How to Speak a Language without Knowing It

Xing Shi, Kevin Knight [1] Heng Ji [2]

[1]Information Sciences Institute
Computer Science Department
University of Southern California
{xingshi, knight}@isi.edu

[2]Computer Science Department
Rensselaer Polytechnic Institute
Troy, NY 12180, USA
jih@rpi.edu

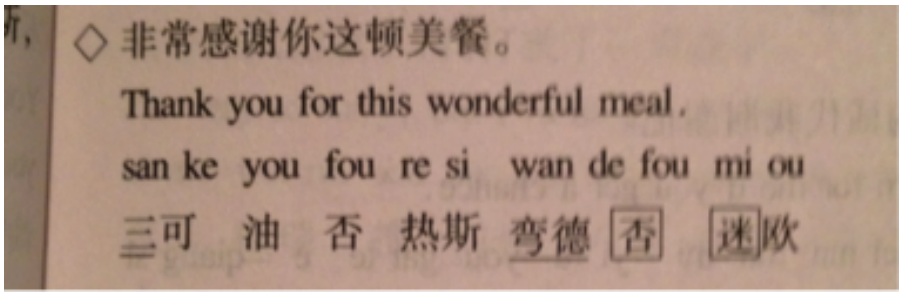
June 24, 2014

Overview

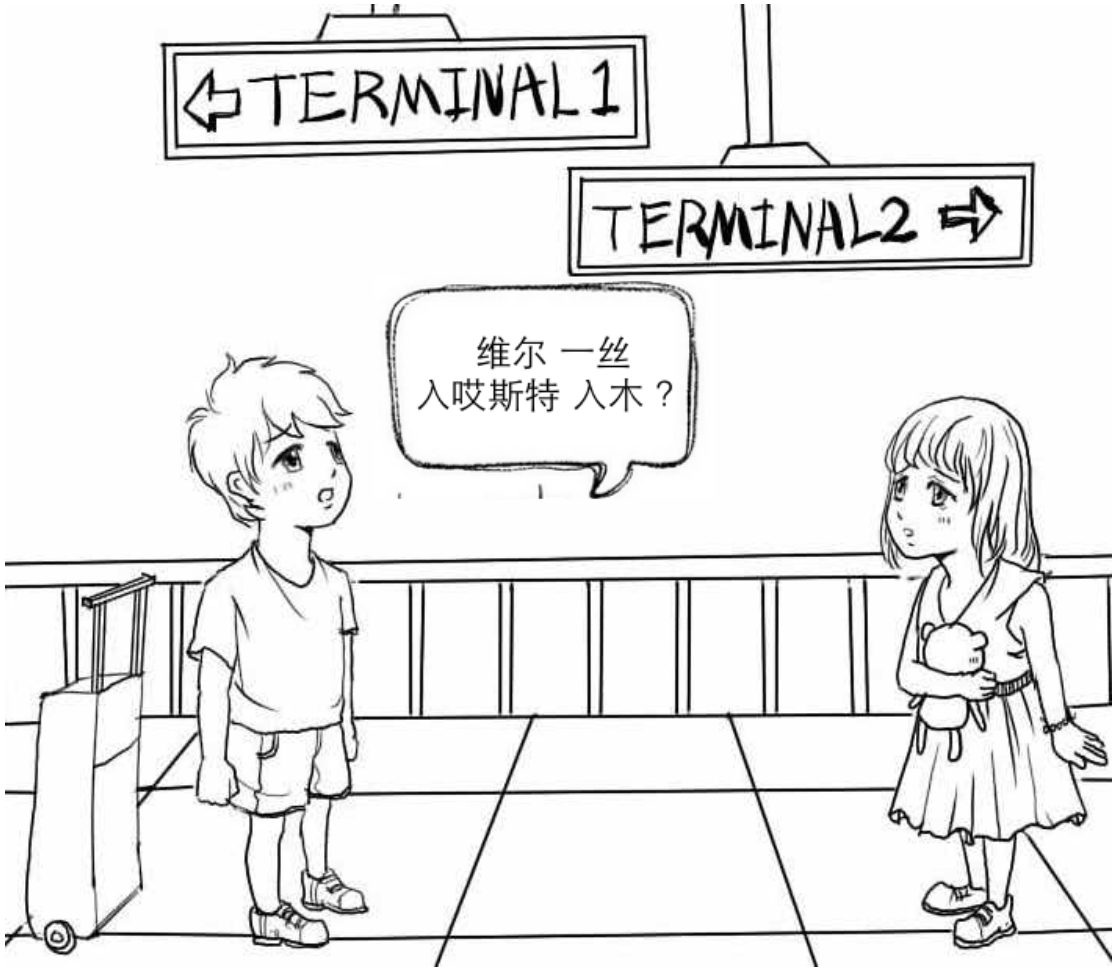
Introduction

- Can people speak a language they don't know ?

