

**FREE WEBINAR – go to <https://ncta.org>
to RSVP or to receive the recording**



**NORTHERN CALIFORNIA
TRANSLATORS ASSOCIATION**
A CHAPTER OF THE AMERICAN TRANSLATORS ASSOCIATION

WEBINAR



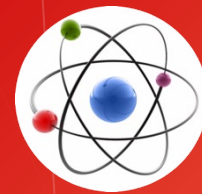
THERE'S A SCAM IN YOUR INBOX!

Presented by Carola F. Berger

Saturday, November 16 | Via Zoom | @NCTAorg



Can You Build Your Own Neural Machine Translation Engine?

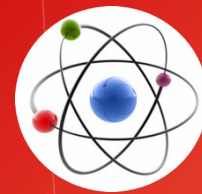


Carola F Berger, PhD, CT
CFB Scientific Translations LLC
<https://www.cfbtranslations.com>
<https://websitesforwords.com>



Can You Build Your Own Neural Machine Translation Engine? –

No.



Carola F Berger, PhD, CT
CFB Scientific Translations LLC
<https://www.cfbtranslations.com>
<https://websitesforwords.com>



Can You Build Your Own Neural Machine Translation Engine? – **No. But you can fine-tune an existing one with astonishing results!**



Carola F Berger, PhD, CT
CFB Scientific Translations LLC
<https://www.cfbtranslations.com>
<https://websitesforwords.com>



DEFINITIONS

■ TM – Translation Memory:

Stores pairs of source-target text segments for later retrieval

■ CAT-Tool – Computer-Assisted Translation Tool:

Software program with various functions to *help human translators*, e.g. terminology/glossary functionality, TM functionality, spell and quality assurance (QA) checks etc.
Not to be confused with

■ MT – Machine Translation:

A computer converts text in the source language into the target language *without human intervention* according to some *algorithm*



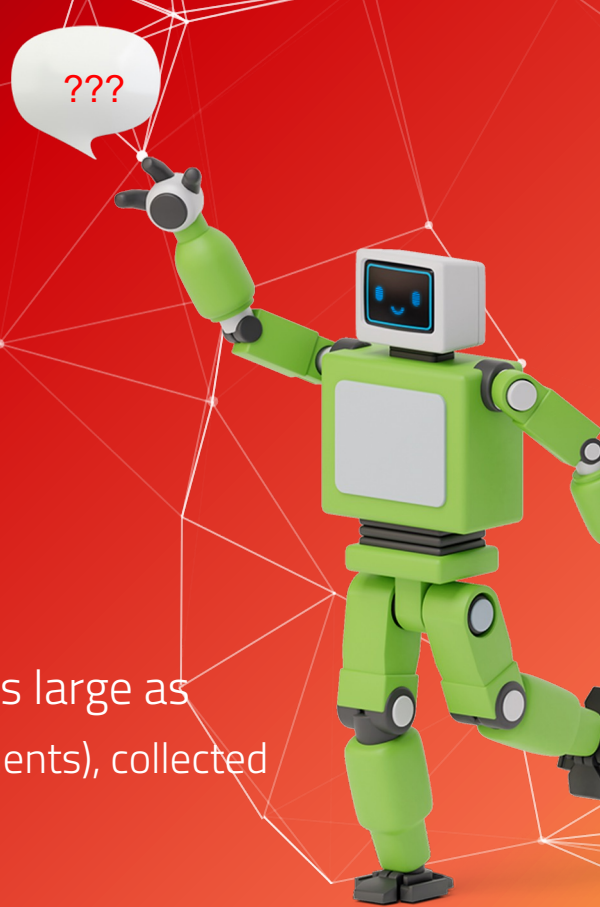
RATIONALE

- Software is free (except your CAT tool), no subscription required
- Preserves privacy, because the MT engine runs locally on your desktop, no interaction with any clouds or servers
- Finetuning with your own TM will lead to better results than using an out-of-the-box pre-trained model



