

**transLectures**

Transcription and Translation of Video Lectures



# Domain Adaptation For Subtitles You Can Take Seriously

**Muhammad Ali Tahir**

**transLectures Workshop, Ljubljana  
April 22nd, 2014**

# Outline

- 1. Task definition for transLectures**
- 2. Main challenges and opportunities**
- 3. Adaptation of acoustic models**
- 4. Adaptation of language models**
- 5. Adaptation of translation models**
- 6. Experimental results**
- 7. Tools for adaptation**
- 8. Conclusions**

# Task Definition

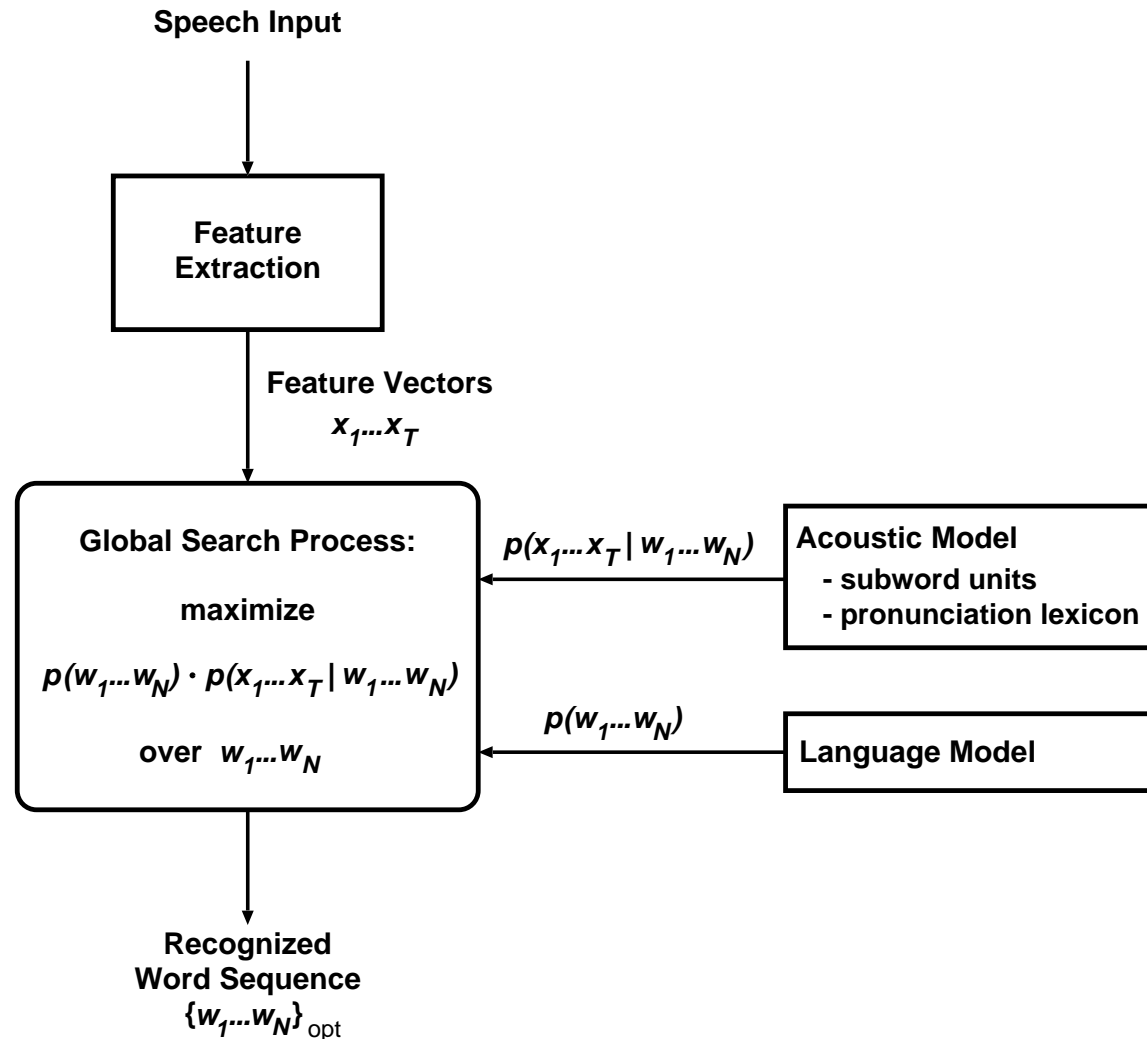
- ▶ **Automatic transcription and translation of video lectures**
- ▶ **Basic components of the system**
  - ▷ **Audio → Text (automatic speech recognition)**
  - ▷ **Text → Text in another language (statistical machine translation)**
- ▶ **Can we make use of the extra information?**
  - ▷ **Slides → Slides text (helpful for speech recognition)**