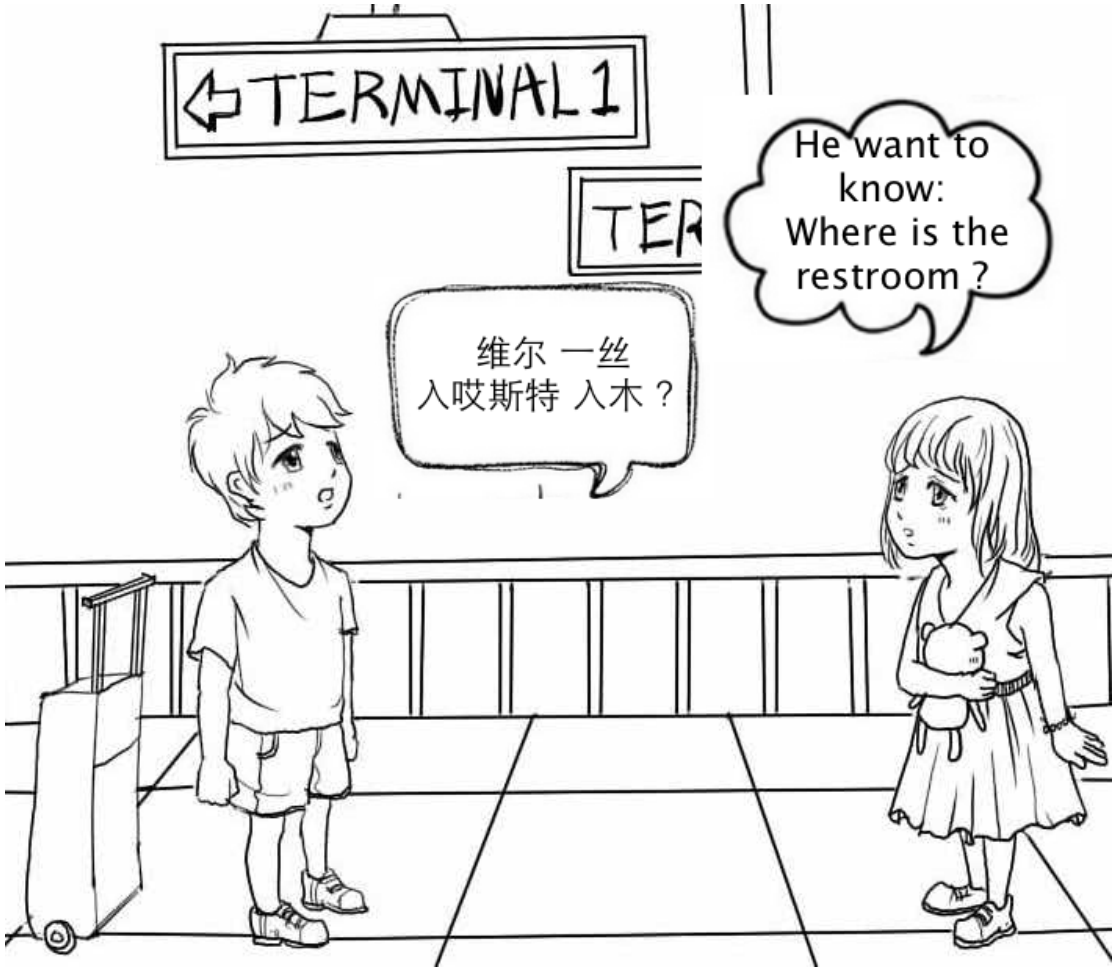
Yes, use a phrasebook



Yes, use a phrasebook



Yes, use a phrasebook



Yes, use a phrasebook

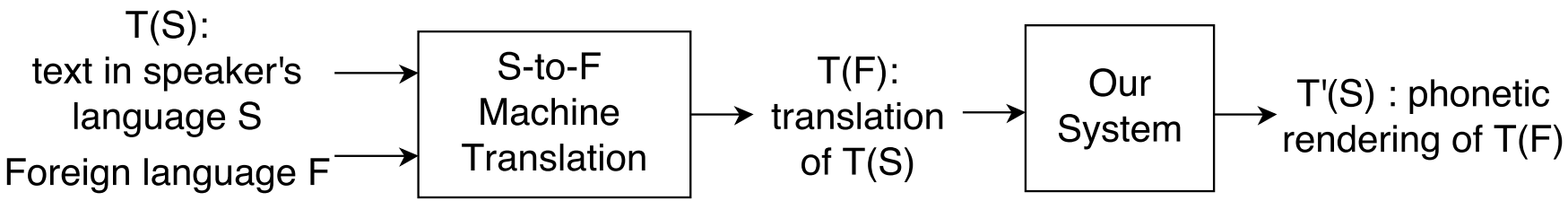
What if
we want to say something beyond the phrasebook ?

Or, a speech-to-speech translator

from: *proto-knowledge.blogspot.com*

However, direct Human interactivity is much more fun !

Our solution

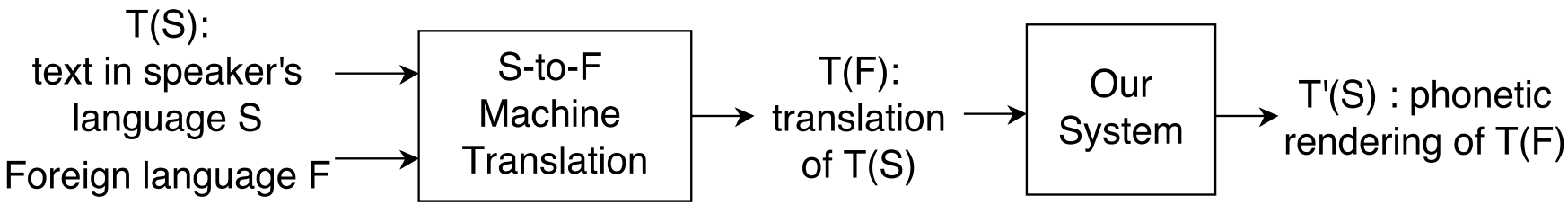


- Easily pronounceable
  - Both input T(S) and output T'(S) are in speaker's language.
- Understandable by listener
  - T'(S) sounds like T(F).
  - T(F) and T(S) has the same meaning.

