Simple, Lexicalized Choice of Translation Timing for Simultaneous Speech Translation

Tomoki Fujita, Graham Neubig, Sakriani Sakti, Tomoki Toda, Satoshi Nakamura

Graduate School of Information Science, Nara Institute of Science and Technology
Background
Speech Translation Systems

- Translate speech from source language to target
Problem: Delay

- Wait for the whole utterance to end before translating

こんにちは、駅はどこですか?

Hello, where is the station?
Solution: Divide into Smaller Chunks

• Choose appropriate timing to start translation

Delay: Reduced
Previous Work: Incremental Dependency Parsing/Manual Rules [Ryu+ 04]

- Utilize knowledge of English/Japanese to derive rules

\[ I \text{ went to the park with your brother } \]

Translate after the first prepositional phrase completes!

- Requires a bilingual linguist to design rules
- Requires an accurate incremental dependency parser
Previous Work: Division on Pauses
[Fugen+ 08, Bangalore+ 12]

- Simply divide on short pauses in the utterance

- Cannot capture relationship between languages

- Result will greatly change with speech speed, disfluencies
Proposed Method

- Utilize the TM directly to choose translation timing
  - + Can be constructed automatically
  - + Uses information about the language pair
  - + Very simple to implement
- Specifically:
  - Choose translation timing at the end of each phrase in the phrase table
  - Utilize reordering probabilities to adjust granularity
  - Adapt the language model to the translation task
Preliminaries
Phrase Based Machine Translation

• Divide the sentence into small phrases and translate

Today I will give a lecture on machine translation.

今日は、機械翻訳の講義を行います。

• Score translations with translation model (TM), reordering model (RM), and language model (LM)
Translation Model Creation

- Perform automatic alignment of bitext
- From aligned text, extract phrases for translation

| ホテルの受付 | ホテル の → hotel |
| ホテルの受付 | ホテル の → the hotel |
| 受付 → front desk | 受付 → front desk |
| ホテルの受付 → hotel front desk | ホテルの受付 → the hotel front desk |
Lexicalized Reordering Model

- Probabilistically models reorderings for increased accuracy of translation
- Given current phrase and next phrase:

<table>
<thead>
<tr>
<th>Monotone:</th>
<th>Swap:</th>
</tr>
</thead>
<tbody>
<tr>
<td>背の高い男</td>
<td>太郎を訪問した</td>
</tr>
<tr>
<td>the tall man</td>
<td>visited Taro</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discontinuous Right:</th>
<th>Discontinuous Left:</th>
</tr>
</thead>
<tbody>
<tr>
<td>私は太郎を訪問した</td>
<td>背の高い男を訪問した</td>
</tr>
<tr>
<td>I visited Taro</td>
<td>visited the tall man</td>
</tr>
</tbody>
</table>

- “monotone” + “discontinuous right” = “right probability”
Proposed Method
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
<td>こんにちは</td>
</tr>
<tr>
<td>where</td>
<td>どこ</td>
</tr>
<tr>
<td>is</td>
<td>ですか</td>
</tr>
<tr>
<td>the</td>
<td>その</td>
</tr>
<tr>
<td>station</td>
<td>駅</td>
</tr>
</tbody>
</table>

**Input String**

hello  where  is  the  station
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

- `hello` → こんにちは
- `where` → どこ
- `where is` → どこですか
- `the` → その
- `the station` → 駅

**Input String**

- `hello`
- `where`
- `is`
- `the`
- `station`
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**
- hello → こんにちは
- where is → どこですか
- the station → 駅
- where → どこ
- the → その

**Input String**
- hello
- where
- is
- the
- station

“hello” phrase exists

↓

wait
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

- **hello** $\rightarrow$ こんにちは
- **where** $\rightarrow$ どこ
- **is** $\rightarrow$ ですか
- **the** $\rightarrow$ その
- **station** $\rightarrow$ 駅

**Input String**

hello where is the station

"hello" phrase exists

wait
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
<td>こんにちは</td>
</tr>
<tr>
<td>where</td>
<td>どこ</td>
</tr>
<tr>
<td>is</td>
<td>ですか</td>
</tr>
<tr>
<td>the station</td>
<td>駅</td>
</tr>
<tr>
<td>the</td>
<td>その</td>
</tr>
</tbody>
</table>

**Input String**

```
hello   where   is   the   station
```

“hello” phrase exists → wait
“hello where” phrase missing → translate
“hello”
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

- `hello` → こんにちは
- `where` → どこ
- `the` → その
- `where is` → どこですか
- `the station` → 駅

**Input String**

- `hello`  
- `where`  
- `is`  
- `the`  
- `station`

“hello” phrase exists  
↓  
wait  

“hello where” phrase missing  
↓  
translate  

“hello”
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
<td>こんにちは</td>
</tr>
<tr>
<td>where</td>
<td>どこ</td>
</tr>
<tr>
<td>where is</td>
<td>どこですか</td>
</tr>
<tr>
<td>the station</td>
<td>駅</td>
</tr>
<tr>
<td>the</td>
<td>その</td>
</tr>
</tbody>
</table>