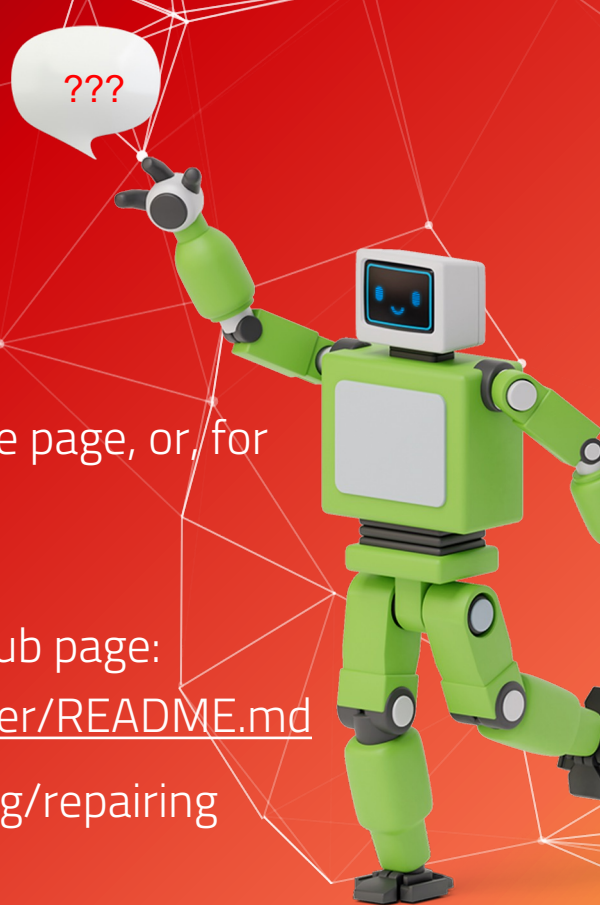
WHAT YOU WILL NEED

- A Windows computer
- Notepad (NOT WordPad)
- OPUS-CAT (free!)
- For best results, a CAT Tool such as:
 - RWS Trados 2017 and newer; or memoQ; or
 - others, including WordFast, OmegaT, XTM, Phrase
- A translation memory – as subject-matter specific and as large as possible (my patent-specific TM contains ~150.000 TUs (segments), collected over 10+ years)
- Olifant (optional, free!) – for editing/repairing your TM



WHAT YOU WILL NEED

- Download OPUS-CAT here:
<https://helsinki-nlp.github.io/OPUS-CAT/install>
- (Optional) Download plugin for your CAT tool on the same page, or, for RWS Trados 2021, 2022, 2024 visit the RWS App Store:
<https://appstore.rws.com/Plugin/11>
For RWS Trados 2017 and 2019 go to OPUS-CAT's GitHub page:
<https://github.com/Helsinki-NLP/OPUS-CAT/blob/master/README.md>
- (Optional) Download/install Olifant (useful tool for editing/repairing translation memories)
<https://okapi.sourceforge.net/Release/Olifant/ReadMe.htm>



STEP 1: EXPORT AND CHECK YOUR TRANSLATION MEMORY



EXPORT TM AS TMX

The screenshot displays the 'Translation Memories' (TM) interface. The 'Translation Memories' list on the left shows a tree structure with 'Patents' selected. A context menu is open over 'Patents', with 'Export' highlighted. The 'Search Details' pane shows filters for 'Filter Name' and 'Source Text'. The 'Field Values' pane shows a table of system fields.

Translation Memories

- General Articles
- IT
- Patents
- Patents
- Patents
- Technical
- test
- test2
- Language

Search Details

TM Filter: [Filter Name] [Source Text]

NOT Condition AND/OR

Use parentheses in this filter

Add... Insert... Edit... Delete

Field Values

Field	Value
Created by	
Created on	
Document structure	
Last modified by	
Last modified on	
Last used by	
Last used on	
Usage count	