# Translectures: Challenges and Opportunities

- ▶ **Lecture recognition and translation:**
  - ▷ Heterogeneous conditions (non-native speakers, echo in lecture rooms ...)
  - ▷ Broad range of topics (technical terms ...)
  - ▷ Need for adaptation methods
  
- ▶ **Large amounts of data for transcription and translation:**
  - ▷ `videlectures.net` and `poliMedia`
  - ▷ Recognition of 7,000 hours of speech
  - ▷ Translation into multiple languages, e. g.,  
from English into {French, German, Slovenian, Spanish}
  - ▷ Focus on efficient adaptation methods
  
- ▶ **Availability of lecture slides**
  - ▷ Exploitation of prior knowledge

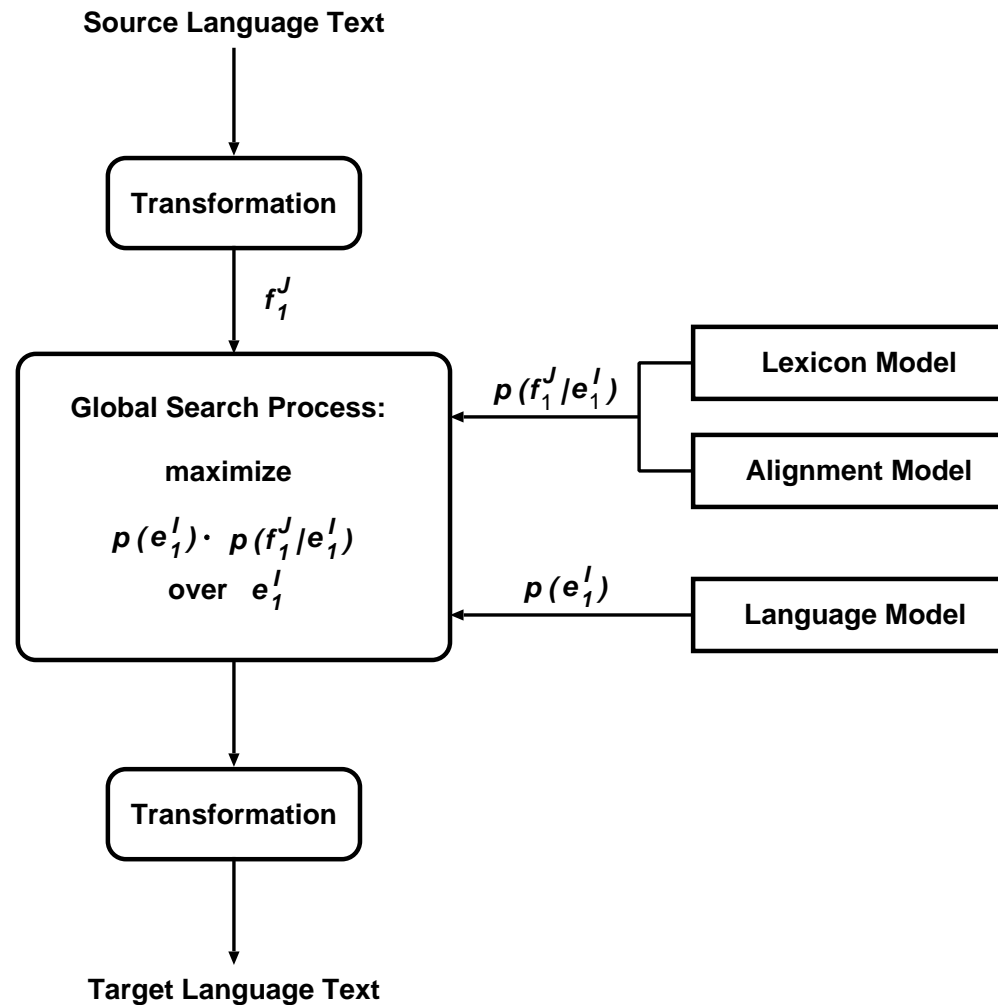
# Automatic Speech Recognition

- ▶ Conversion of audio into text
- ▶ Four main components
  - ▷ Signal analysis
  - ▷ Acoustic model
  - ▷ Language model
  - ▷ Search



# Statistical Machine Translation

- ▶ Conversion of text in one language into another language
- ▶ Four main components
  - ▷ Translation model (lexicon + alignment)
  - ▷ Language model
  - ▷ Transformations (pre-/post-processing)
  - ▷ Search



# Performance Metrics

## ► Automatic speech recognition

### ▷ Word error rate (WER)

$$\frac{\text{deletions} + \text{insertions} + \text{substitutions}}{\text{reference words}}$$

		E	L	E	P	H	A	N	T
	0	1	2	3	4	5	6	7	8
R	1	1	2	3	4	5	6	7	8
E	2	1	2	2	3	4	5	6	7
L	3	2	1	2	3	4	5	6	7
E	4	3	2	1	2	3	4	5	6
V	5	4	3	2	2	3	4	5	6
A	6	5	4	3	3	3	3	4	5
N	7	6	5	4	4	4	4	3	4
T	8	7	6	5	5	5	5	4	3

## ► Statistical machine translation

### ▷ Bilingual evaluation understudy (BLEU)

# Detailed Project Goals

## ▶ Partners involved:

- ▶ RWTH Aachen University (RWTH)
- ▶ European Media Laboratory GmbH (EML)
- ▶ Polytechnic University of Valencia (UPVLC)
- ▶ Xerox Research Center Europe (XEROX)

## ▶ Massive adaptation of models in ASR and SMT

- ▶ Adaptation of acoustic models
- ▶ Adaptation of language models
- ▶ Adaptation of translation models