Our solution

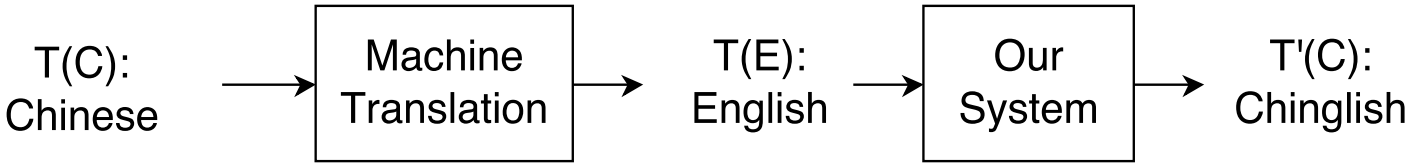
Demo

Our solution



谢谢你 **Thank you** **三可 有**

Data



- A collection of 1312 <Chinese, English, Chinglish> phrasebook tuples. [1]
- 1182 for training, 65 for development and 65 for test.

Chinese	已经八点了
English	It's eight o'clock now
Chinglish	意思埃特额克劳克闹 (yi si ai te e ke lao ke nao)
Chinese	这件衬衫又时髦又便宜
English	this shirt is very stylish and not very expensive
Chinglish	迪思舍特意思危锐思掉利失安的闹特危锐伊克思班西五

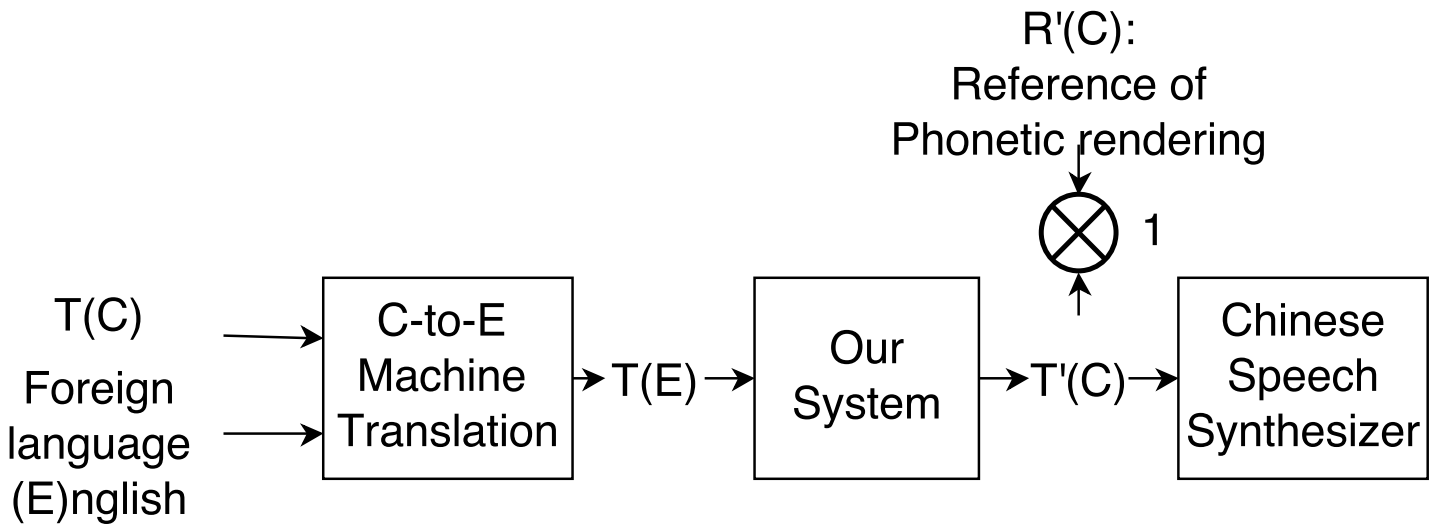
[1] Dataset at `http://www.isi.edu/natural-language/mt/chinglish-data.txt`

Data

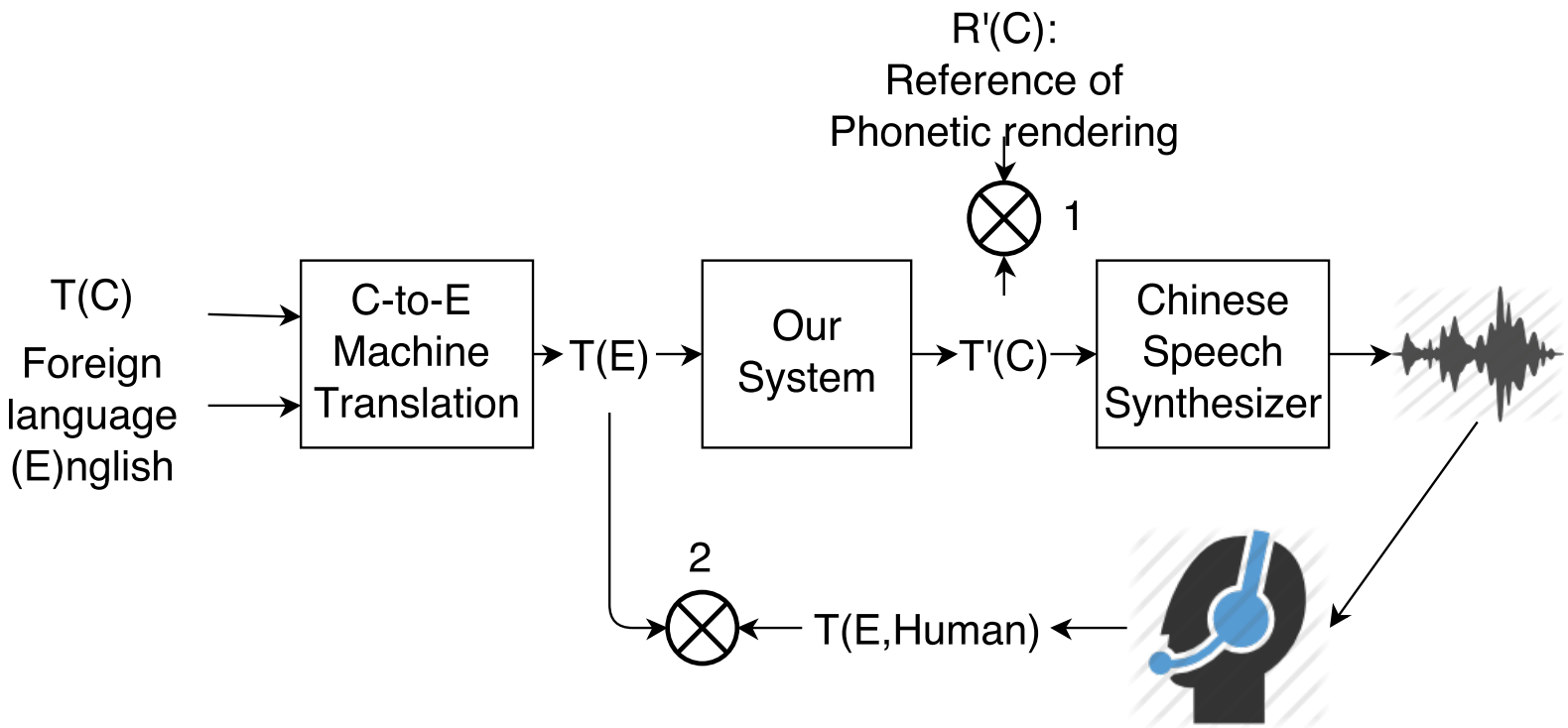
Frequency Rank	Chinese	Chinglish
1	de	si
2	shi	te
3	yi	de
4	ji	yi
5	zhi	fu

