ATA GLD Members in Europe
9th Annual Workshop
February 2-4, 2024

The New Kid on the Block: First Experiences with GenAI in Translation
Dipl.-Dolm. Johanna Klemm, JK Language Services Europe LTD,
ATA-certified EN-DE, Kenilworth, UK
About me

• Dipl.-Dolm. Johanna Klemm, JK Language Services Europe LTD
• ATA-certified EN-DE, Kenilworth, UK
• Freelance Translator and Conference Interpreter for over 20 years
• German (native), English, French, Portuguese
• Medical, marketing, business
• Several major technology shifts during my professional life – Internet, TM, dictation software, (N)MT and now AI?
Goals

• Basic concepts and questions
• Forms of AI use for translation
• Concrete examples for AI in a CAT tool
• Value of the human translator
• Discussion
To use AI or not to use AI?

By ESA/Hubble, CC BY 4.0,
https://commons.wikimedia.org/w/index.php?curid=8788068
(N)MT vs. GenAI

(N)MT
- Mature technology
- Purpose:
  - Transforms text between languages
  - Specializes in accurate language conversion
- Faster, uses fewer resources
- Can translate long documents
- Is static
- Is part of an environment designed for the translation workflow
- Generates professional deliverables (xliff files)

GenAI
- New technology
- Purpose:
  - For conversational interaction, all-purpose tool
  - Designed for versatile text generation (context-oriented)
- Slower, uses more processing power
- Has token limitations
- Can generate different translations/fine-tune text, is trainable
- Has a general-purpose environment/chat interface
- Generates only „simple“ text or tables

Developed in conversation with the organizers and participants of the techforwords AI Bootcamp January
Uses of GenAI in translation

• Support tasks
• AI for translation in novel environments
• AI embedded in a CAT tool
Support tasks (not in a CAT tool)

Examples:
- Prepare source text
- Research – definitions, explanations
- Identifying terminology, glossaries
- Alternatives, synonyms, metaphors
- Fine-tuning translation

To do:
- Explore different options, ChatGPT/Copilot (Bing Chat)/Bard
- Learn how to write prompts or get prompts
- ChatGPT Translate: How to Write Your Translation Prompt (instantaiprompt.com)
- 2304.02182.pdf (arxiv.org)
- Try out novel tools, e. g. https://www.deepl.com/write
Bessere Texte im Handumdrehen
Schreiben Sie klar, präzise und fehlerfrei

The research organisation sends the encrypted audio file of the focus group to a transcriber via a secure online portal. This person will transcribe the interviews. A pseudonym will be used instead of your real name on the transcripts. Once the transcriber has returned the transcripts to the research organisation, they will delete the audio files.
AI for translation (and other tasks) in novel environments

• Captioning software (CaptionHub, Sonix)
  • AI voice recognition (creating the transcript and captions)
  • Manual review of source captions
  • MT or AI translation of captions
  • Manual review of translation

• To do:
  • Explore these new environments
  • Explore possibilities for most productive way of working (import/export)
AI embedded in a CAT environment

• Direct access to ChatGPT from a CAT tool
  • Definition/Explanation
  • Translation
  • Alternatives/Rephrasing
  • Spell-check/Proof-reading
  • Etc.

• Options to overcome shortcomings of machine translation
  • Easier trainability
  • Alternatives
  • Built-in proofreader
  • Context/cohesion

To do:
• Explore AI access in traditional CAT tools
• Explore new options:
  • Additional Tools: CoTranslatorAI, Custom.MT
  • New CATs: Wordscope, Bureau Works
Example: Wordscope
Bureau Works

• Localization platform
• Online tool
• Features:
  • Translation management (TMs, glossaries, tasks)
  • Automated project management
  • BWX AI
  • Connectors/Integrations
Example: Bureau Works AI

- Combines glossaries, TMs, MT and ChatGPT
  - Translate
  - Alternative Suggestion
  - Proofread
  - Check Smells
  - Fix Tags

https://support.bureauworks.com/hc/en-us/articles/18130511494797-All-about-our-AI-integration
https://www.loom.com/share/9eb2dbb2ea5746f7a4fdba2956a79c4 (Gabriel Fairman)
Example: Bureau Works

• Easier trainability
• Alternatives
• Built-in proofeditor
• Overcome cohesion problems
Use case: informed consent for research

What is the aim of this research?
We want to investigate to most relevant, feasible and effective ways to improve:
• adherence to public health messaging to COVID-19 and
• public health understanding and trust during the COVID-19 response.