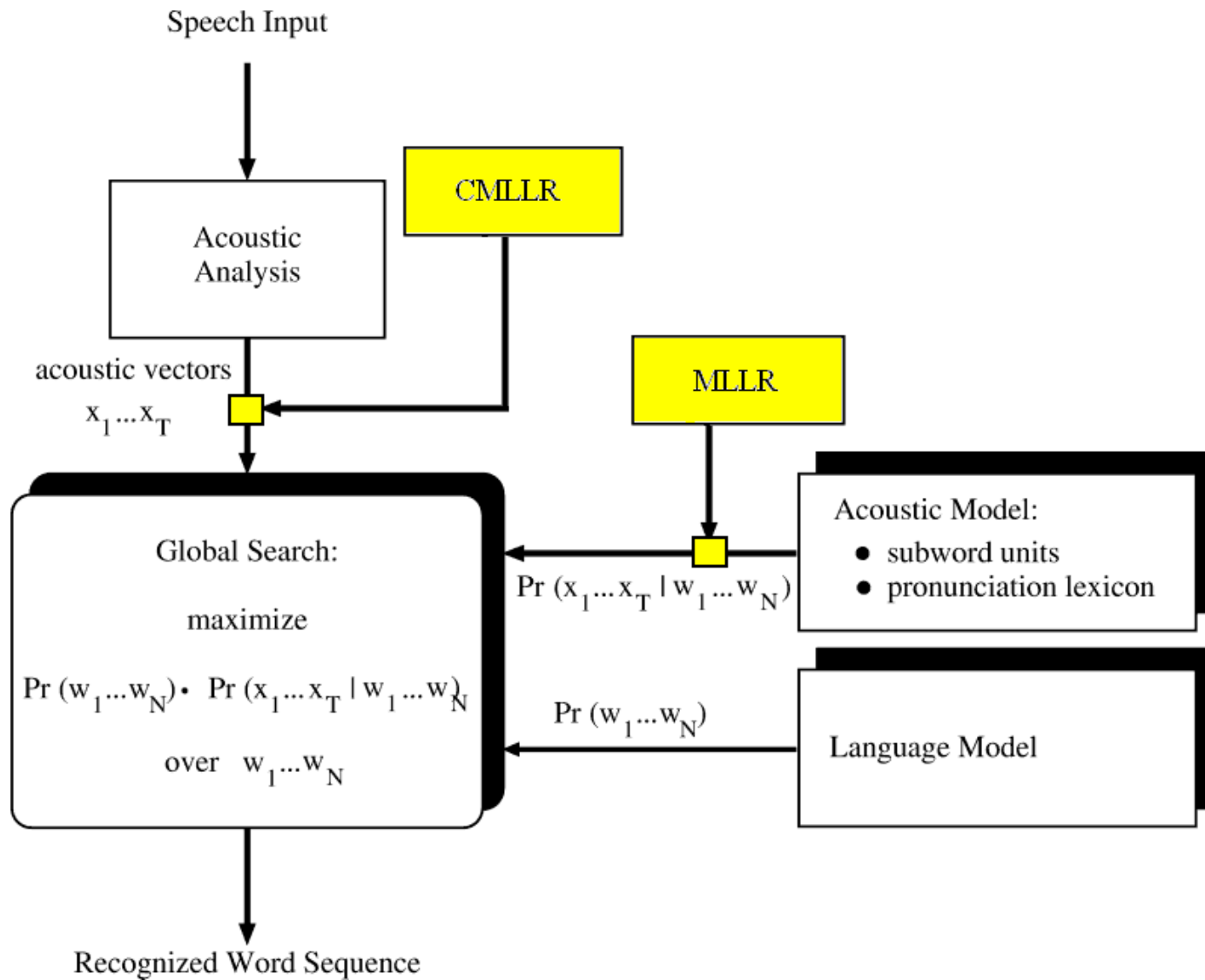
# Adaptation of Acoustic Models

- ▶ **Investigations on acoustic model adaptation methods**
  - ▷ **CMLLR (constrained maximum likelihood linear regression)**
  - ▷ **MLLR (maximum likelihood linear regression)**
  - ▷ **MAP (maximum a-posteriori adaptation)**
  
