Stochastic Finite-State models for Spoken Language Machine Translation

Srinivas Bangalore and Giuseppe Riccardi

{srini,dsp3}@research.att.com

Speech and Image Processing Services Lab
AT&T Labs – Research

WORKSHOP ON EMBEDDED MT SYSTEMS, MAY 4, 2000

Overview

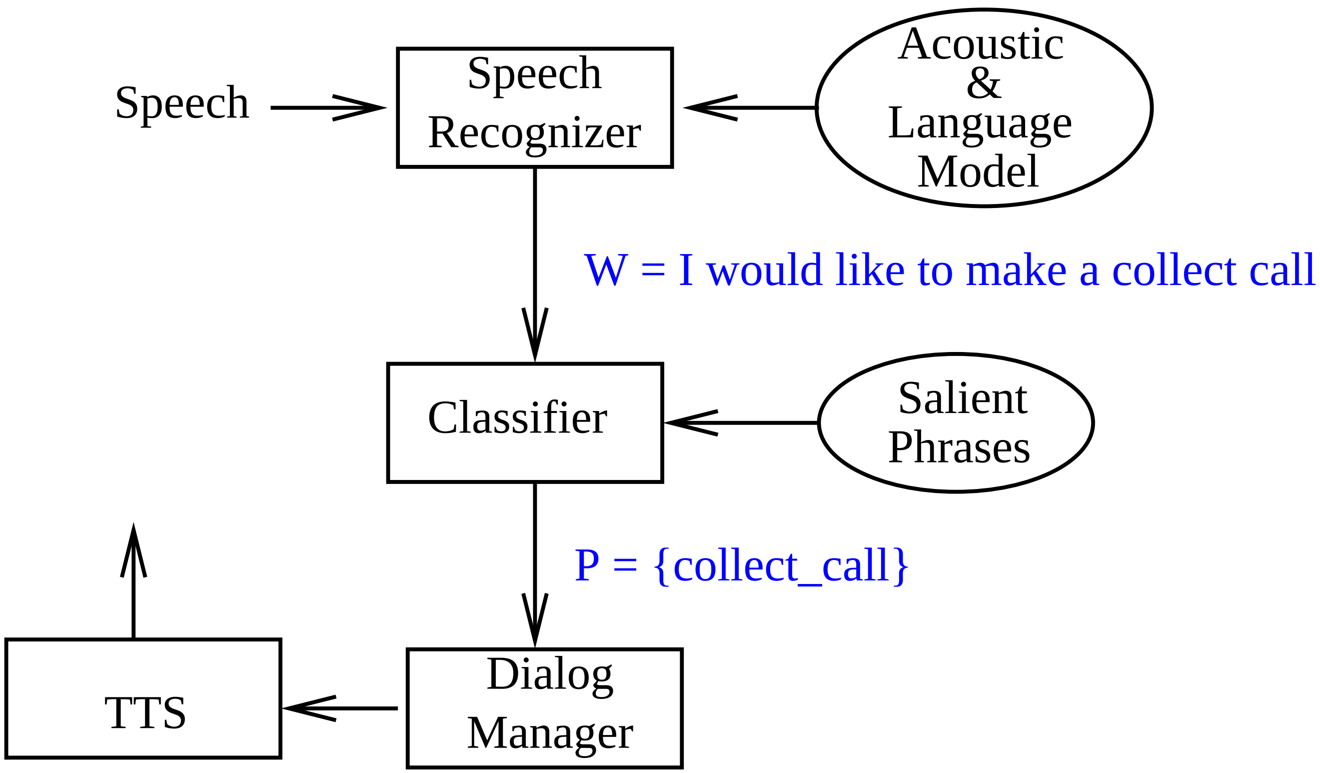
- Motivation
- Application Setting
- Learning Translation Models
- Experiments and Results

Motivation

- Finite-state Transducers (FST) are
  - Composable: allows for modular system development
  - Learnable from data
  - Efficient at runtime: Viterbi decoding
- Domain-specific language may be representable by an FST.
- Earlier work on FST-MT
  - Vilar et.al(1998)
  - Knight and Al-Onaizan(1998)
- Evaluation of translation in context of an application.

Application: How May I Help You?

- A Spoken Dialog System for call type classification.