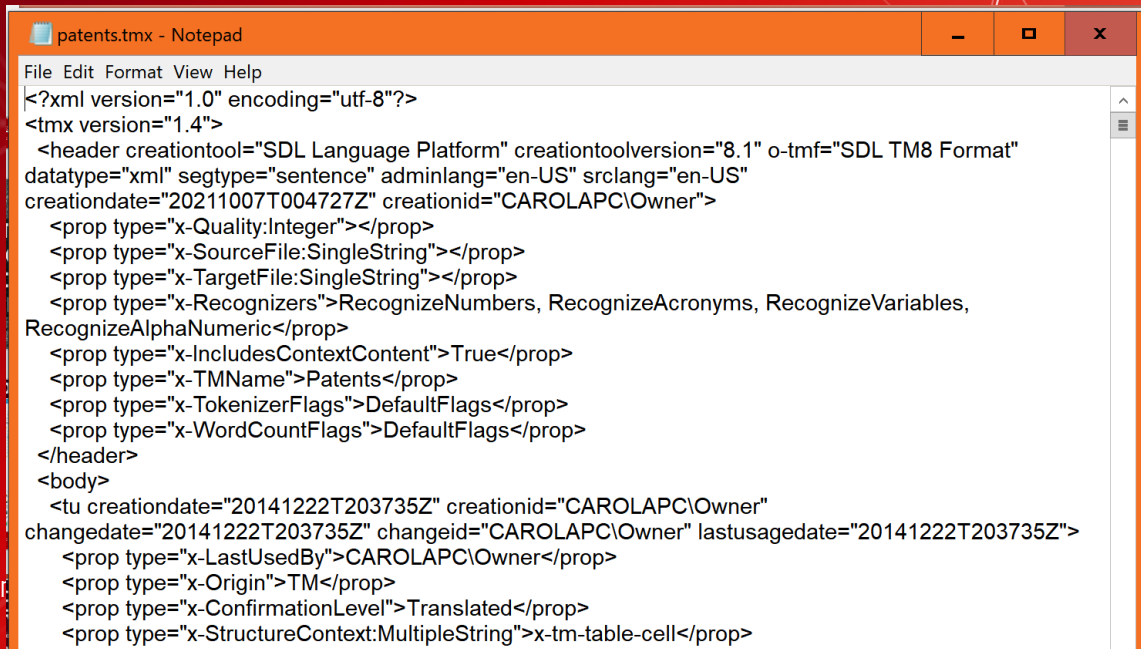


OPTION 1 NOTEPAD

Open (all file formats) the TMX.

Warning, large TMXs may take a while to open (~150.000 segments are 200+ MB).

Edit > Find > Search for "&#" type special characters

A screenshot of a Notepad window titled 'patents.tmx - Notepad'. The window displays XML code for a TMX file. The code includes a header section with metadata like creation tool, version, and date, followed by a body section containing a translation segment. The XML is well-formatted with indentation.

```
File Edit Format View Help
<?xml version="1.0" encoding="utf-8"?>
<tmx version="1.4">
  <header creationtool="SDL Language Platform" creationtoolversion="8.1" o-tmf="SDL TM8 Format"
datatype="xml" segtype="sentence" adminlang="en-US" srclang="en-US"
creationdate="20211007T004727Z" creationid="CAROLAPC\Owner">
    <prop type="x-Quality:Integer"></prop>
    <prop type="x-SourceFile:String"></prop>
    <prop type="x-TargetFile:String"></prop>
    <prop type="x-Recognizers">RecognizeNumbers, RecognizeAcronyms, RecognizeVariables,
RecognizeAlphaNumeric</prop>
    <prop type="x-IncludesContextContent">True</prop>
    <prop type="x-TMName">Patents</prop>
    <prop type="x-TokensizerFlags">DefaultFlags</prop>
    <prop type="x-WordCountFlags">DefaultFlags</prop>
  </header>
  <body>
    <tu creationdate="20141222T203735Z" creationid="CAROLAPC\Owner"
changedate="20141222T203735Z" changeid="CAROLAPC\Owner" lastusedate="20141222T203735Z">
      <prop type="x-LastUsedBy">CAROLAPC\Owner</prop>
      <prop type="x-Origin">TM</prop>
      <prop type="x-ConfirmationLevel">Translated</prop>
      <prop type="x-StructureContext:MultipleString">x-tm-table-cell</prop>
```



Find

Find what: &#

Find Next

Cancel

Direction

☐ Up ☒ Down

☐ Match case

☐ Wrap around

<prop type="x-ConfirmationLevel">Translated</prop>

</prop>

width of the apparatus is on the order of 10

MHz.

</t

<tu

<

Größe

</t

</tu>

<tu creationdate="20151026T163717Z" creationid="CAROLAPC\Owner"

changedate="20151026T163717Z" changeid="CAROLAPC\Owner" lastusedate="20151026T163717Z">

<prop type="x-LastUsedBy">CAROLAPC\Owner</prop>

<prop type="x-Origin">TM</prop>

<prop type="x-ConfirmationLevel">Translated</prop>

<tuv xml:lang="en-US">

<seg>a transimpedance amplifier; and</seg>

</tuv>

<tuv xml:lang="de-DE">

<seg>einen Transimpedanzverstärker; und</seg>

</tuv>

</tu>

<tu creationdate="20151026T163717Z" creationid="CAROLAPC\Owner"

changedate="20151026T163717Z" changeid="CAROLAPC\Owner" lastusedate="20151026T163717Z">

<prop type="x-LastUsedBy">CAROLAPC\Owner</prop>

<prop type="x-Origin">TM</prop>

<prop type="x-ConfirmationLevel">Translated</prop>

<tuv xml:lang="en-US">

<seg>Kelvin sensing switches electrically connected to an output of the transimpedance amplifier,

wherein at least one of the Kelvin sensing switches comprises a Tjunction switch network.</seg>

</tuv>

OPTION 2 OLIFANT

- Install according to:
<https://okapi.sourceforge.net/Release/Olifant/ReadMe.htm>
- Open TMX after removing special characters in option 1
- If further errors persist, Olifant will show you an error log



