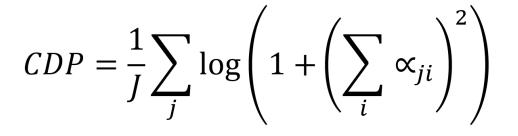# Visualizing Neural Machine Translation Attention and Confidence

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

$$CDP = \frac{1}{J}\sum_j \log\left(1 + \left(\sum_i \propto_{ji}\right)^2\right)$$

$$AP_{out} = -\frac{1}{I}\sum_i\sum_j \propto_{ji} \cdot \log \propto_{ji}$$

$$AP_{in} = -\frac{1}{I}\sum_j\sum_i \propto_{ij} \cdot \log \propto_{ij}$$

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

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

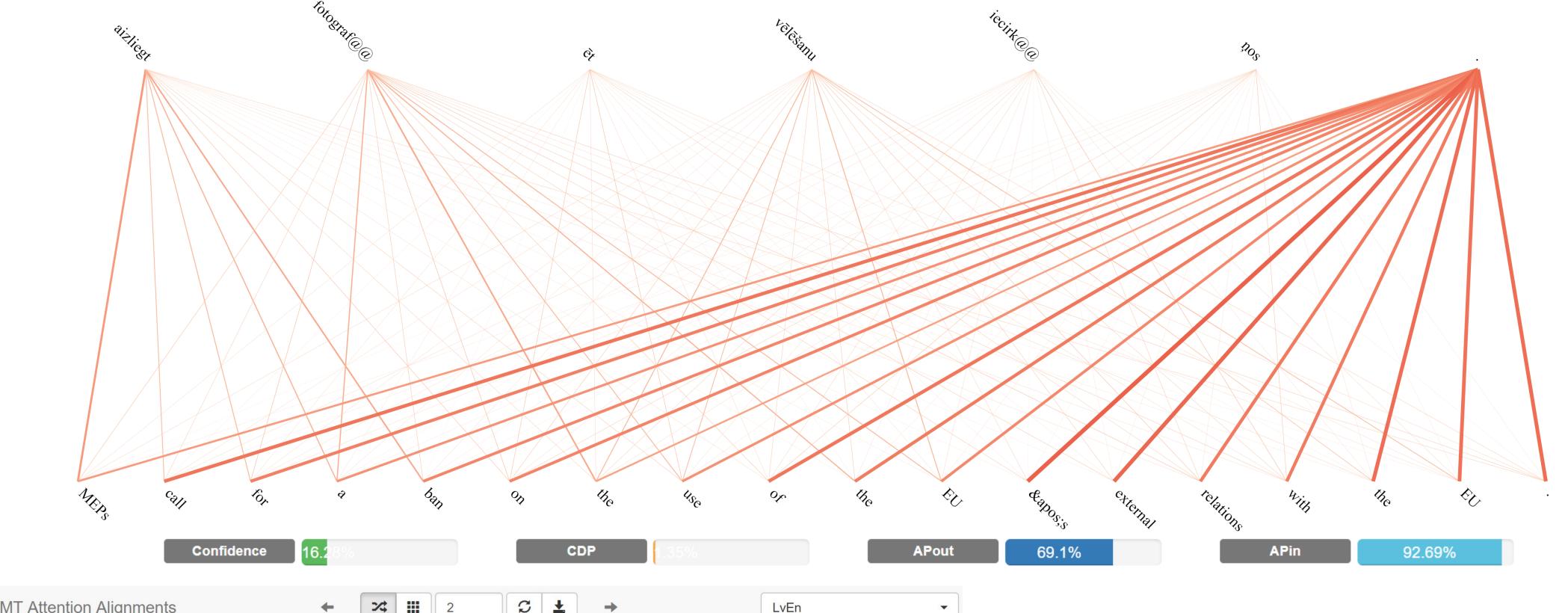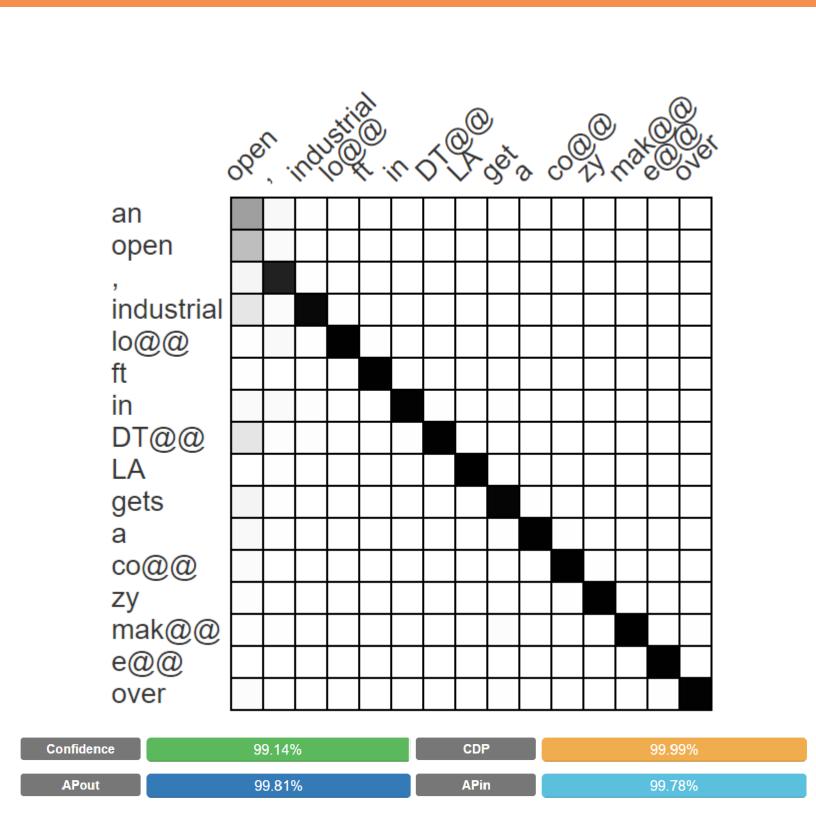{J, l} - source sentence length; i - output token index; j - input token index; α - attention weight; X - confidence score; C - constant 1 for CDP, 0.05 for others
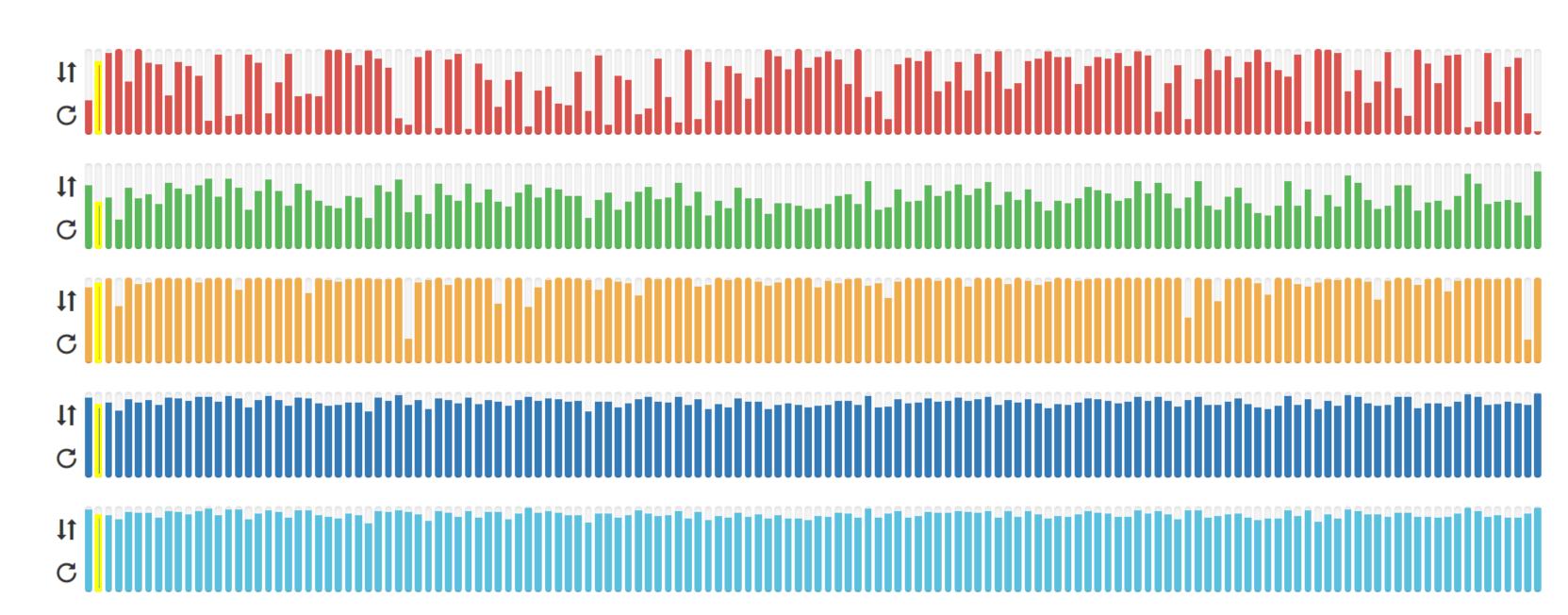
Terminal Visualisations

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

| Confidence | 16.2 | | CDP | | | APout | 69.1% | | APin | 92.69% |

| Confidence | 99.14% | | CDP | 99.99% |
| APout | 99.81% | | APin | 99.78% |

NMT Attention Alignments   ← ⤬ ▦ 2 ↻ ⤓ →   LvEn ▾

Source: šonedēļ kāda lielveikala kāpņutelpā atrasts miris 28 gadus vecs pavārs , kurš nesen pārcēlies uz Sanfrancisko .

Translation: This week someone in the supermarket stepped up dead 28 years old, who has recently moved to San Francisco.

| Confidence | 55.23% | | CDP | 93.96% | | APout | 85.95% | | APin | 89.94% |

Features

Works with attention alignment data from
. Nematus
. Neural Monkey
. AmuNMT (fork github.com/barvins/amunmt)
. OpenNMT

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   Poster   Demo

ej.uz/nmt-github   ej.uz/nmt-poster   ej.uz/nmt-attention

Acknowledgements