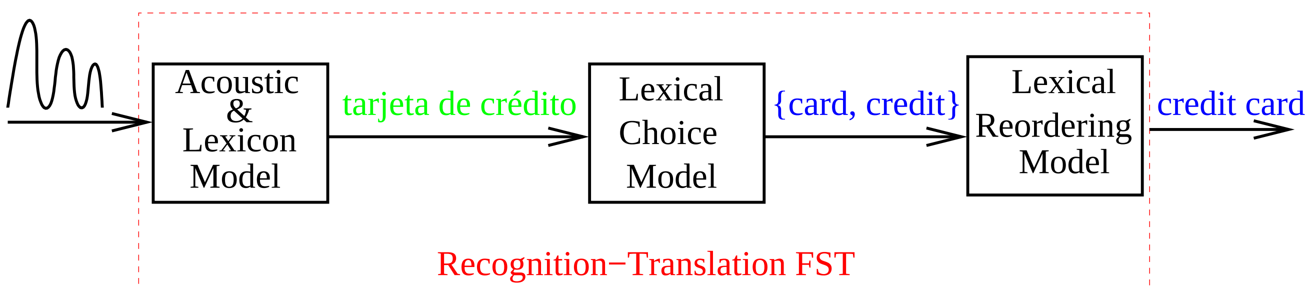
Customer Utterances

- Prompt: How may I help you?
- Utterance Length from 1-183 words, average of 18 words.
- Out-of-vocabulary rate per sentence is about 30%.
- Examples
  - Yeah I need the area code for rockmart georgia
  - Yes I'd like to make this long distance call area code x x x x x x x x x x
  - Yeah I'm wondering if you could place this call for me I can't seem to dial it it don't seem to want to go through for me

Speech Translation

- Previous approaches to Speech Translation
  - Source language ASR
  - Translation Model

- Finite-state Model based Speech Translation
  - Source Language Acoustic Model
  - Lexical Choice Model
  - Lexical Reordering Model



Learning Lexical Choice Models

Bitext

- English utterances recorded from customer calls.
- Manually translated into Spanish and Japanese.
- “Bunsetsu” like tokenization for Japanese.

English: I need to make a collect call
Japanese: 私は　コレクト　コールを　かける　必要があります

Spanish: ajá quiero usar mi tarjeta de crédito
English: yeah I wanna use my credit card

Alignment

Algorithm proposed by `Alshawi, Bangalore, Douglas in ACL98`

- Collect correlation information among reference-target word-pairs.
- Hierarchical decomposition of the source and target language sentences.
- Tree-alignment search using correlation costs.

English: I need to make a collect call

Japanese: 私は　コレクト　コールを　かける　必要があります
Alignment: 1 5 0 3 0 2 4

Spanish: ajá quiero usar mi tarjeta de crédito

English: yeah I wanna use my credit card
Alignment: 1 3 4 5 7 0 6

Bilanguage

- Bilanguage: Each token consists of a source language with its aligned target language word.
- Ordering of tokens: source language order or target language order.
- English ordered Bilanguage

I_私は need_必要があります to_ϵ make_コールを a_ϵ collect_コレクト call_かける

- Spanish ordered Bilanguage

ajá_yeah ϵ_I quiero_wanna usar_use mi_my tarjeta_card de_ϵ crédito_credit

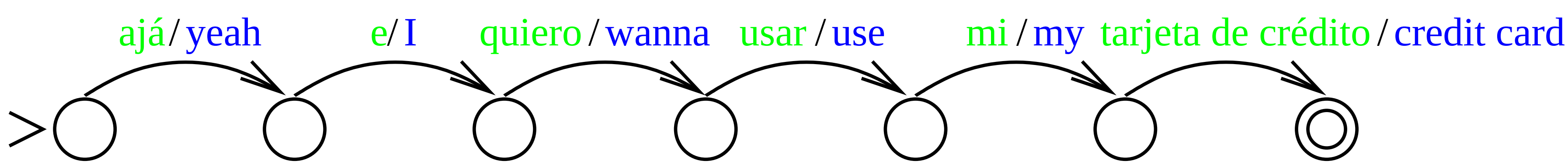
Learning Bilingual Phrases

- Weighted mutual information metric.
- Using suffix array representation to compute frequency.
- Length of a phrase: tunable parameter.
- Reordering within phrases.

私は コレクト コールを かける 必要があります	I need to make a collect call
はい あなたは いただけますか	yes could you
tarjeta de crédito	credit card
una llamada de larga distancia	a long distance call

Lexical Choice Transducer

- Language Model: N-gram model built on phrase-chunked bilanguage.
- A combination of phrases and words maximize predictive power and minimize number of parameters
- Resulting finite-state automaton on bilanguage vocabulary is converted into a finite-state transducer.