Log - Olifant

HelpClearSave As...StopKeep on topClose

```
charCount, EntityType& entityType)
    at System.Xml.XmlTextReaderImpl.ParseText(Int32& startPos, Int32& endPos,
Int32& outOrChars)
    at System.Xml.XmlTextReaderImpl.ParseText()
    at System.Xml.XmlTextReaderImpl.ParseElementContent()
    at System.Xml.XmlReader.WriteNode(XmlTextWriter xtw, Boolean defattr)
    at System.Xml.XmlReader.ReadInnerXml()
    at Olifant.TMXReader.ProcessSeg()
    at Olifant.TMXReader.ReadItem()
Error: The last entry read was the entry number 2240.
Number of entries read = 2240 (Added = 2240)
Duration = 00:00:00.7970837
--- 21-Oct-2024 16:03:29 ----- End Task
Error count = 2
===== End Process
```

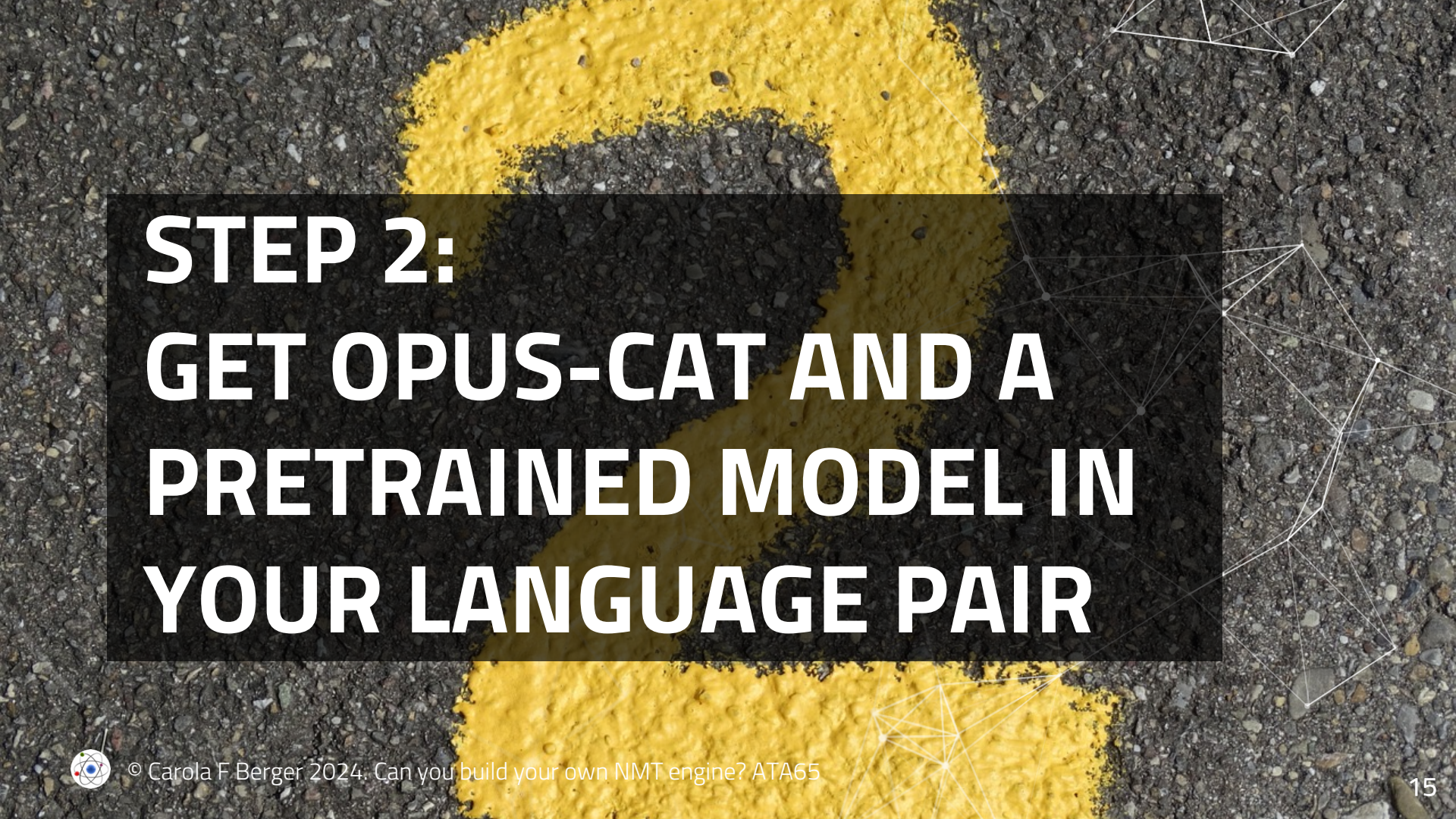
7	☐	<bpt i="1" type="pt23" x="1" />A head section according to Claim 1, wherein another bristle bundle (...)	<bpt i="1" type="pt23" x="1" />Kopfteil nach Anspruch 1, wobei ein weiteres Borstenbündel (18) in d...
8	☐	A head section (14, 50) according to one of the preceding Claims, wherein the supporting section has ...	Kopfteil (14, 50) nach einem der vorangehenden Ansprüche, wobei das stützende Teil eine kreisfömi...
9	☐	Start translation here:	Start translation here:
10	☐	Record ID 83354433	Datensatz-ID 83354433
11	☐	Page <bpt i="1" type="x-TextPlaceholderph3" x="1" />18<bpt i="1" /> of <bpt i="2" type="x-TextPla...	Seite <bpt i="1" type="x-TextPlaceholderph3" x="1" />18<bpt i="1" /> von <bpt i="2" type="x-TextPla...
12	☐	HEAT EXCHANGER	Wärmetauscher
14	☐	Field	Gebiet
15	☐	The present disclosure relates to a heat exchanger and method for operation of a heat exchanger.	Diese Offenbarung betrifft einen Wärmetauscher und ein Verfahren zum Betrieb eines Wärmetauschers.
16	☐	Background and Summary	Hintergrund und Darstellung
17	☐	Heat exchanger designs, such as automotive heat exchangers, may utilize rolled and/or folded fins to	Wärmetauscher-Ausführungen, wie Wärmetauscher für Kraftfahrzeuge, können mithilfe von gewalzte...

