Visualizing Neural Machine Translation Attention and Confidence

Matīss Rikters

Faculty of Computing,
University of Latvia
matiss@lielakeda.lv

Mark Fishel

Institute of Computer Science,
University of Tartu
fishel@ut.ee

Ondřej Bojar

Charles University,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
bojar@ufal.mff.cuni.cz

Confidence Scores

$$CDP = \frac{1}{J}\sum_{j}\log\left(1 + \left(\sum_{i}\propto_{ji}\right)^2\right)$$

$$AP_{out} = -\frac{1}{I}\sum_{i}\sum_{j}\propto_{ji}\cdot\log\propto_{ji}$$

$$AP_{in} = -\frac{1}{I}\sum_{j}\sum_{i}\propto_{ij}\cdot\log\propto_{ij}$$

$$confidence = CDP + AP_{out} + AP_{in}$$

$$percentage = e^{-C(X^2)}$$

Terminal Visualisations


```
Coverage Deviation Penalty:
Input Absentmindedness Penalty:
Output Absentmindedness Penalty:
Confidence:
        -0.2329971657 (94.72%)
        -0.7235557273 (97.42%)
        -0.6299903148 (98.04%)
        -1.5865432078 (88.17%)
```

Browser Visualisations
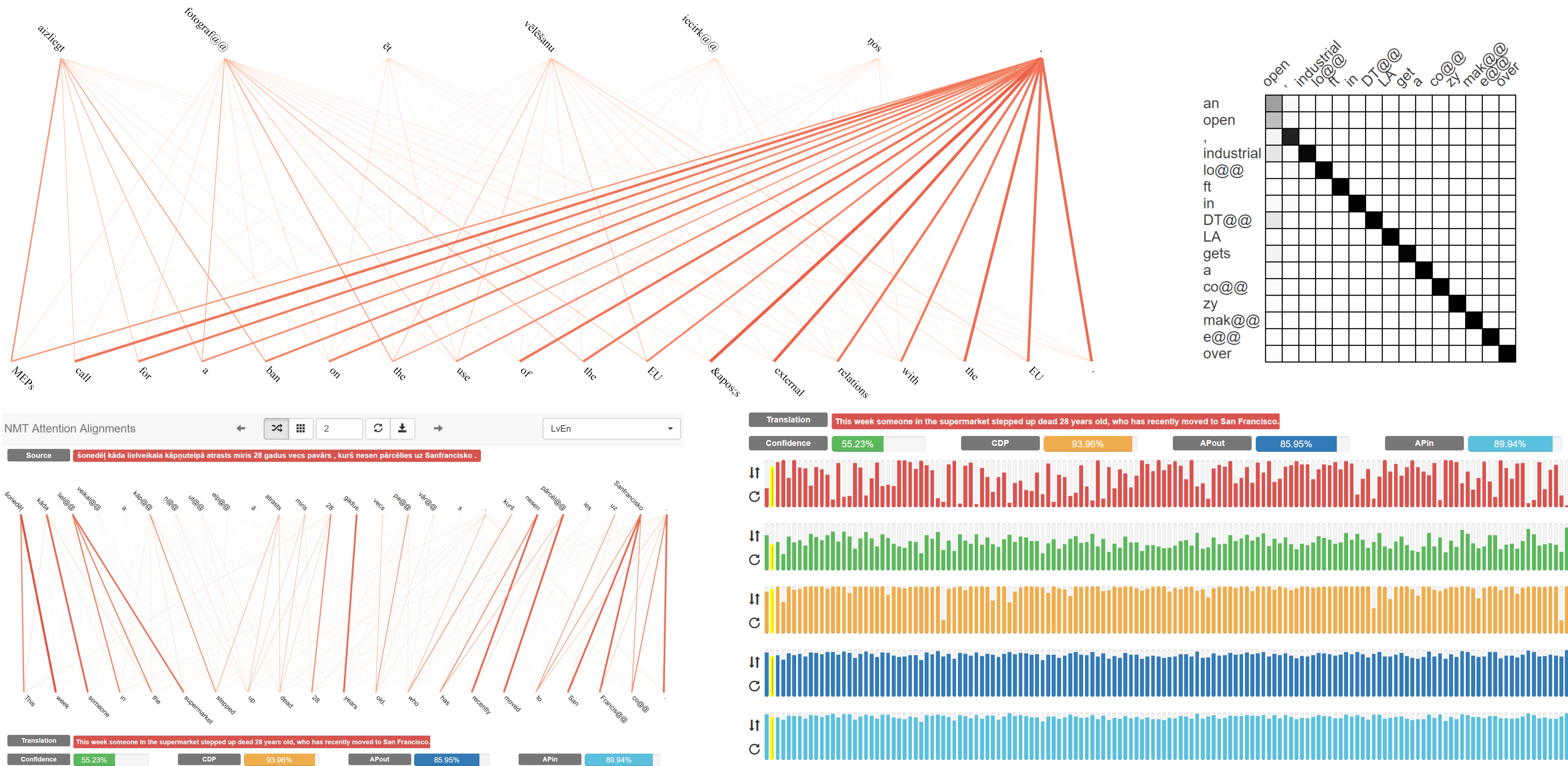
Features

Works with attention alignment data from

. Nematus

. Neural Monkey

. AmuNMT (fork github.com/barvins/amunmt)

Visualise translations in

. Linux Terminal or Windows PowerShell

. Web browser

. Line form or matrix form

. Save as PNG

. Sort and navigate dataset by confidence scores

GitHub

ej.uz/nmt-github

Poster

ej.uz/nmt-poster

Demo

ej.uz/nmt-attention

Acknowledgements