Table : Top 5 frequent syllables in Chinese and Chinglish

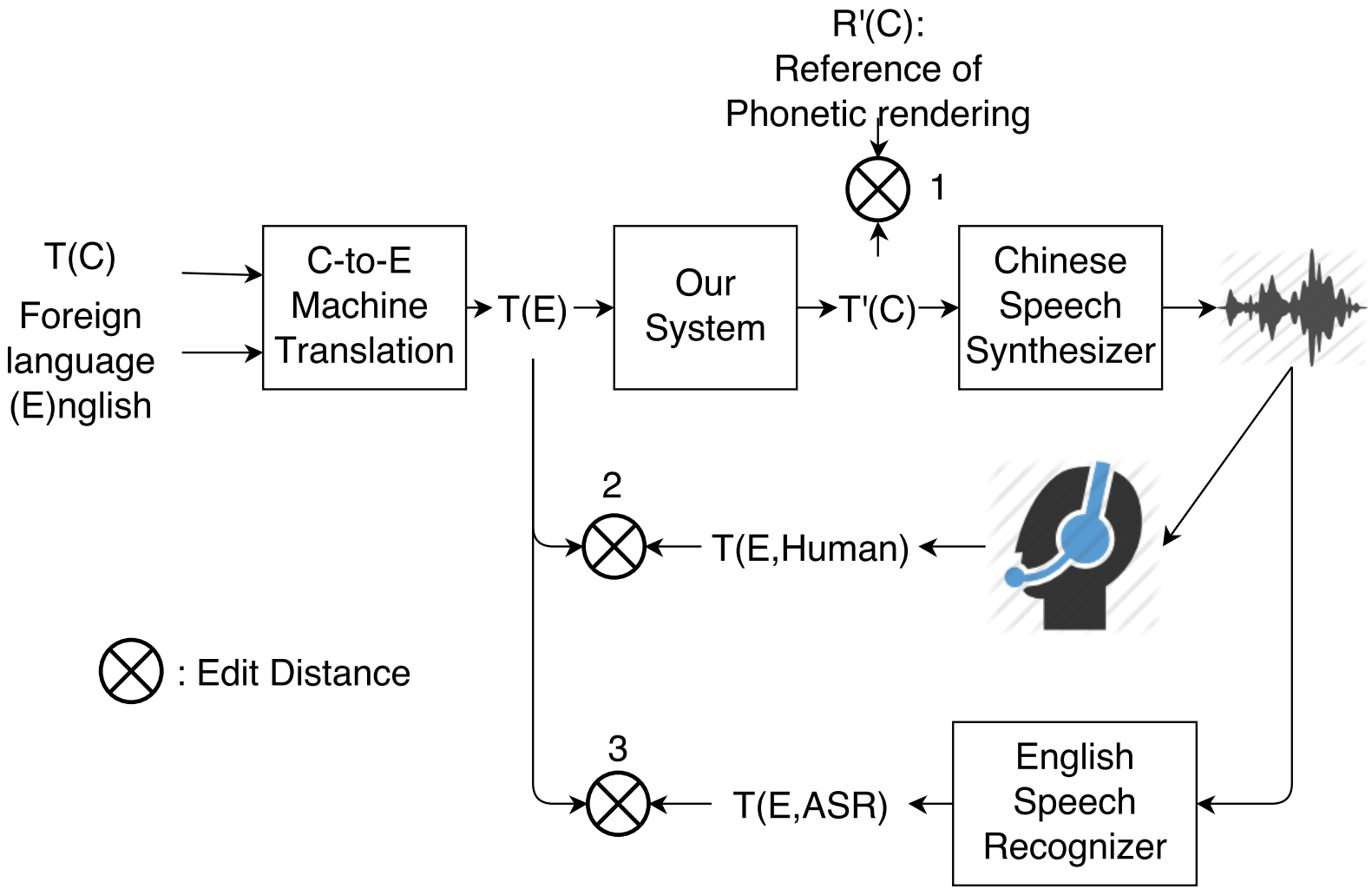
Evaluation


R'(C):
Reference of
Phonetic rendering
1
T(C)
Foreign
language
(E)nglish
C-to-E
Machine
Translation
T(E)
Our
System
T'(C)
Chinese
Speech
Synthesizer