1 / 2240

Total = 2240

Filter is OFF



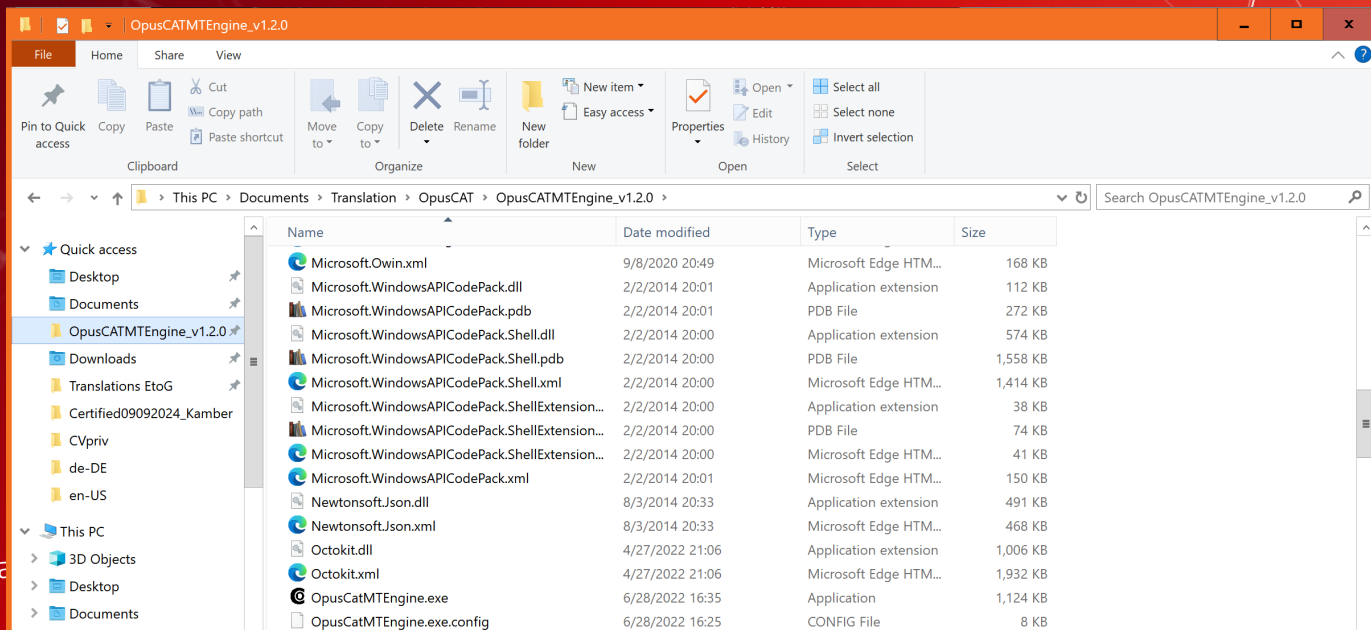


STEP 2: GET OPUS-CAT AND A PRETRAINED MODEL IN YOUR LANGUAGE PAIR

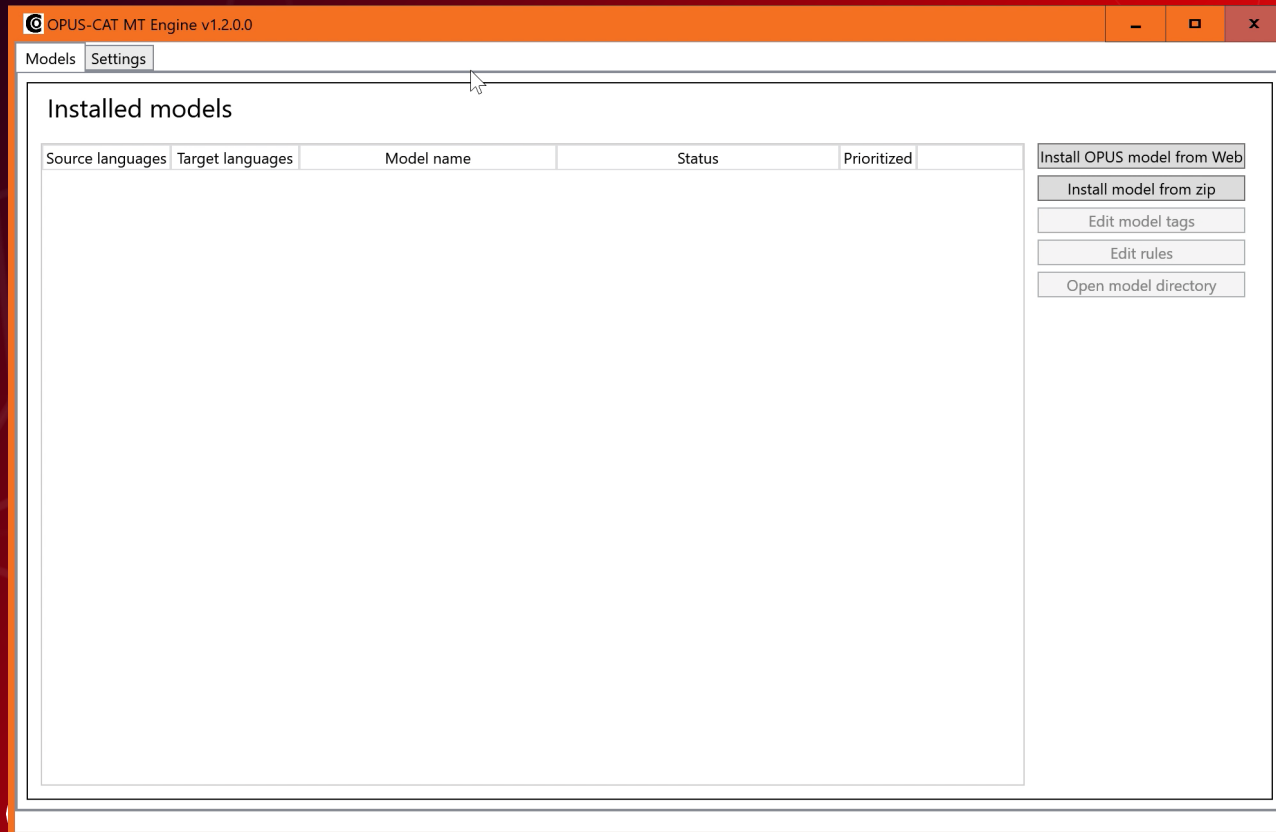


GET OPUS-CAT AND A PRETRAINED MODEL

- Download from:
<https://helsinki-nlp.github.io/OPUS-CAT/install>
- Unzip zip file
- Run "OpusCatMTEngine.exe" in OpusCAT folder



GET OPUS-CAT AND A PRETRAINED MODEL



GET OPUS-CAT AND A PRETRAINED MODEL

OPUS-CAT MT Engine v1.2.0.0

Models Settings Online models ✕

Fetching list of online models, please wait...

Source filter Target filter Model name filter

Ger

☒ OPUS-MT models ☒ Newest only ☐ Models with errors
☒ Tatoeba models ☒ Bilingual models ☐ Transformer-big/tiny models
☐ Multilingual models (usually lower quality than bilingual models, use with care)

Source languages	Target languages	Model name	Installation progress	Model origin
Afrikaans	German	opus-2021-02-18		Tatoeba-MT-models/
Belarusian	German	opus-2020-09-10		Tatoeba-MT-models/
Bulgarian	German	opus-2021-02-18		Tatoeba-MT-models/
Catalan	German	opus-2021-02-19		Tatoeba-MT-models/
Czech	German	opus-2021-02-19		Tatoeba-MT-models/
Danish	German	opus-2021-02-18		Tatoeba-MT-models/
Modern Greek (1	German	opus-2021-02-19		Tatoeba-MT-models/
English	German	opus-2021-02-22		Tatoeba-MT-models/
English	German	opus+bt-2021-04-13		Tatoeba-MT-models/
English	Low German	opus+bt-2021-04-14		Tatoeba-MT-models/
Esperanto	German	opus-2021-02-18		Tatoeba-MT-models/
Estonian	German	opus-2021-02-18		Tatoeba-MT-models/
Basque	German	opus-2021-02-19		Tatoeba-MT-models/