- ▶ **Adaptation by neural networks**

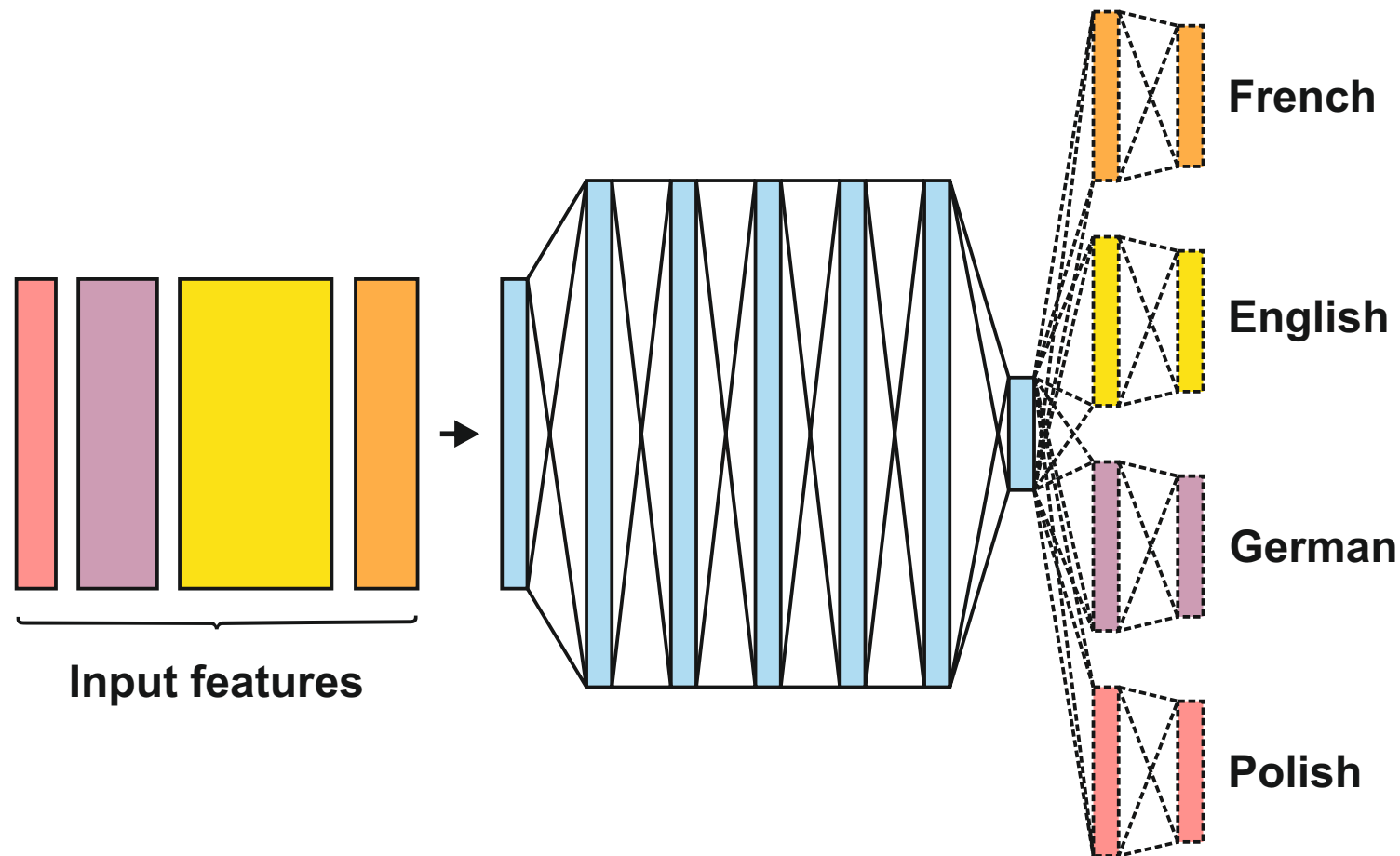
**Directions of research:**

  - ▷ **Shallow vs. deep**
  - ▷ **Tandem vs. hybrid**
  - ▷ **Mono- vs. multilingual**

# Adaptation of Acoustic Models



# Multilingual Neural Network Adaptation



- ▶ Adaptation with data from other languages
- ▶ Dimensionality reduction by bottleneck layer

# Languages and Training Data

<b>Repository</b>	<b>Language</b>	<b>AM Hours</b>	<b>LM Words</b>
<b>videolectures.net</b>	<b>English</b>	<b>1006</b>	<b>6600 M</b>
	<b>Slovenian</b>	<b>90</b>	<b>75 M</b>
<b>poliMedia</b>	<b>Spanish</b>	<b>390</b>	<b>1600 M</b>
	<b>Catalan</b>	<b>49</b>	<b>141 M</b>

# Results for Acoustic Model Adaptation

## ► Improvements by acoustic model adaptation only

Language	without AM adaptation	with AM adaptation
English	27.9 %	22.8 %
Spanish	17.0 %	15.6 %
Slovenian	45.4 %	38.1 %

## ► Additional gains by system combination (Spanish)

- ▷ System 1 WER: 17.0%
- ▷ System 2 WER: 15.6%
- ▷ System 1 + System 2 WER: 15.3%

# Adaptation of Language Models

- ▶ **Adaptation by interpolation with in-domain data**
- ▶ **Refinement: adaptation with slides**
  - ▷ **Adapt LM with *all* available slides**
  - ▷ **Adapt LM with *specific* slides of the given lecture**
  - ▷ **Improved OCR (with help from other project)**
- ▶ **Language model cache**

# Results for Language Model Adaptation

## ► Word error rate results for language model adaptation

Language	AM Adaptation	AM+LM Adaptation
Catalan	—	35.5 %
English	22.8 %	21.2 %
Spanish	15.6 %	13.5 %
Slovenian	38.1 %	36.2 %