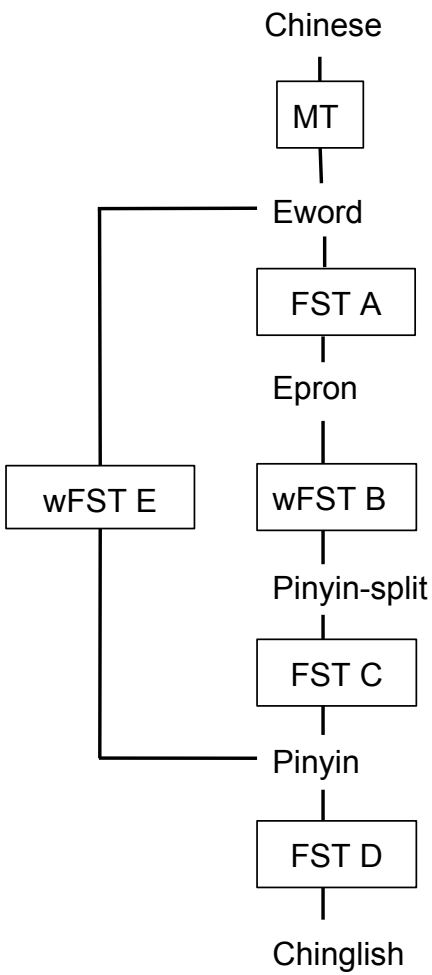
Evaluation


R'(C):
Reference of
Phonetic rendering
1
T(C)
Foreign
language
(E)nglish
C-to-E
Machine
Translation
T(E)
Our
System
T'(C)
Chinese
Speech
Synthesizer
2
T(E,Human)

Evaluation


R'(C):
Reference of
Phonetic rendering
1
T(C)
Foreign
language
(E)nglish
C-to-E
Machine
Translation
T(E)
Our
System
T'(C)
Chinese
Speech
Synthesizer
2
T(E,Human)
: Edit Distance
3
T(E,ASR)
English
Speech
Recognizer

Model: Cascade FSTs



谢谢你

translate.google.com

Thank you

CMU Pron Dict
(Weide,2007)

TH EY N K Y UW

s an k e y ou

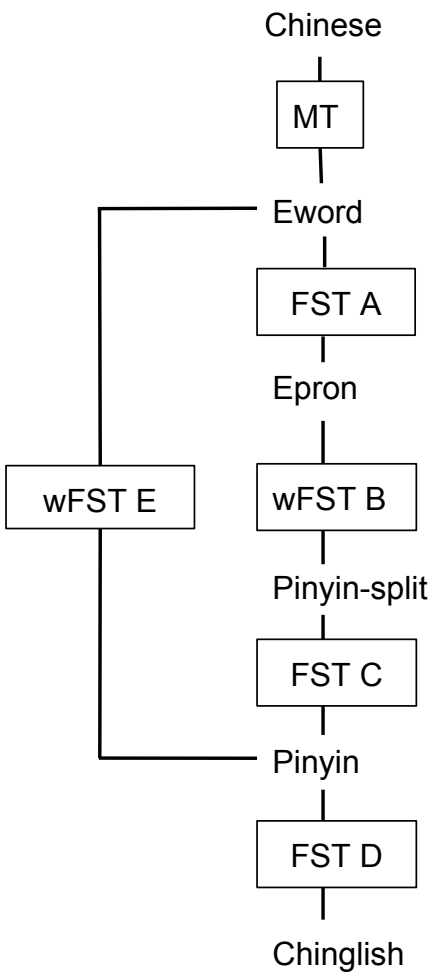
Deterministic Rules

san ke you

Pron Dict

三可由

Model: Cascade FSTs



谢谢你

translate.google.com

Thank you

CMU Pron Dict
(Weide,2007)

TH EY N K Y UW

Need to learn from data

s an k e y ou

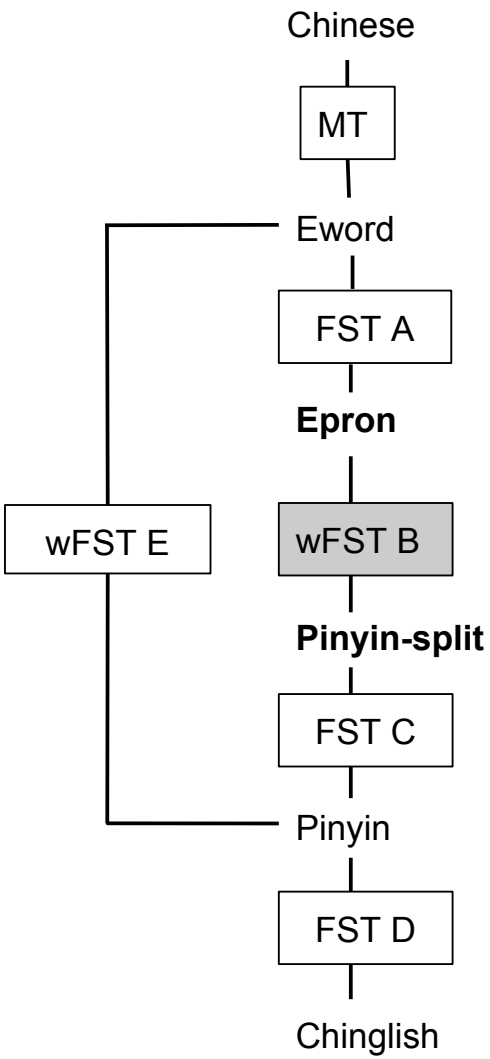
Deterministic Rules

san ke you

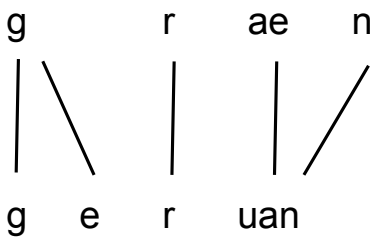
Pron Dict

三可由

Phoneme-based model



- Construct <Epron, Pinyin-split> training pairs.
- Mapping schema: 1-to-1, 1-to-2 and 2-to-1.



