

# OpenNMT for Human Translation

Terence Lewis, MyDutchPal Ltd

OpenNMT Workshop Paris, March 2018



# My four areas of focus as an LSP



- Access (or workflow)
- Handling OOV's
- Customization (specialization)
- Feedback to the NN (incremental training)



# Ways to access OpenNMT

- Connectors (plugins)
- Dedicated client software
- Web page
- Command line

## Our connectors

- memoQ (Jon Olds)
- SDL Trados (Jon Olds)
- CafeTran (Tom Hoar, Slate)
- OmegaT (Tom Hoar, Slate)

Options

X

Category

- Default resources
- Appearance
- Spelling and grammar
- Advanced lookup settings
- Keyboard shortcuts
- Locations
- Updates and news
- Machine translation**
- Terminology plugins
- TM plugins
- Miscellaneous

Description

Manage third-party machine translation plugins.

## Machine translation

Offer machine-translated results while working in the translation grid

- Only if there is no good TM match available
- Only if there is no exact TM match available
- Always

### Installed machine translation plugins

Plugin	Author	Preferred	Status
 MyDutchPal Plugin for memoQ	(C) MyDutchPal Ltd	<input checked="" type="checkbox"/>	Enabled
 AltLang MT Plugin	(C) Prompsit	<input type="checkbox"/>	Not configured
 CrossLang Gateway MT	Copyright 2013 CrossLang NV	<input type="checkbox"/>	Not configured
 GoogleMT machine translation plugin	(C) Kilgray 2009	<input type="checkbox"/>	Not configured

OK Apply Cancel Help

Onze oplossing	Our solution		Voor de A-Pier ...	1	For the A-Pier si...
<b>Brandmeldinstallatie A-Pier</b>	<b>Fire detection system A-Pier</b>	0%	Voor de A-Pier worden zes Sinteso brandmeldcentrales in drie staande Rittal kasten geleverd, conform de Schiphol standaard.	1	For the A-Pier six Sinteso fire alarm control centres are supplied in three standing Rittal cabinets, in accordance with the Schiphol standard.
		0%	De twee brandmeldcentrales in één Rittal kast zijn via het FC-net met elkaar verbonden.		The two fire alarm control centres in one Rittal cabinet are connected via the FC-net with each other.
		0%	De bedieningsunits voor de brandmeldinstallaties zijn in de brandmeldcentrale geplaatst.		Voor de A-Pier worden zes Sinteso brandmeldcentrales in drie staande Rittal kasten geleverd, conform de Schiphol standaard.
		0%	Op de melderlussen worden de automatische melders, de handmelders en de I/O-boxen aangesloten.		Voor de A-Pier worden zes Sinteso brandmeldcentrales in drie staande Rittal kasten geleverd, conform de Schiphol standaard.
		0%	De A-Pier wordt voorzien van multisensor melders (rookmelders) en handmelders.		For the A-Pier six Sinteso fire alarm control centres are supplied in three standing Rittal cabinets, in accordance with the Schiphol standard.



The screenshot shows the CafeTran Espresso 2017 software interface. The top menu bar includes 'File', 'Edit', 'View', 'Action', 'Translate', 'Filter', 'QA', 'Task', 'Resources', and 'Help'. Below the menu is a toolbar with icons for 'Resources', 'Machine Tr.', 'Project source', 'Project target', 'TM source', and 'TM target'. The main window is titled 'Translate' and contains two panes: 'Project source' and 'Project target'. The 'Project source' pane displays the text 'TenderNed.' and 'announced on TenderNed.' with a red box highlighting 'TenderNed.'. The 'Project target' pane displays the translated text 'GVB volgt voor deze aanbesteding een onderhandelingsprocedure met aankondiging, volgens deel 1, 2 en 4 van het ARN2016].'. Below the target text is a toolbar with icons for 'Next', 'Previous', 'Find', 'Replace', 'Copy', 'Paste', 'Delete', and 'Translate'. At the bottom of the interface are tabs for 'Slate Desktop' and 'ProjectTM', and a 'Matchboard' panel on the right.

MyDutchPal - Neural MT Gateway

File Edit Tools Help

Dutch-English  English-Dutch Custom Dictionary Phrase Table Translate Clear File Translator

De Service Integrator is verantwoordelijk voor Incident-monitoring en Incident-control. Dit betreft alle Incidenten zoals die worden gemeld of geconstateerd en/of bij beherende partijen worden geconstateerd. Als een Incident kan worden afgehandeld binnen gestelde afhandeltijden beperkt deze rol zich tot bewaking en rapportage. Onder van te voren afgesproken condities zal de service integrator actief escaliatiemanagement gaan uitvoeren of, waar nodig en afgesproken, het escaliatiemanagement overdragen aan NS I&O.

The Service integrator is responsible for Incident-monitoring and Incident-control. This relates to all Incidents such as those that are reported or recorded and/or at managing parties are recorded. If an Incident can be handled within stated time limits the service integrator will carry out actively escalation management or, where necessary and agreed, the service system will transfer the escalation management to NS I&O.