## ► Improvements on top of acoustic adaptation

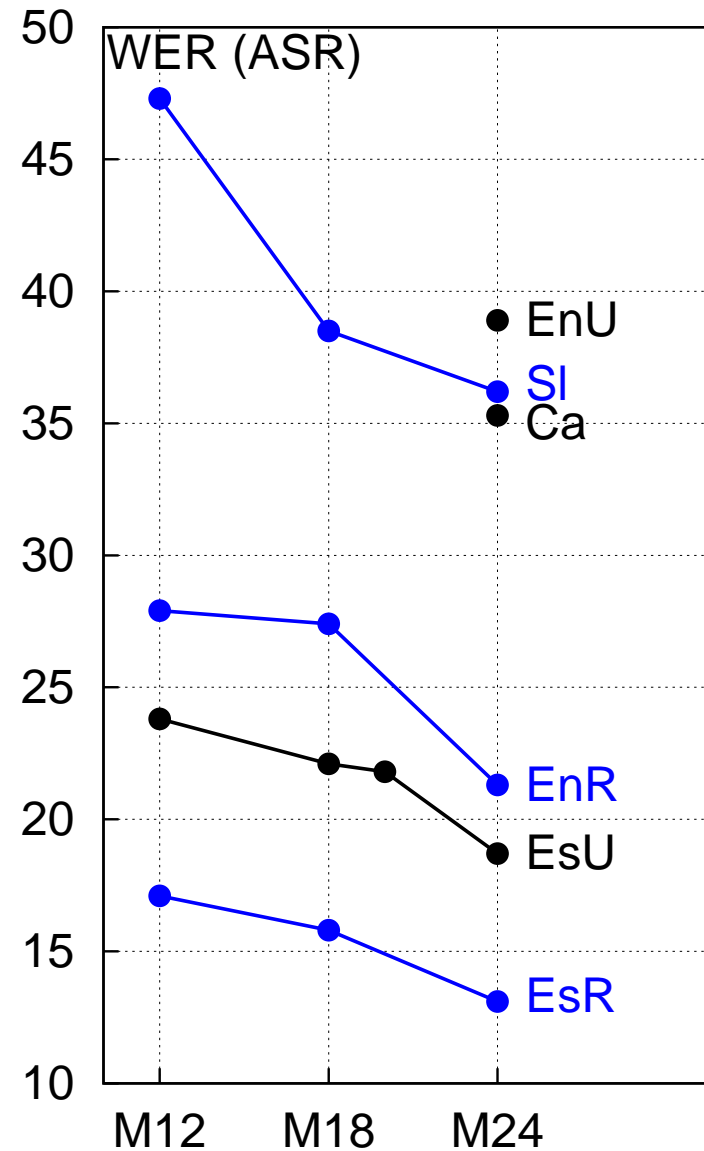
# Impact of More Data vs. Adaptation

- ▶ How much do additional data help?
- ▶ Case study for English ASR, word error rate results

System	WER
Baseline	27.9 %
+AM Adaptation	23.1 %
+Additional Data	22.8 %
+LM Adaptation	21.2 %



# Improvements for Automatic Speech Recognition



# Adaptation of Translation Models

- ▶ **Efforts on LM adaptation methods:**
  - ▷ **Adaptation with more in-domain data**
  - ▷ **Domain specific LM array**
  - ▷ **Lexical coverage features**
  
- ▶ **Data weighting and data selection**
  
- ▶ **Measure similarity to target domain**
  - ▷ **Using the language model**
  - ▷ **Using the translation model**
  
