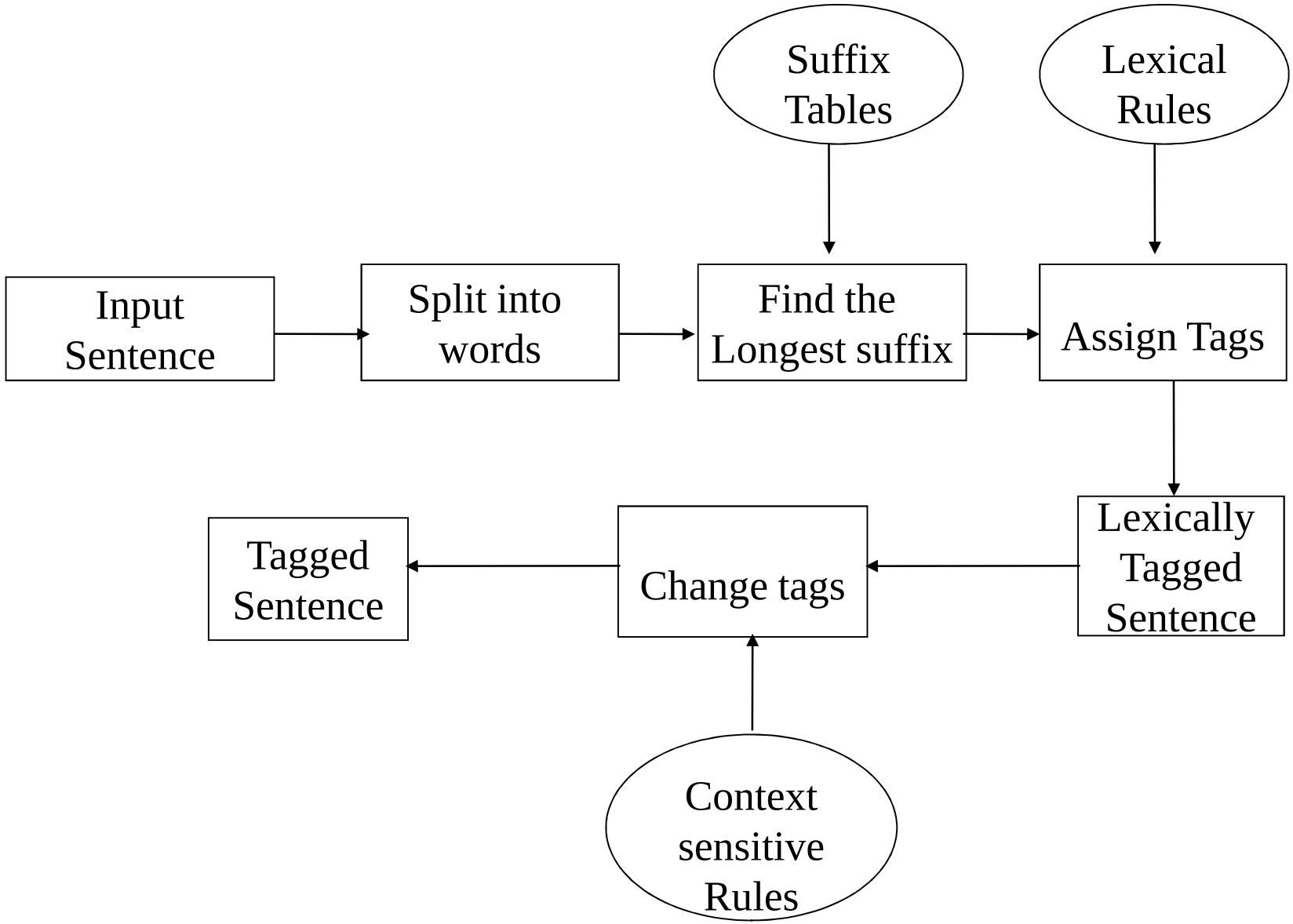
Sample Subtags

1	Acc	Accusative
2	Ins	Instrumental
3	Dat	Dative
4	Soc	Sociative
5	Loc	Locative
6	Gen	Genitive
7	Abl	Ablative
8	Voc	Vocative
9	Ben	Benefactory
10	Sel	Selective
11	Hrt	Hortative
12	Inc	Inclusive
13	Cond	Conditional
14	Cont	Continuous
15	Neg	Negative

Tagger for TTS

- This POS tagger is a rule based one.
- We do not use a root word dictionary.
- The tagger is based on a two stage architecture.
- The first stage has the lexical rules and the second stage has the context sensitive rules.
- The lexical rules act at the word level and the context sensitive rules act at the sentence level.

Architecture of POS Tagger



Lexical Rules

- Separate tables are created for programming purpose with the list of suffixes identified.

- Lexical Rule: Acts in the word level.

  2*1+1*1, NN+pl.acc

  The suffixes indexed 2*1 (Suffix Table 2 Column 1 - kaL) and 1*1 (Suffix Table 1 Column 1 -ae) occur in a sequence, the word will be tagged as Noun+Plural+Accusative.

- There are 13 such tables which list 103 suffixes identified and put in.

Context Sensitive Rules

- Acts in the sentence level
- Example
  - ‘If a sentence starts with a verb, change it to noun’
  - If the first word of a sentence is wrongly tagged as a verb in the first level, it will be corrected in the second level.

- Sentence Splitter:
  - Embedded in the POS tagger
  - Splits Paragraphs into sentences
  - Uses heuristic rules.

