Paying Attention to Multi-Word Expressions in Neural Machine Translation

Matīss Rikters[1] and Ondřej Bojar[2]

[1]University of Latvia, Faculty of Computing

[2]Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

The 16th Machine Translation Summit

Nagoya, Japan

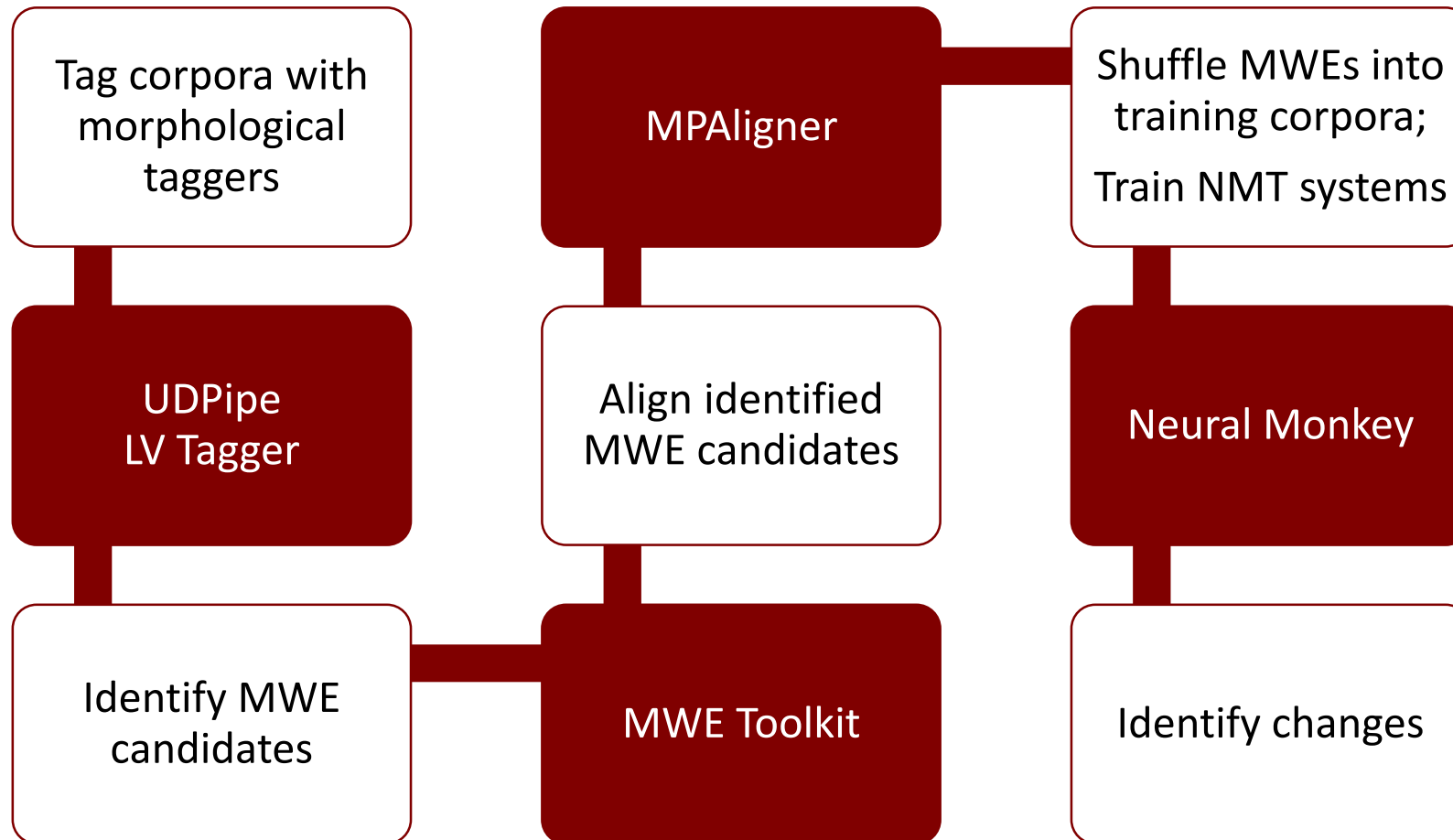September 22, 2017

Contents

- Introduction

- Related Work

- Workflow

  - Data

  - NMT Systems

- Experiments

  - Results

  - Manual Inspection

  - Attention Inspection

- Conclusions

Introduction

Related Work

Workflow

Tag corpora with morphological taggers

MPAligner

Shuffle MWEs into training corpora;
Train NMT systems

UDPipe
LV Tagger

Align identified MWE candidates

Neural Monkey

Identify MWE candidates
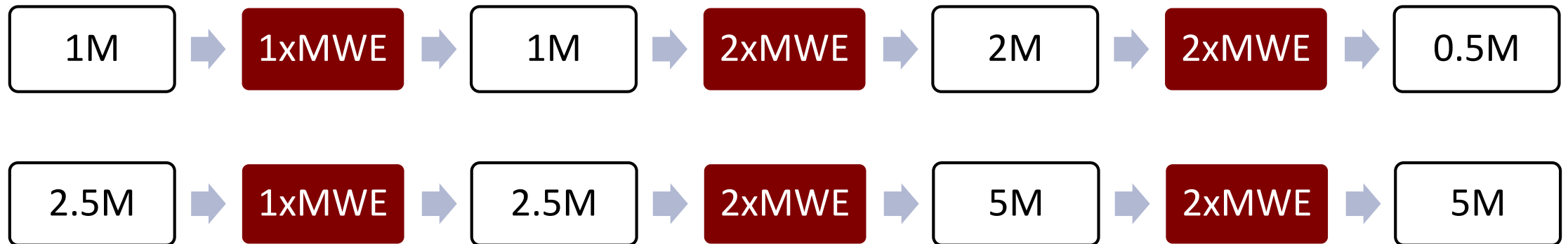
MWE Toolkit

Identify changes

Data

- WMT17 News Translation Task
  - Training
    - En → Lv – 4.5M sentences for the baseline; 4.8M after adding MEWs/MWE sentences
    - En → Cs – 49M sentences for the baseline; 17M after adding MEWs/MWE sentences
  - Evaluation
    - En → Lv – 2003 sentences in total; 611 sentences with at least one MWE
    - En → Cs – 6000 sentences in total; 112 sentences with at least one MWE

| 1M | ➡ | 1xMWE | ➡ | 1M | ➡ | 2xMWE | ➡ | 2M | ➡ | 2xMWE | ➡ | 0.5M |

| 2.5M | ➡ | 1xMWE | ➡ | 2.5M | ➡ | 2xMWE | ➡ | 5M | ➡ | 2xMWE | ➡ | 5M |

NMT Systems

Experiments

Results

Languages	En → Cs		En → Lv	
Dataset	Dev	MWE	Dev	MWE
Baseline	13.71	10.25	11.29	9.32
+MWE phrases	-	-	**11.94**	**10.31**
+MWE sentences	**13.99**	**10.44**	-	-

Manual Inspection

Alignment Inspection

Conclusions

Code & Presentation