

Introduction

- ▶ spoken language translation
 - ▶ recognition of spoken language with an ASR system
 - ▶ translation of the ASR output with an SMT system
- ▶ mismatch between training data and input of the SMT system
 - ▶ SMT system is trained on written language data
 - ▶ but the input is ASR output
- ▶ create bilingual corpus with ASR output as source language
 - ▶ source language data are recognized with an ASR system
 - ▶ train an SMT system on this new corpus

Automatically Transcribed Text in Training

two approaches using automatically transcribed text:

- ▶ direct translation of ASR output
 - ▶ target language data: translation of the manual transcriptions
- ▶ translate ASR output to manual transcription
 - ▶ target language data: manual transcription
 - ▶ postprocessing of the ASR output as translation process
 - ▶ translate postprocessed ASR output with text MT system
- ▶ in addition
 - ▶ combine manual with automatic transcriptions
 - ▶ extract n -Best lists from lattices to create a larger corpus

data sources to create new bilingual corpus:

- ▶ **TED talks**
 - ▶ audio recordings \Rightarrow manual transcriptions \Rightarrow translations
- ▶ **WIT³**
 - ▶ TED talks transcriptions and translations as bilingual corpus

Sentence Alignment

- ▶ ASR system does not provide sentence-wise segmentation
- ▶ but SMT systems are trained on bilingual parallel corpora
- ▶ automatic re-segmentation algorithm [Matusov et al., IWSLT 2005]
 - ▶ create Levenshtein alignment between the recognition and its manual transcription
 - ▶ sentence segmentation of the manual transcription is transferred to the recognition
- ▶ for the direct translation approach:
 - ▶ replace the manual transcription with its translation

ASR and SMT system

- ▶ ASR system
 - ▶ based on [Sundermeyer et al., ICASSP 2011]
 - ▶ language model data: GigaWord v4, TED, acoustic transcriptions
- ▶ SMT system
 - ▶ RWTH's open-source translation toolkit, phrase-based decoder [Wuebker et al., CoLing 2012]

Experimental Evaluation

- ▶ evaluated on the IWSLT 2012 English-to-French SLT task
- ▶ punctuation and case information are restored during translation [Matusov et al., IWSLT 2006]
- ▶ **MANUAL-TRANSCRIPTION**: bilingual corpus provided by WIT³
 - ▶ remove punctuation and casing on the source side (pseudo ASR output)
- ▶ **AUTOMATIC-TRANSCRIPTION**: bilingual corpus with ASR output as source language data
 - ▶ 1028 talks (around 250 hours of speech)
- ▶ phrase table and data combination
 - ▶ union (\cup MANUAL-TRANSCRIPTION)
 - ▶ two phrase tables (\circ MANUAL-TRANSCRIPTION)
 - ▶ training data concatenation ($+$ MANUAL-TRANSCRIPTION)

setup	tst2010	
	BLEU	TER
MANUAL-TRANSCRIPTION (baseline)	21.1	62.5
AUTOMATIC-TRANSCRIPTION	21.3	62.3
\cup MANUAL-TRANSCRIPTION	21.2	62.2
\circ MANUAL-TRANSCRIPTION	21.6	62.1
$+$ MANUAL-TRANSCRIPTION	21.5	61.8
AUTOMATIC-TRANSCRIPTION (1-Best)	21.1	62.4
AUTOMATIC-TRANSCRIPTION (10-Best)	21.0	62.3
AUTOMATIC-TRANSCRIPTION (20-Best)	21.2	62.4

- ▶ ASR output **POSTPROCESSING**
- ▶ compared with **IMPLICIT** [Matusov et al., IWSLT 2006]
- ▶ best SMT system from [Peitz et al., IWSLT 2012]

method	dev2010		tst2010	
	BLEU	TER	BLEU	TER
IMPLICIT	19.2	67.8	22.5	61.6
POSTPROCESSING	20.1	67.2	23.4	60.7

Corpus Example

Corpus	
manual transcription	So you can double efficiency with a 60 percent internal rate of return .
pseudo ASR output	so you can double efficiency with a 60 percent internal rate of return
automatic transcription	so you can you can double efficiency with a sixty percent internal rate of return
manual translation	Donc vous pouvez doubler votre efficacite nergtique avec un Taux de Rendement Interne de 60 % .

Translation Example

Input/Translations	
automatic transcription	and you know i i thought well i 'm i 'm like living in a science fiction movie
manual transcription	and I thought like , " Wow . I am like living in a science fiction movie .
POSTPROCESSING	and , you know , I thought , " Well , I 'm like living in a science fiction movie .
IMPLICIT translation	et , vous savez , je me suis dit : " Eh bien , je suis comme je suis vivant dans un film de science-fiction .
POSTPROCESSING translation	et , vous savez , j' ai pens : " Eh bien , je suis vivant dans un film de science-fiction .
reference translation	et IÃ j' ai pens : " Wow . c' est comme si je vivais dans un film de science-fiction .