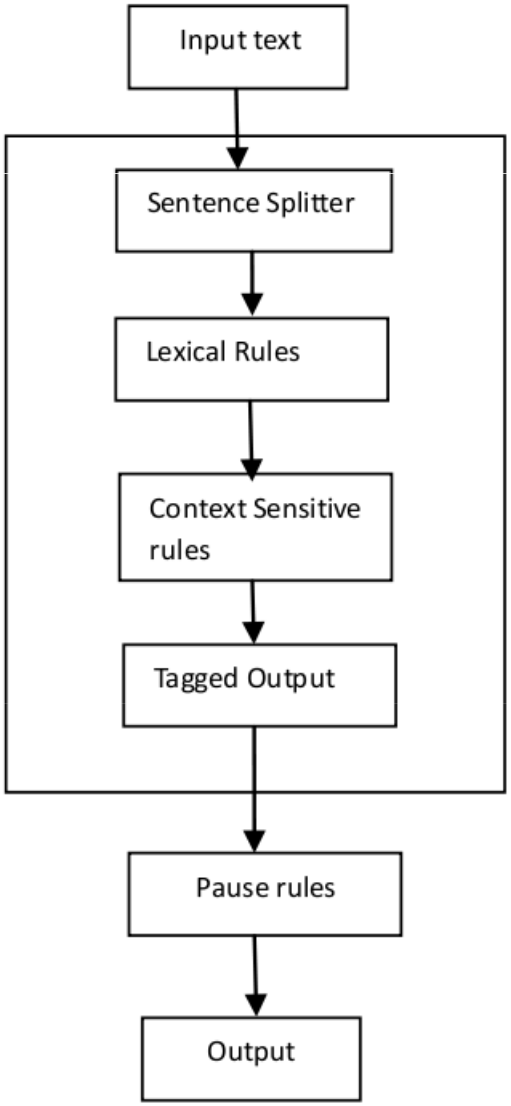
Pause Model

- Native speakers introduce pauses with the knowledge of the language acquired.
- TTS system, those pauses need to be inserted by the system at the right places.
- A wrong pause inserted between two words may make the synthesized speech unnatural.
  - The< np> book<np>is<p> on<np>the<np>table.
  - The<np>book<np>is<np> on<p>the<np>table.
- Deterministic – Rule based
- Uses list of words and POS tags for identifying pause

Continued...

- Six levels of pause have been identified, which determine the duration of the pause.
- <P0>, <P1>, <P2>, <P3>, <P4>, and <PW>.
- Order:
  - <P0> - No Pause
  - <P1> - Lowest Pause
  - <PW> - Common Pause between each word
  - <P2> - Medium Pause (Pause after a comma)
  - <P3> - Significant Pause (Pause after a Semicolon)
  - <P4> - Highest Pause (Pause Between Sentences)
- Wherever <P0> ... <P4> is not identified, <PW> is assumed

Pause Model : System Architecture



Rules for Pause

1. There is no pause (or may be very minimal pause) between a number and the words such as ‘mani’, ‘latcam’, ‘kOTi’. There is a list of words defined for this rule. Any NN (+PL) after a number does not have a pause.

Exception:

plus 1 – there should be no pause between the words.

<month name> <No> - there should be no pause between the words.

<No> <month name> - there should be a small pause between the words.

rUpaay 100 - there should be a small pause between the words.

Continued...

2. If the previous word has an accusative/dative marker, and the current word is a postposition, there is no pause between the current and the previous words.

Ex : avanai <P0> pola,

avanukku <P0> pin

3. If the previous word ends in a consonant (k, c, t, p) and the current word starts with the same letter, there is no pause between them.

Ex : akkaraec <P0> cImae

Continued...

4. If the previous word has the POS “ADV” and the current word has the POS “VB”, then there is no pause in between them.

Ex : itu oru ankamaaka tikazkiRatu.

puttakangkaL inkee uLLana.

5. There should be a pause before quantifier. The POS tag for a quantifier will be Q.

Ex : (azhakiya kiraamamum)<P3>(oru periya ooraaTciyum)

All the numbers are considered as quantifiers.

Exception: 3 aayiram – there is no pause before aayiram.

6. There should be a pause before and after the following words.

maRRum, allatu, aanaal, aakavee, enavee

Sample Output

சி.பி.எஸ்.இ.	முறையில்	படிக்கும்	மாணவர்களுக்கு	அடுத்த
C.B.S.E	stream+loc	studying	students+pl+dat	next
NN	NN+loc	VB+fut+3sn	NN+pl+dat	AJP
<P0>	<PW>	<PW>	<PW>	<P0>

கல்வி	ஆண்டு	முதல்	10ம்	வகுப்பு	தேர்வு	கிடையாது
education	year	from	10th	class	exam	No
NN	NN	PP	Q	NN	NN	VB+pst+3sn+neg
<P0>	<P0>	<P2>	<P0>	<P0>	<PW>	<P4>

சுவற்றின் மேல் சித்திரங்கள் உள்ளன.

After Pause Rules...

- The DSP module will produce speech wave form with appropriate pauses as given by the NLP module.

- There are 15 rules for pause.
- Various levels of pause are also identified.
- There are more rules for <P0>

- Changing all <PW> to <P0> reduces the number of rules.

Conclusion

- Evaluation done based on Mean Opinion Score (MOS)
- The synthesized speech with and without pause model are compared.
- Pause model which uses a POS tagger is found to improve the naturalness of the synthesized speech.
- Problems occur when wrong tags are identified with the POS tagger.
  - Ex : Aru – river or six?
  - Aru – six has a pause before and river does not have a pause before.
    - Ex: pawAni ARu kAvEriyin kiLai Akum
    - pawAni Aru mAttiraikaL cAppiTTAL.
- More rules are to be identified to insert pauses.

Thankyou