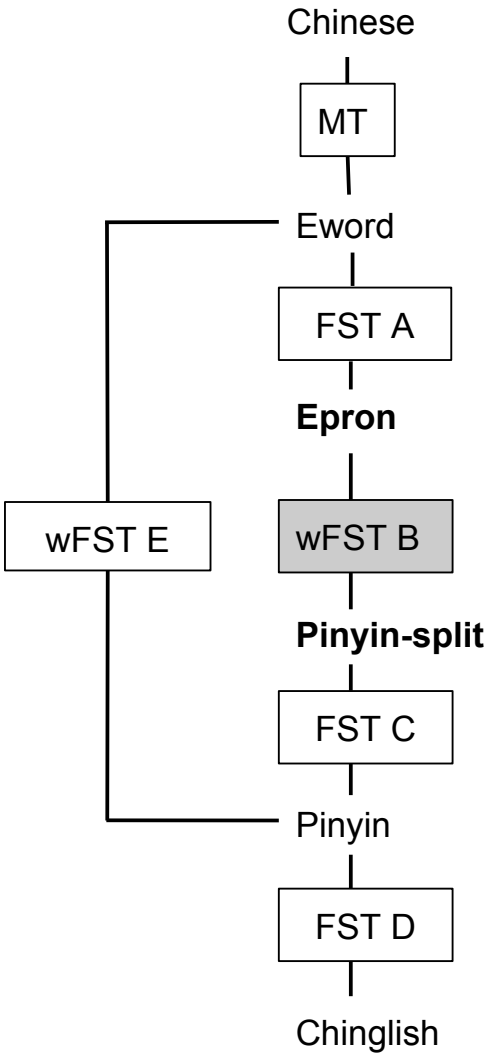
- EM to learn parameters in wFST B, e.g. $P(\text{g e}|g)$.
- Viterbi alignments:

Phoneme-based model

labeled Epron	Pinyin-split	$P(p\|e)$
d	d	0.46
	d e	0.40
	d i	0.06
	s	0.01
ao r	u	0.26
	o	0.13
	ao	0.06
	ou	0.01

Table : Learned translation tables for the phoneme based model

Phoneme-based model



- Alignment using phoneme-based model is fine.
- When decoding test data, choices of target phonemes are context sensitive.

Decoding "grandmother":

g	r	ae n	d	m	ah	dh	er
g e	r	an	d e	m u	e	d	e

reference Pinyin-split sequence:

g e r uan d e m a d e

Phoneme-phrase-based model

- Intuition: model the substitution of longer sequences [2].

  Viterbi alignment using Phoneme-based model:

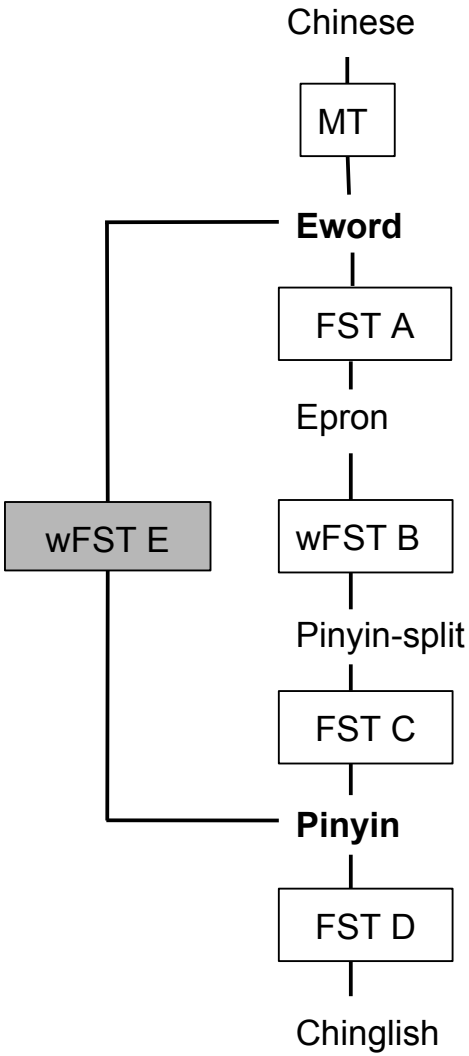
g	r	ae n	d	m	ah	dh	er
g e	r	uan	d e	m	a	d	e

  Extract phoneme phrase pairs:

$$\texttt{g} \rightarrow \texttt{g e}$$
$$\texttt{g r} \rightarrow \texttt{g e r}$$
$$\ldots$$
$$\texttt{r} \rightarrow \texttt{r}$$
$$\texttt{r ae n} \rightarrow \texttt{r uan}$$
$$\ldots$$

[2] (Koehn et al., 2003)