What can I expect from taking part in this research study?
The field agency will invite you to take part in this study. They will also invite you to be part of a focus group. They will discuss the study with you and ask for your verbal consent to take part. If you agree to take part, they will then co-ordinate and establish a time that suits you to take part in the focus group.
You will be asked to complete a consent form and email it back directly to the research team.
This focus group will take approximately one hour and will involve a conversation with other people, using the teleconference software, Microsoft Teams. Interviews will be audio-recorded using Microsoft Voice recorder.
The focus group will be an informal conversation on your experiences and understanding of COVID-19, with topics including:
• how you and your family have coped during the pandemic,
• which sources you rely on and trust to gain information on COVID-19, and
• your views and opinions on responses to the pandemic.
You will receive €40 for your participation in the form of a pre-paid card.
If you wish to stop at any point during the focus group you may do so without giving an explanation.

Legal requirements:
• Client consent
• Terms of Use of AI provider
• Personal data

Translational/textual requirements:
• Mechanical issues:
  • Machine-readable, monolingual text
  • Cut-off sentences
  • Tags
• Content issues:
  • Purpose (informative, not motivational, formal etc.)
  • Shift in audience
  • Cultural influence on topic/metaphors/way of presenting things
• Language issues:
  • Language not neutral
  • Incorrect writing (complexity)
  • Form of address
  • Formatting conventions
Example: Bureau Works – easier trainability
Example: Bureau Works – alternative suggestions

1. What is the aim of this research?

2. We want to investigate to most relevant, feasible and effective ways to improve:
   • adherence to public health messaging to COVID-19 and

3. Suggested translation: Was ist das Ziel dieser Forschung?

Explanation: The original translation "Was ist das Ziel dieser Studie?" is grammatically correct and conveys the meaning accurately. However, the word "Studie" specifically refers to a scientific study or survey, whereas the term "Forschung" is a more general term that encompasses all types of research. Therefore, using "Forschung" instead of "Studie" provides a broader and more inclusive translation.
Example: Bureau Works – built-in proofreader

The field agency will invite you to take part in this study.
Example: Bureau Works – cohesion problems

21 The field agency will store your name and contact details in an encrypted file on a password-protected computer for the purposes of co-ordinating the focus group interviews.
Use case: challenges

Participants had in-person or phone call visits with the doctors once a week.

Suggested translation: Die Teilnehmer hatten einmal pro Woche persönliche oder telefonische Arztbesuche.

Explanation: In the original translation, the word "Besuche" was used, which means "visits." However, in this context, it is more appropriate to use the word "Arztbesuche," which specifically refers to visits to the doctors. This helps to clarify that the visits were related to medical appointments.
Use case: challenges

1. Participants had in-person or phone call visits with the doctors once a week.
## Use case: challenges

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2</strong></td>
<td><strong>We call ourselves Birds!</strong></td>
<td><strong>Wir nennen uns selbst Birds!</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>3</strong></td>
<td><strong>We work fast, grow fast, build fast and focus on impact.</strong></td>
<td><strong>Wir arbeiten schnell, wachsen schnell, bauen schnell und konzentrieren uns auf Wirkung.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>4</strong></td>
<td><strong>We're go-getters, industry leaders and roll-up-your-sleeves-and-make-it-happen kind of people.</strong></td>
<td><strong>Wir sind Macher, Branchenführer und packen an und setzen es um.</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Use case: challenges

2  We call ourselves Birds!

Wir sind die Birds!