install model locally

Total amount of filtered models: 28



STEP 3: FINE-TUNE WITH YOUR TMX





STEP 4: ADD PLUGIN TO YOUR CAT- TOOL AND TRANSLATE HAPPILY EVER AFTER



INSTALL CAT-TOOL PLUGIN

- Download plugin for your CAT tool (with the exception of RWS Trados) by following the instructions:
<https://helsinki-nlp.github.io/OPUS-CAT/install>
- For RWS Trados 2021, 2022, 2024 visit the RWS App Store:
<https://appstore.rws.com/Plugin/11>
- For RWS Trados 2017 and 2019 go to OPUS-CAT's GitHub page:
<https://github.com/Helsinki-NLP/OPUS-CAT/blob/master/README.md>



RWS Trados

The screenshot shows the SDL Trados Studio interface with the 'Project Settings - Patents DE into EN' dialog box open. The 'OPUS-CAT settings' dialog box is also open, showing the 'General' tab. The 'Connection' tab is selected, and the 'Pre-order MT for next 5 segments (max 10)' checkbox is checked. The 'Model tag' is set to 'patents'. The 'Connection status' section indicates that MT models are available for all language pairs of the project, and a model with tag 'patents' for de-en is available. A link to 'OPUS-CAT MT online documentation' is provided. The 'Save' button is highlighted.

SDL Trados Studio - Patents DE into EN

File Home View and Save View Add-Ins Help

Go to last segment in project Bookmark settings Open For Review Open For Sign-off Explore Containing Folder Add Files Add Folders

Project Settings - Patents DE into EN

OPUS-CAT settings

General Connection

☒ Pre-order MT for next 5 segments (max 10)

☒ Show MT as suggestion origin

Model tag: patents

Connection status

MT models available for all the language pairs of the project
Model with tag patents for de-en is available.

[OPUS-CAT MT online documentation](#)