Lexical Reordering

- Output of the lexical choice transducer: sequence of target language phrases.
- Output: `like to make I'd call a calling card please`
- Words in phrases are in target language word order.
- However, phrases need to be reordered in target language word order.
- Reordered: `I'd like to make a calling card call please`

Lexical Reordering Models

- On-going work on developing various lexical reordering models.
- String-based model
  - Approximating a permutation lattice.
- Tree-based model
  - Impose a tree structure on a sentence.
  - Reordering using tree-local reordering rules.
- Compose with a target language N-gram model to retrieve the most likely sentence.

Experiments and Evaluation

- Data Collection:
  - The customer side of operator-customer conversations transcribed
  - Transcriptions were then manually translated into Spanish
- Training Set: 5812 Spanish-English sentence pairs
- Test Set: 829 Spanish sentences.
- Six different translation models
  - Unigram, Bigram, Trigram models
  - With and Without phrases

Accuracy of Lexical Choice

- Metric:
  - Treat the reference string and result string as a bag-of-tokens
  - Measure Recall and Precision figures

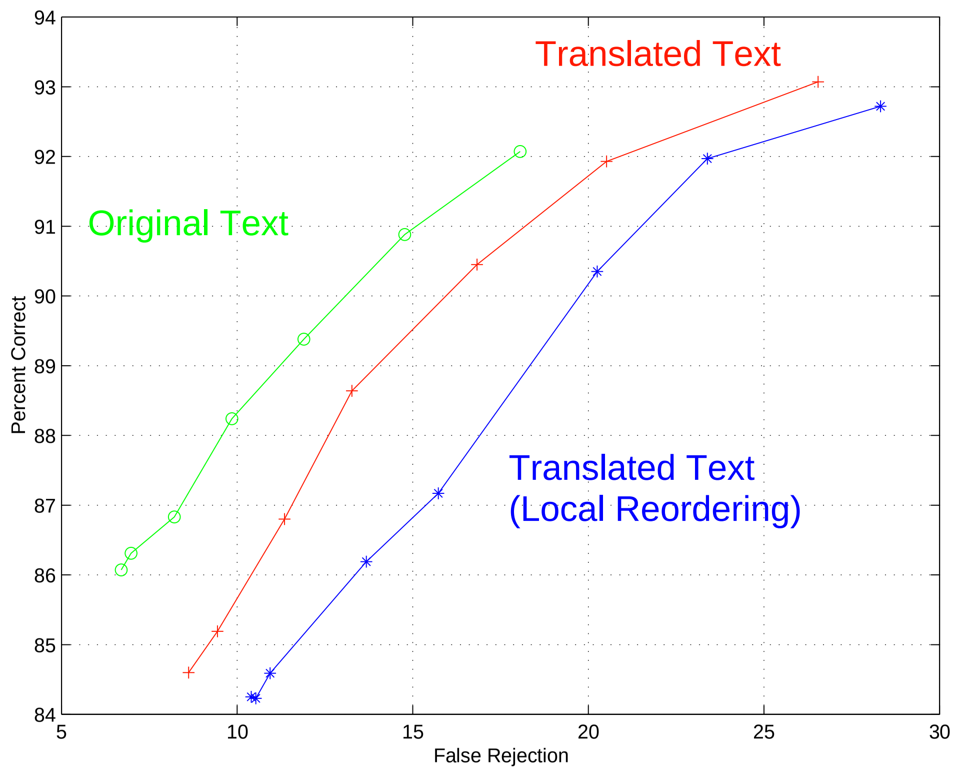
Trans VNSA order	Recall (R)	Precision (P)	F-Measure (2*P*R/(P+R))
Unigram	30.7	86.4	45.4
Bigram	69.0	84.4	75.8
Trigram	66.8	79.0	72.1
Phrase Unigram	45.6	86.6	59.7
Phrase Bigram	72.2	83.4	77.4
Phrase Trigram	71.9	83.1	77.1

Call Type Classification

- Exisiting Classifier:
  - Automatically trained on 8000 English transcription-Call type pairs.
  - 14 call-types plus rejection.
  - `area_code, billing_credit, rate, time` ...
- Dialog Module
  - Confirmation dialog
  - Clarification dialog
  - Completion dialog

Call Type Classification Accuracy

- Classify Spanish test set transcriptions based on their English translations into one of 14 call types.
- Plot false rejection against correct classification rate.



Summary

- FSTs provide a framework for integration of modules.
- Call-type accuracy using translated English text is almost the same as original English text.
- Enables applications to be multilingual.
- Integrated finite-state model for speech translation.