# Neural MT Gateway

- Takes:  
docx, odf, xlxs files (etc)
- Generates:  
original format + tmx file
- Translates tmx file directly

The <UNK> is installed by first fitting a <UNK> and a <UNK> to the <UNK> which is situated under the <UNK> on the left side of the upper <UNK> of the <UNK> located behind the <UNK>.

We can reduce OOV's by:

- Increasing the model's vocabulary
- Incremental training
- Using subword units (BPE, etc)
- Using custom compound splitter
- Using “–phrase\_table” option

fabrieksdirecteur = fabriek directeur  
(handled by compound splitter in intermediate  
server)

File Edit View Search Terminal Help

EMC-effecten ||| EMC\_effects  
EMC-eigenschappen ||| EMC\_properties  
EMC-eisen ||| EMC\_requirements  
EMC-emissie ||| EMC\_emission  
EMC-filter ||| EMC\_filter  
EMC-gecertificeerd ||| EMC-certified  
EMC-gedrag ||| EMC\_behaviour  
EMC-getest ||| EMC-tested  
EMC-gevoelige ||| EMC-sensitive  
EMC-goten ||| EMC\_ducts  
EMC-kabels ||| EMC\_cables  
EMC-kasten ||| EMC\_cabinets  
EMC-keuringen ||| EMC\_inspections  
EMC-maatregelen ||| EMC\_measures  
EMC-management ||| EMC\_management  
EMC-matrix ||| EMC\_matrix  
EMC-norm ||| EMC\_standard  
EMC-normen ||| EMC\_standards  
EMC-normset ||| set\_of\_EMC\_standards  
EMC-omgeving ||| EMC\_environment  
EMC-oogpunt ||| EMC\_perspective  
EMC-plan ||| EMC\_plan  
EMC-platform ||| EMC\_platform  
:■



“Oh, and here’s a list of the  
words we want translated  
our way...”

Src: De analyse van het product is betrekkelijk complex gezien de eigenschappen van de **verbindingen**.

Mod: The analysis of the product is relatively complex in view of the properties of the **compounds**.

Src: De **verbindingen** tussen lab en **fabriek** zijn slecht.

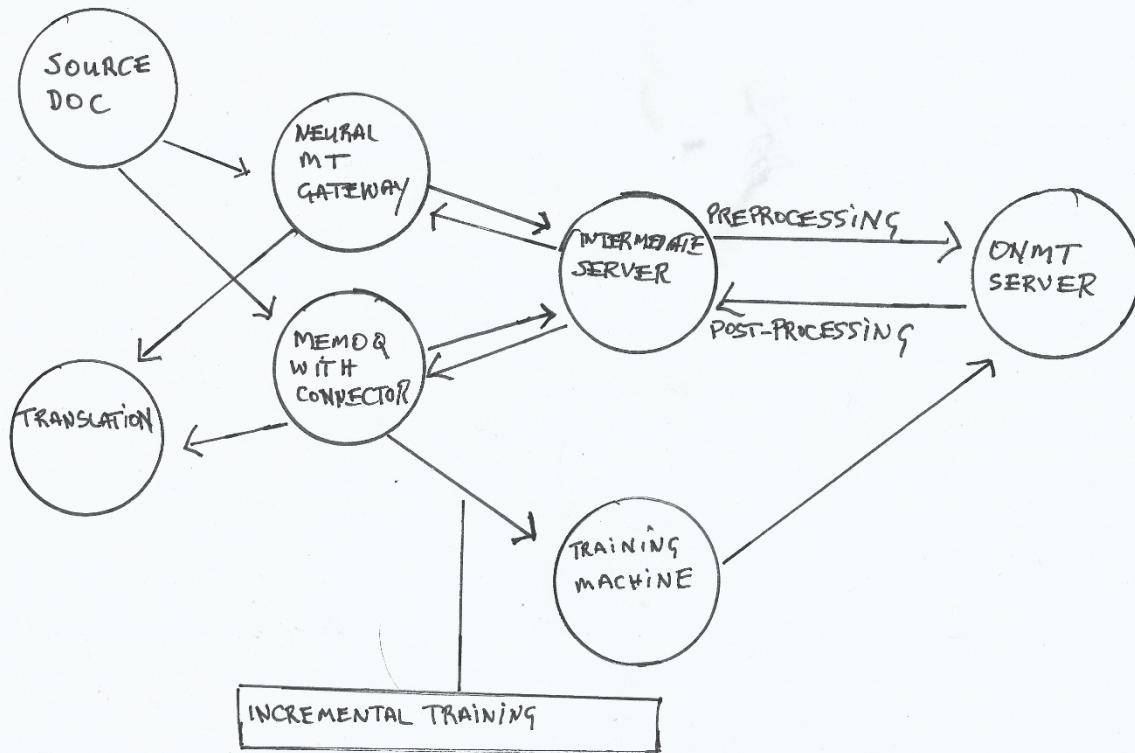
Mod: The **connections** between lab and **factory** are bad.

CustDic: The **connections** between lab and **plant** are bad.

Src: In dit geval zal een nieuwe **raad van bestuur** gekozen worden.

Mod: In this case a new **board of directors** will be chosen.

PT: In this case a new **board of bazziwicks** will be chosen.



For more information on our services see:

[www.mydutchpal.com](http://www.mydutchpal.com)

Thank you