Word-based Model



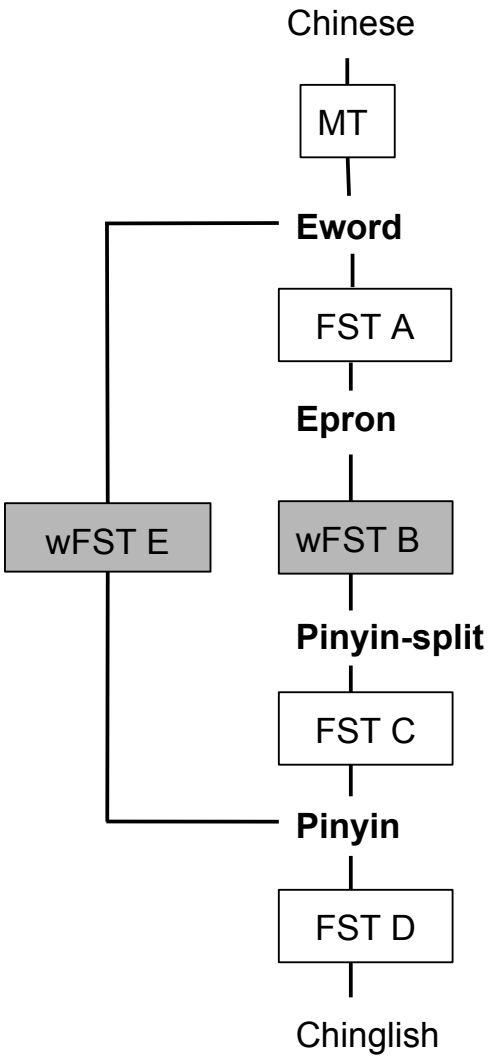
- Construct <Eword,Pinyin> training pairs.
- Mapping schema: 1-to-[1,7].
- EM to learn parameters in wFST E, i.e. $P(\text{nai te}|\textit{night})$.
- Viterbi alignments:

wake	up
wei ke	a pu

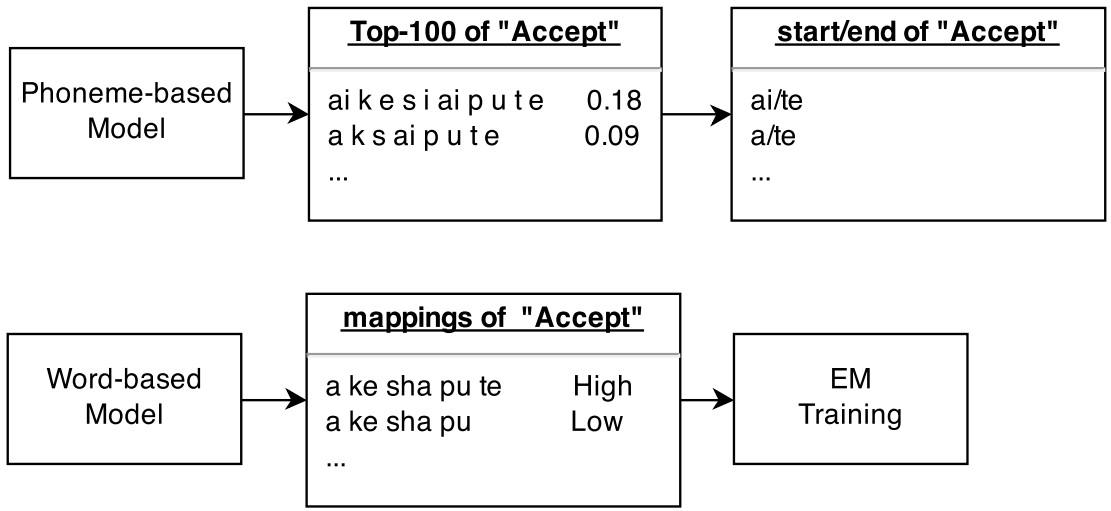
- Error happen due to sparsity: "tips" and "ti pu si" only appear once.

accept	tips
a ke sha pu	te ti pu si

Hybrid training



- Intuition: Combine two models during **training** phrase.
- Use phoneme-based model to help word-based model:

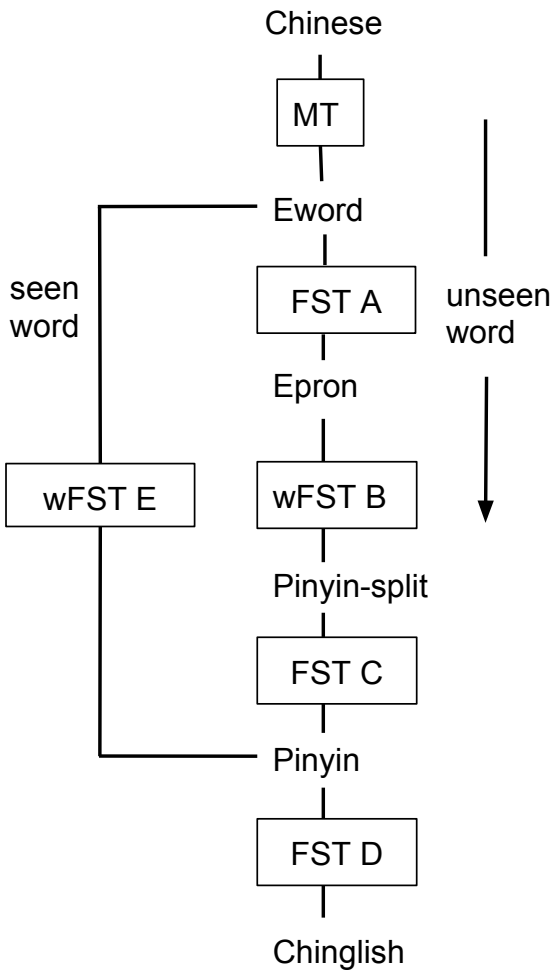


- Errors are fixed:

accept	tips
a ke sha pu te	ti pu si

Hybrid decoding

- Intuition: Combine two models during **decoding** phrase.

