

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

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

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

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

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

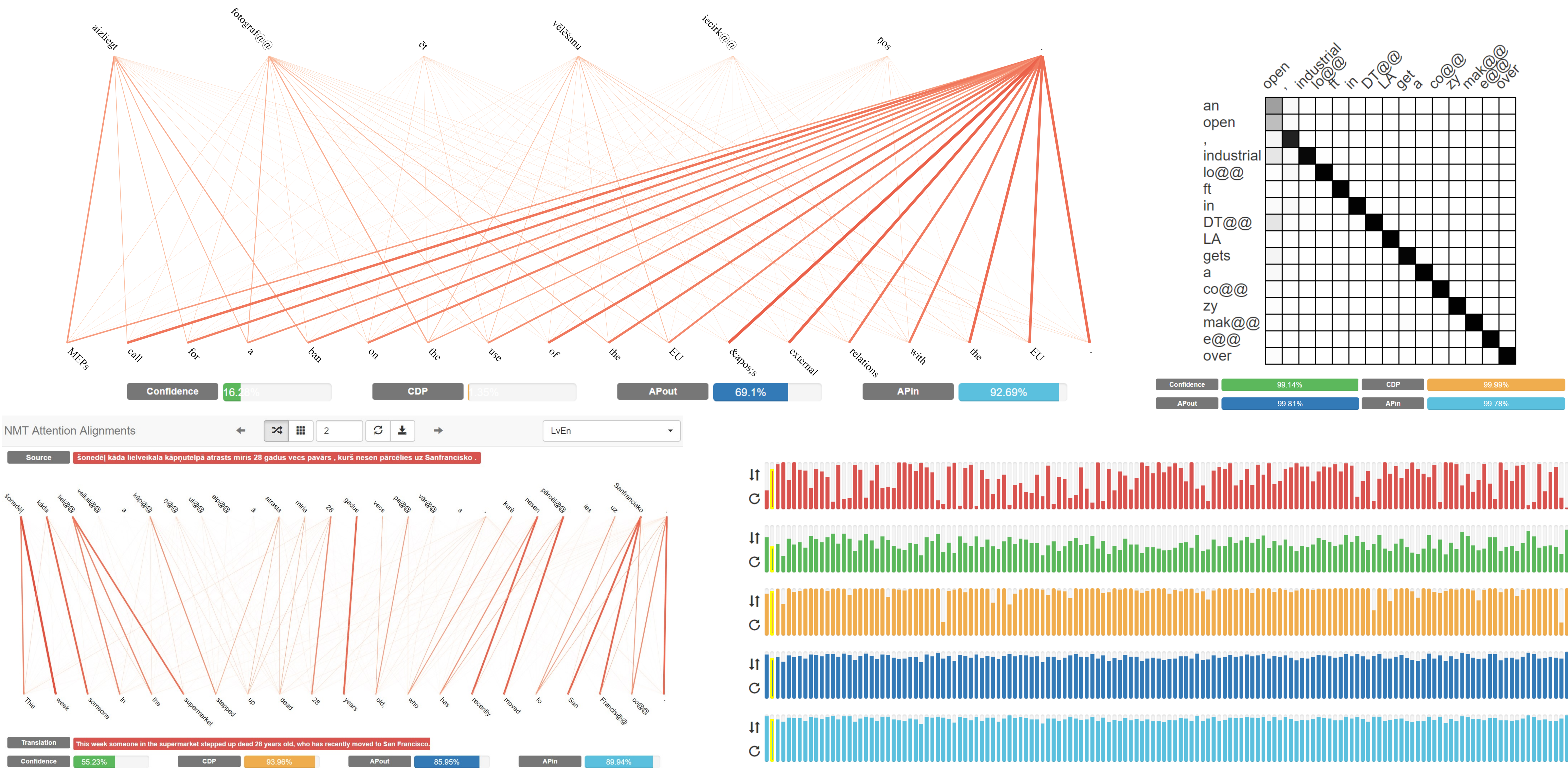
$$percentage = e^{-C(X^2)}$$

$\{J, I\}$ - 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



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

Poster

Demo



ej.uz/nmt-github



ej.uz/nmt-poster



ej.uz/nmt-attention

Acknowledgements



This research was supported by the ICT COST Action IC1207 ParseME: Parsing and multi-word expressions - towards linguistic precision and computational efficiency in natural language processing, the grant H2020-ICT-2014-1-645442 (QT21) and Charles University Research Programme "Progres" Q18+Q48.