**Input String**

```
hello  where  is  the  station
```

- “hello” phrase exists
  - wait
- “hello where” phrase missing
  - translate “hello”
- “where is” phrase exists
  - wait
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

Phrase Table

<table>
<thead>
<tr>
<th>English</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
<td>こんにちは</td>
</tr>
<tr>
<td>where</td>
<td>どこ</td>
</tr>
<tr>
<td>where is</td>
<td>どこですか</td>
</tr>
<tr>
<td>the station</td>
<td>駅</td>
</tr>
<tr>
<td>the</td>
<td>その</td>
</tr>
</tbody>
</table>

Input String

hello where is the station

“hello” phrase exists → wait
“hello where” phrase missing → translate “hello”
“where is” phrase exists → wait
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

Phrase Table
- hello → こんにちは
- where → どこ
- where is → どこですか
- the → その
- the station → 駅

Input String
- hello where is the station

- "hello" phrase exists
  - wait
- "hello where" phrase missing
  - translate "hello"
- "where is" phrase exists
  - wait
- "where is the" phrase missing
  - translate "where is"
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

<table>
<thead>
<tr>
<th>Phrase Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello → こんにちは</td>
</tr>
<tr>
<td>where → どこ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Input String</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
</tr>
</tbody>
</table>

```
“hello” phrase exists  ↓  wait
“hello where” phrase missing  ↓  translate “hello”
“where is” phrase exists  ↓  wait
“where is the” phrase missing  ↓  translate “where is”
```
Method One: Choosing Translation Timing with Phrases

- Input words one at a time from ASR
- While words exist in phrase table, don't translate yet

**Phrase Table**

- `hello` → こんにちは
- `where` → どこ
- `where is` → どこですか
- `the` → その
- `the station` → 駅

**Input String**

```
hello  where   is   the   station
```

```
“hello” phrase exists
↓ wait

“hello where” phrase missing
↓ translate “hello”

“where is” phrase exists
↓ wait

“where is the” phrase missing
↓ translate “where is”

“the station” utterance ends
↓ translate “the station”
```
Problem with Method One

- Has the potential to degrade translation accuracy:

**Normal phrase-based translation:**

こんにちは 駅 は どこ ですか
Hello, where is the station

**Translation with early timing:**

こんにちは 駅 は どこ ですか
Hello, the station where is it
Method Two: Adjusting Timing with Reordering Probabilities

- First, temporarily choose strings according to method one
- Next, if that phrase's right probability exceeds a threshold, actually translate the words in the cache

Example (threshold = 0.8):

<table>
<thead>
<tr>
<th>hello</th>
<th>where</th>
<th>is</th>
<th>the</th>
<th>station</th>
</tr>
</thead>
<tbody>
<tr>
<td>“hello” phrase exists</td>
<td>“hello where” phrase missing</td>
<td>“where is” phrase exists</td>
<td>“where is the” phrase missing</td>
<td>“the station” utterance ends</td>
</tr>
<tr>
<td>wait</td>
<td>choose “hello”</td>
<td>wait</td>
<td>choose “where is”</td>
<td>translate “where is the station”</td>
</tr>
<tr>
<td>right probability is 0.9 &gt; 0.8</td>
<td>right probability is 0.6 &lt; 0.8</td>
<td>do not translate yet</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Threshold 1.0 = traditional, 0.0 = method one
Problem with Method Two

- LMs are traditionally trained on sentences
  - This is not appropriate for translating shorter chunks
  - e.g.: The translator will try to “finish” sentences

こんにちは は どこ ですか
Hello . Where is the station ? Where is it .
Method Three: Language Model Adaptation

- Before learning the language model, **split the training data** according to the same criterion

**Traditional LM Training**

- Hello, where is the station.
- My name is John.

**Proposed Method**

- Hello, where is the station.
- My name is John.
Experiments
Experimental Setup

- **Four Types of Experiments:**
  - Japanese-English BTEC Travel Conversation (ja-en)
  - Japanese-English BTEC with 11+ Words (ja-en 11+)
  - English-Japanese BTEC Travel Conversation (en-ja)
  - French-English WMT News (fr-en)

- **Evaluation Measures:**
  - Accuracy
    - 14-ref BLEU for BTEC, 1-ref BLEU for News
    - Manually-graded acceptability
  - Delay (Seconds)
Result One: Comparison Across Settings

- Delay decreases in all settings
- Better delay/accuracy tradeoff for long sentences, similar languages
Result Two: Compare with Pause-based Segmentation

- In faster settings proposed method best
- In slower settings pause-based method best
Manual Evaluation

- Decrease in manual evaluation as well, but less obvious than evaluated by BLEU
Conclusion
Conclusion

- Proposed method for choosing timing in speech translation using phrase table and reordering model
  - Considers reordering tendencies across languages
  - Simple and language independent
  - Competitive accuracy
- Future work:
  - Combination of prosodic, reordering, and syntactic cues
  - How do we evaluate translations?
Thank You!
Example:

サーフィン に いい 場所 を 教え て ください

please tell me a good surfing place

サーフィン に いい 場所 を 教え て ください

for surfing | please tell me a good place