

# Using OpenNMT in a CAT environment

Panos Kanavos, Lexorama Ltd

OpenNMT Workshop Paris, March 2018



- Model Description
- Trados Plugin Description
- Using OpenNMT with Trados
- Conclusions
- Further work and ideas

- English to Greek Model trained with the following options:
  - 5 source features (case feature, lemma, POS, client, domain)
    - Sennrich, Rico & Haddow, Barry. (2016). Linguistic Input Features Improve Neural Machine Translation. 83-91. 10.18653/v1/W16-2209.
    - Kobus, Catherine & Crego, Josep & Senellart, Jean. (2017). Domain Control for Neural Machine Translation. 372-378. 10.26615/978-954-452-049-6\_049.
  - 1 target feature (case feature)
  - BRNN encoder type, 800 RNN size, 4 layers, 13 epochs
- Training corpus:
  - ~5 million sentences: Europarl, EMEA (medical), SETIMES (news), DGT, IT (private data), Marketing/Management (private data)
- TreeTagger server
- REST server with TreeTagger hook

# Trados Plugin Description



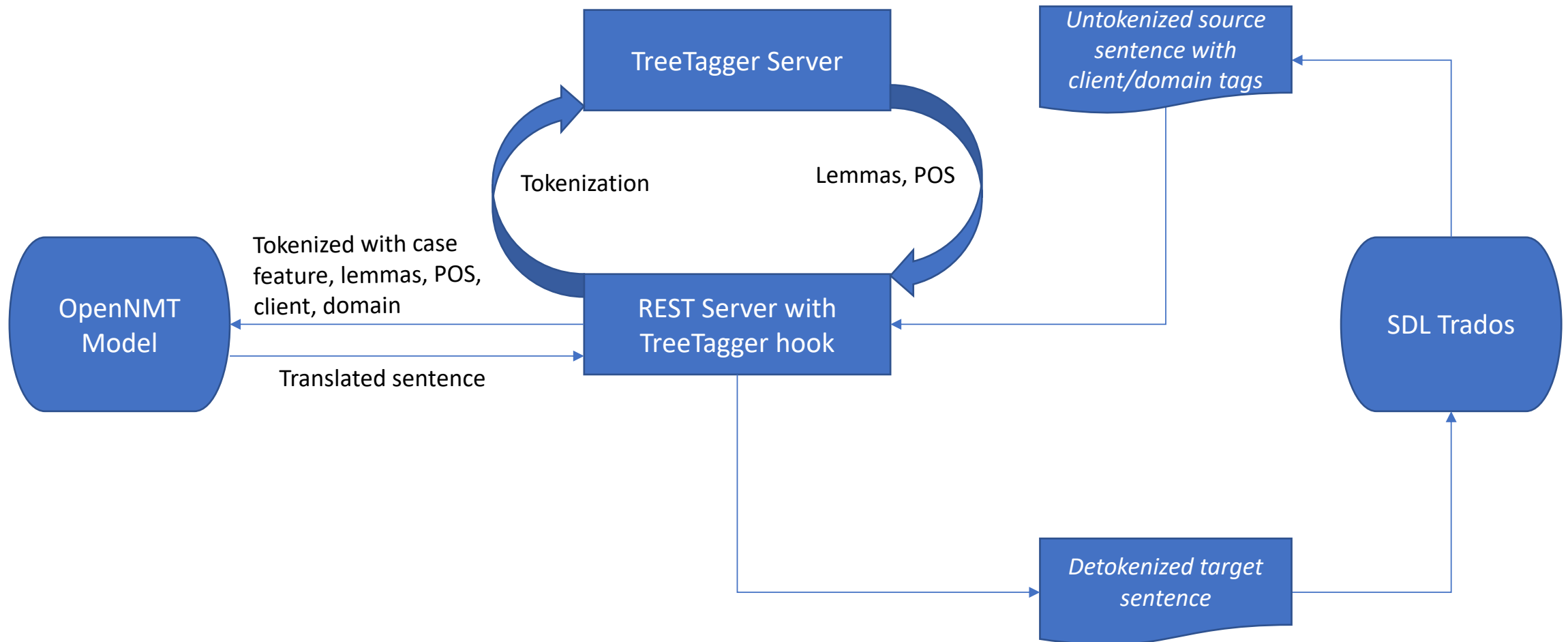
- A translation provider created with SDL Trados' APIs
  - Compatible and tested with Trados 2014/2017
- Implements latest end-user functionality from OpenNMT
  - Custom source features, e.g client and domain
- Best output with server-side tokenization (`-joiner_annotate`) and case handling (`-case_feature`)

Available at:

GitHub: <https://github.com/OpenNMT/Plugins/>

SDL AppStore: <https://appstore.sdl.com/language/app/opennmt/818/>

# Using OpenNMT with Trados



- Sample SW manual project (LibreOffice's small .odt manual)
- Style and terminology control with custom source features
  - From minor style adaptations to significant syntax changes and word reordering

*(10' live demonstration using Trados with OpenNMT)*

- Seamless integration
  - With any CAT tool
- Output consistency
  - Complex and “inconsistent” functions (e.g tokenization) performed in server
- Custom source features make a big difference
  - Reduces the need for multiple models



- TODO: Transfer internal tags in the translations
- Real-time adaptation with user translations?
- How much more could we fine-tune and control the model?
  - e.g. CLIENT-DIVISION-DEPARTMENT-DOMAIN-SUBDOMAIN

THANK YOU FOR YOUR ATTENTION!