Translation Smells

Übersetzungsfehler Die englische Phrase 'We call ourselves Birds!' wurde als 'Wir sind die Birds!' übersetzt. Eine genauere Übersetzung wäre 'Wir nennen uns Birds!'

Hallo, wir sind die Birds!

Translation Smells

Hinzufügung Von Wörtern Das Wort 'Hallo' wurde in der deutschen Übersetzung hinzugefügt, obwohl es im englischen Originaltext nicht vorhanden ist.
Use case: challenges

We work fast, grow fast, build fast and focus on impact.
Value of the human translator – what makes the "human touch" so valuable?

• The translator understands the intention of your communication and will make sure that you can achieve it, even if this means changing the target audience or the medium. In short, your translator will make sure that you look good to your target audience...

• The translator acts as the first person in the target country to see the communication in question. A translator is almost like a mini focus group - they will notice what is awkward, inappropriate or missing.

• The translator will help you smooth out any elements of your communication that are culturally awkward or inappropriate. Often this happens without the client even realizing.

• The translator will proofread your text like no other proofreader. They need to really understand the text to be able to translate it, so anything unclear (or omitted) will be picked up - helping you to improve your communication in the source language.
Thank you!

And special thanks to Karen Leube, Jaqueline Jugenheimer and André Pires da Silva for their encouragement!
Further reading

• https://arxiv.org/abs/2302.09210v1
• https://inten.to/blog/gpt-3-translation-capabilities/
Back-up slide:
Full list of factors to consider for MT

- Legal requirements:
  - Client consent
  - Terms of Use of MT provider
  - Personal data
  - Know how and where your use of MT is recorded

- Technical requirements:
  - Selection of appropriate MT engine
    - Language pair/locale?
    - Option to train?
    - Fluency of output?
  - Integration into the tool of your choice

- Translational/textual requirements:
  - Mechanical issues:
    - Machine-readable, monolingual text
    - Cut-off sentences
    - Tags
  - Content issues:
    - Purpose (informative, not motivational, formal etc.)
    - Shift in audience
    - Cultural influence on topic/metaphors/way of presenting things
  - Language issues:
    - Language not neutral
    - Incorrect writing (complexity)
    - Form of address
    - Formatting conventions
Back-up slide: Clarifications/terms

• **Artificial intelligence** (AI) is the intelligence of machines or software, as opposed to the intelligence of humans or other animals. ([https://en.wikipedia.org/wiki/Artificial_intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence))

• **Generative artificial intelligence** (generative AI, GAI, or GenAI) is artificial intelligence capable of generating text, images, or other data using generative models, often in response to prompts. ([https://en.wikipedia.org/wiki/Generative_artificial_intelligence](https://en.wikipedia.org/wiki/Generative_artificial_intelligence))
• A **large language model (LLM)** is a language model notable for its ability to achieve general-purpose language generation. LLMs acquire these abilities by learning statistical relationships from text documents (…).
  (https://en.wikipedia.org/wiki/Large_language_model)

• **Machine translation** is use of either rule-based or probabilistic (i.e. statistical and, most recently, neural network-based) machine learning approaches to translation of text or speech from one language to another, including the contextual, idiomatic and pragmatic nuances of both languages. (https://en.wikipedia.org/wiki/Machine_translation)
Back-up slide:
Shortcomings of Machine Translation

• Content
  • Mistranslations
  • Convoluted wording
  • Added/missing information

  ➢ read the text aloud
  ➢ check the translation length (manually or with QA)

• Mechanical errors
  • Glitches
  • Tags
  • Formatting conventions

  ➢ QA
  ➢ QA
  ➢ regular expressions

• Stylistic errors
  • Repetitions
  • Sentence structure too literal
  • Inconsistencies
  • Terminology
  • Cohesion ("it")
  • Form of address

  ➢ terminology functions, filtering
  ➢ check the translation length (manually or with QA)
  ➢ terminology functions, filtering
  ➢ terminology functions
  ➢ filter for “It”
  ➢ filter for appropriate words (“Du”)

• Tracking of use of MT: you need to know where this information is stored
  ➢ know your CAT tool