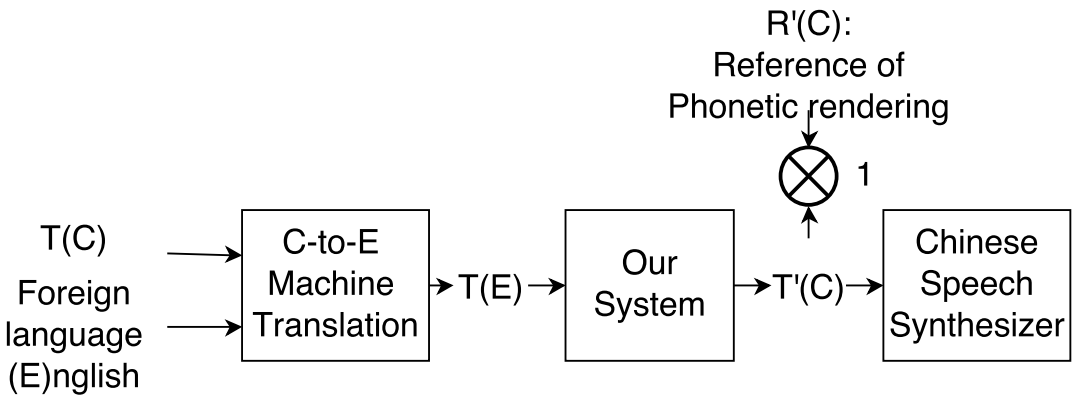


Experiments: Sample system output

Chinese	等等我
Reference English	wait for me
Reference Chinglish	唯特 佛 密 (wei te fo mi)
Hybrid Chinglish	位忒 佛 密 (wei te fo mi)
Human-dictated English	wait for me
ASR English	wait for me
Chinese	年夜饭都要吃些什么
Reference English	what do you have for the Reunion dinner
Reference Chinglish	沃特 杜 又 海夫 佛 则 锐又尼恩 低呢
Hybrid Chinglish	我忒 度 优 嗨佛 佛 得 瑞优你恩 低呢
Human-dictated English	what do you have for the reunion dinner
ASR English	what do you high for 43 Union Cena

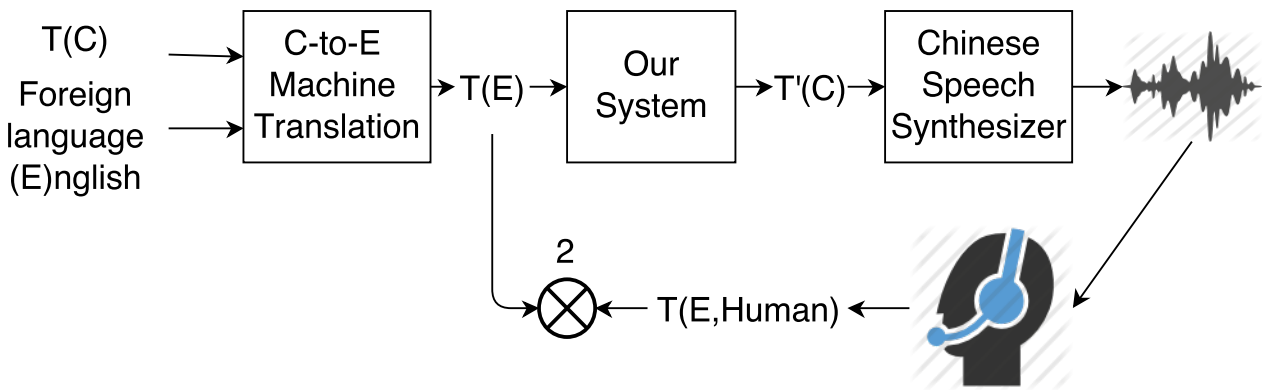
Experiments: English-to-Pinyin decoding accuracy

Model	Coverage	Error Rate on covered text	Error Rate
Word based	29/65	0.042	0.664
Word-based hybrid training	29/65	0.029	0.659
Phoneme based	63/65	0.583	0.611
Phoneme-phrase based	63/65	0.136	0.194
Hybrid training/decoding	63/65	0.115	0.175



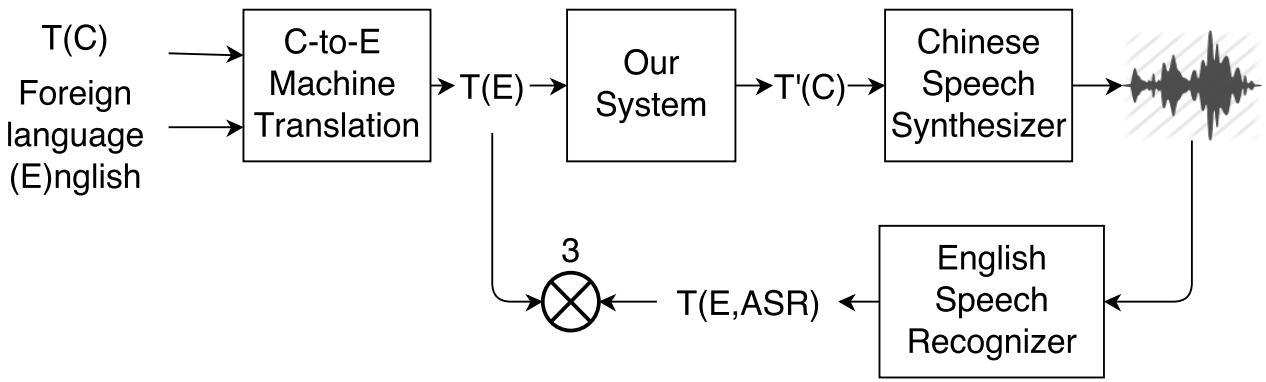
Experiments: Human Dictation Accuracy

Model	Error Rate vs. reference English
Dictation from Reference Chinglish	0.477
Phoneme based	0.696
Hybrid training and decoding	0.496



Experiments: No Human in the Loop

Model	Error Rate vs. reference English
Word based	0.925
Word-based hybrid training	0.925
Phoneme based	0.937
Phoneme-phrase based	0.896
Hybrid training and decoding	0.898



Conclusion & Future work

Conclusion

- Goal: Help people speak foreign languages
  - Provide native phonetic spellings that approximate the sounds of foreign phrases
  - Use a cascade of FSTs
  - Improve the model by adding phrases and combining models in both training and decoding phase

For future:

- More Language Pairs

Thank you! & QA

Demo: `http:\\cage.isi.edu:8080`