Save Cancel

File details layout

Layout

Type Iden.. Path

processing... \

processing... \

Enabled Languages Lookup Penalty Concordance Updat

Enabled Languages Lookup Penalty Concordance Updat

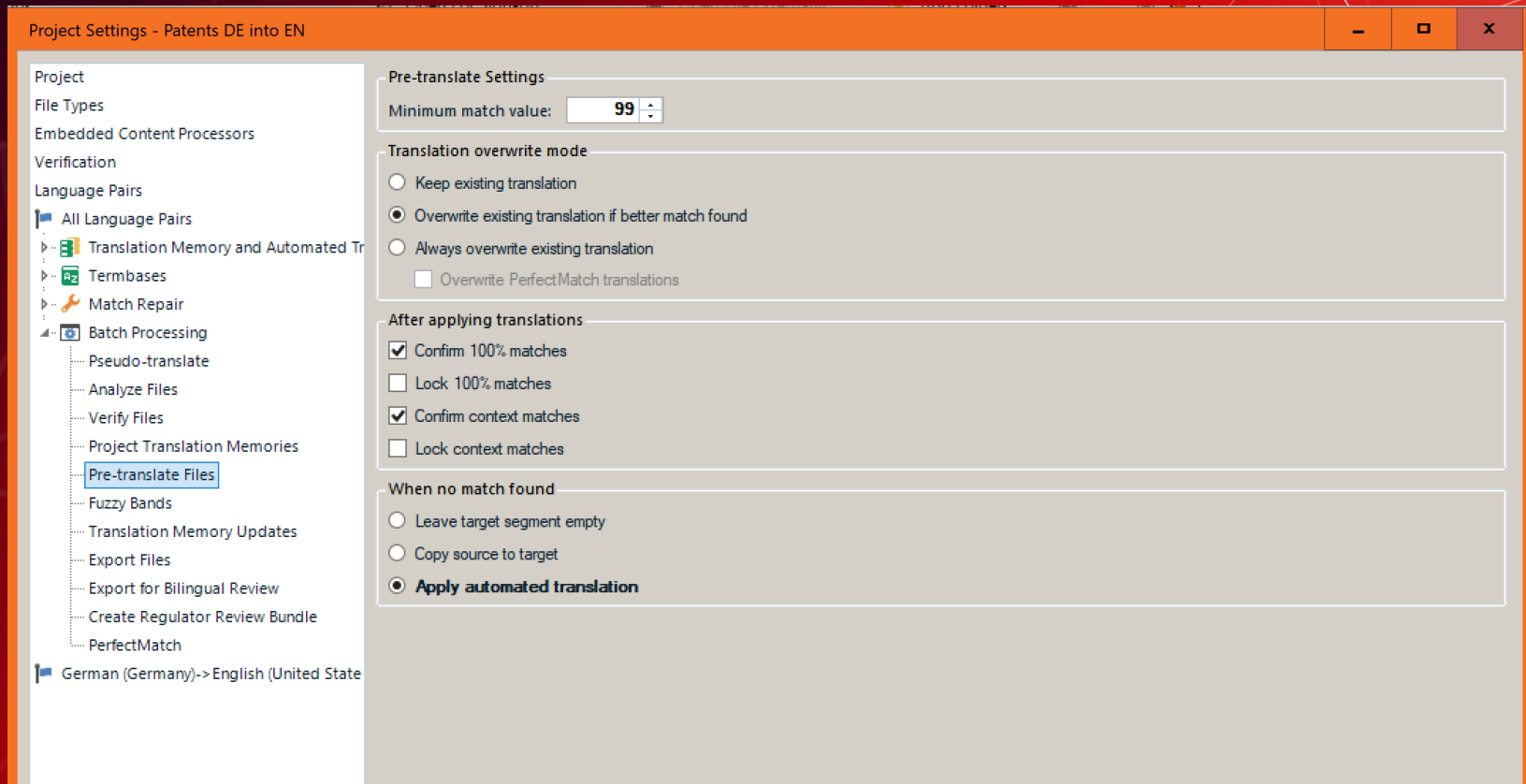
OK Cancel Help

TWO OPTIONS ON HOW TO USE MT

- Pre-translate using MT, then you are effectively post-editing the Opus-CAT MT output
- Or, you translate normally, using the Opus-CAT suggestions as you type, sort of like autocomplete while typing on a smartphone (machine-augmented translation)



POST-EDITING OPUS-CAT MT



Editor < Translation Results - Patents DE into EN_AnyTM Patents.OPUS-CAT Term Recognition

FAHRRADSCHALTWERK MIT Project Settings

	FAHRRADSCHALTWERK MIT EINSTELLANZEIGE	mit Patents
--	---------------------------------------	-------------

1	FAHRRADSCHALTWERK MIT EINSTELLANZEIGE	AT	DRIVING FEATURE WITH INSTRUCTION	mtSystem opus-cat
---	---------------------------------------	----	----------------------------------	----------------------

OPUS OPUS-CAT

[Translation Results - Patents DE int...](#)
[Fragment Matches - Patents DE int...](#)
[Concordance Search](#)
[Comments](#)
[TQAs \(0\)](#)
[Messages \(0\)](#)
[Term Recognition](#)
[Termbase Search](#)

FAHRRADSCHALTWERK MIT EINSTELLANZEIGE.docx.sdxliif [Translation]

Mac\Home\Documents\Studio 2017\Projects\Patents DE into EN\de-DE\FAHRRADSCHALTWERK MIT E

Mac\Home\Documents\Studio 2017\Projects\Patents DE into EN\de-DE\FAHRRADSCHALTWERK MI

1	FAHRRADSCHALTWERK MIT EINSTELLANZEIGE			
2	1.	100%	1.	P
3	Schaltwerk (54) für ein Fahrrad (10), umfassend			P
4	- eine B-Knuckle-Baugruppe (56), welche zur Montage an einem Fahrradrahmen eingerichtet ist,			P
5	- eine P-Knuckle-Baugruppe,			P
6	- eine Parallelogramm-Baugruppe, über welche die B-Knuckle-Baugruppe (56) und die P-Knuckle-Baugruppe (58; 158; 258) beweglich miteinander verbunden sind,			P
7	- eine Käfigbaugruppe (60; 160; 260), in der eine Kettenleitrolle (68) und eine Kettenspannrolle (66) drehbar gehalten sind, wobei die Käfigbaugruppe (60; 160; 260) an der P-Knuckle-Baugruppe (58; 158; 258) um eine P-Achse drehbar gelagert ist, gekennzeichnet durch eine Drehwinkel-Erfassungseinrichtung (72-80; 172, 174, 276, 290), welche dafür eingerichtet ist, einen Drehwinkel (C) zwischen der P-Knuckle-Baugruppe (58; 158; 258) und Käfigbaugruppe (60; 160; 260) zu erfassen oder/und anzuzeigen.			P
8	2.	100%	2.	P
	Schaltwerk (54) nach Anspruch 1, dadurch gekennzeichnet, dass die			

ANY QUESTIONS?

You can reach me at:

- info@cfbtranslations.com
- <https://www.CFBtranslations.com>
- <https://www.websitesforwords.com>
- Slides and other material at:
<https://bit.ly/YourownNMT>



THANKS!