- ▶ **Data selection based on text that is being translated**

# Language Pairs and Training Data

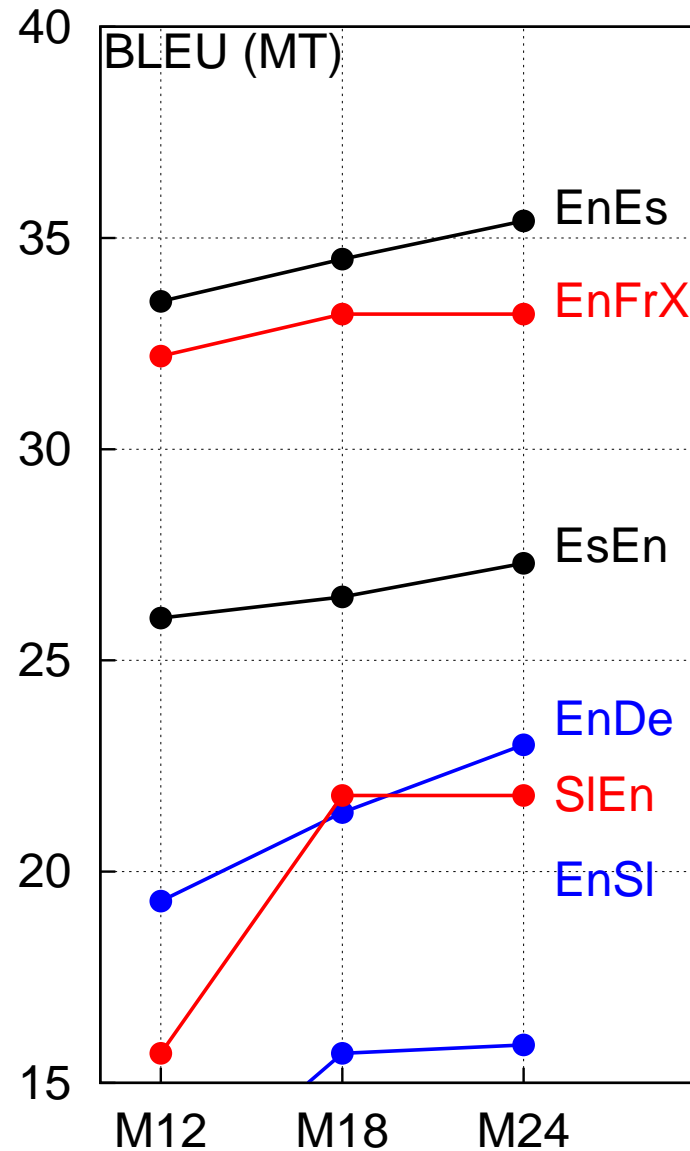
Repository	Task	Sentences
poliMedia	Es → En	17.9 M
	En → Es	17.9 M
	En → Fr	38.7 M
videolectures.net	En → De	12.8 M
	En → Sl	13.4 M
	Sl → En	13.4 M

# Results for Translation Model Adaptation

## Translation quality in BLEU

<b>Task</b>	<b>Baseline</b>	<b>TM Adaptation</b>
<b>Es→En</b>	<b>26.0</b>	<b>27.3</b>
<b>En→Es</b>	<b>33.5</b>	<b>35.4</b>
<b>En→Fr</b>	<b>32.4</b>	<b>35.0</b>
<b>En→De</b>	<b>19.3</b>	<b>23.0</b>
<b>En→Sl</b>	<b>12.0</b>	<b>15.9</b>
<b>Sl→En</b>	<b>15.7</b>	<b>21.8</b>

# Improvements for Statistical Machine Translation



# Tools for Adaptation

- ▶ **EML**
  - ▷ **ASR: EML Transcription Platform**
- ▶ **RWTH**
  - ▷ **ASR: RASR**
  - ▷ **SMT: Jane**
- ▶ **UPLVC**
  - ▷ **ASR: TLK**
- ▶ **XEROX**
  - ▷ **SMT: TunaTon Toolkit**

# Conclusions

## ▶ ASR

- ▶ Large improvements of 20-30 % relative for all languages
- ▶ English and Spanish: accurate enough transcriptions
- ▶ AM adaptation improved by neural networks
- ▶ LM adaptation improved by slides (better OCR)
- ▶ Initial release of UPVLC TLK toolkit

## ▶ SMT

- ▶ Either high quality scores or large improvements
- ▶ Translation model adaptation improved by data selection and data weighting

## ▶ Future Work

- ▶ Improved Slovenian ASR, Slovenian and German SMT

**Thank you for your attention!**



# Ongoing Work

