IATIS Training Event on Neural Machine Translation
Chat Protocol

(Morning Session, first Zoom meeting)

10:42:08 Organiser:
Audience: you are welcome to post your questions for our speakers in this chat at any time. These questions will be discussed in the Q&A session after each talk.

(After first talk by Josef van Genabith)

10:57:27 Participant:
(To Josef) Thank you for this very interesting lecture!

10:57:49 Participant:
(To Josef) Wonderful. Thank you so much.

10:57:54 Participant:
(To Josef) How is it possible to create training data for low-sourced languages that NMT can benefit from?

10:58:01 Participant:
(To Josef) Awesome lecture Thank you

10:58:07 Participant:
(To Josef) Thank you for this brilliant lecture.

10:58:08 Participant:
(To Josef) Thank you very much! This was an amazing presentation!

10:58:22 Participant:
(To Josef) Thank you very much!

10:58:47 Participant:
(To Josef) Thank you very much!

10:58:55 Participant:
(To Josef) Is there a limit for the context relevant translations that the current models can produce?

10:59:12 Participant:
(To Josef) Great lecture, thank you!

10:59:17 Participant:
(To Josef) This was a wonderful and very clear overview, thank you!

10:59:36 Participant:
(To Josef) Where does the pragmatics fit in MT and NMT systems because I can see that these systems are based on coding/encoding models rather than inferential models?

11:01:59 Participant:
(To Josef) Thanks for the very interesting presentation! Could you share the slides or the sources you recommended afterwards?

11:02:00 Participant:
(To Josef) Yes, fantastic lecture. Very informative, thanks! Can you perhaps talk a bit about the 'new'

development recently announced by TAUS regarding the use of a few clicks to train an MT engine?

11:02:28 Participant:
(To Josef) Thanks for such a clear and wonderful presentation

11:02:39 Participant:
(To Josef) thanks a lot!

11:03:15 Participant:
(To Josef) Really interesting and stimulating presentation. Thanks a lot!!

11:06:15 Josef van Genabith:
Yes, very relevant to human translation professionals to tailor MT to their specific demands. Other offers in that space includes Google AutoML where you fine tune an existing model to your data. Sth. also offered by DeepL and others.

11:20:08 Participant:
@Josef van Genabith: thanks for the response! Yes, I have been reading about Google AutoML too. But how accurate are these claims about "premium quality level of real-time translation where the training of the selected MT engine is done with a few clicks using the readily-available domain-specific datasets" (TAUS)? I mean, are professional translators likely to become "dummy engineers" too? I know, too general a question, but I wonder what the implications for professional translators will be in terms of evolving job profiles

11:30:39 Josef van Genabith:
Yes, a very interesting question! At Saarland Univ. the human translators on the Translation Science and Technology MA degree have a course in basic Python, two courses in using basic NLP (based on NLTK) and a Translation Technology course (TM, MT, terminology …). In the Translation Technology course one of the things they learn is use Google AutoML to finetune a generic Google model to Harry Potter data and then they evaluate the performance of the autotuned system on some other Harry Potter data. All the LT and the programming courses are tailored to the human translation professional trainees. The courses we teach to our Computer Science and NLP/CL students are very different: those students learn to build the systems. The translation students learn about the basic ideas and how to use the systems.

(after the second talk by Christian Eisold)

11:31:28 Josef van Genabith:
@Christian Eisold: nice talk! How did you say you detect over-editing?

11:36:47 Participant:
(To Christian) Can you elaborate a little more on how you deal with "personal preferences of the human post-editors" in order to homogenize the results? Is it simply the result that is given most often is regarded the best solution?

11:39:35 Participant:
@Josef van Genabith: Excellent and insightful approach to learning how to use G AutoML to finetune generic models: Harry Potter, above all! :). Will check out the MA program at Saarland and attempt to monitor the evolving nature of the engineering tasks increasingly enabled for and probably expected of professional translators in the MT loop.

11:40:02 Participant:
(To Christian) Thanks for this very interesting presentation! Regarding Retraining: How do you select the data for retraining to make sure that the quality is increased?

11:41:26 Participant:
(To Christian) I am curious to know how long the entire process takes, from the initial "my boss wants us to use MT" to go-life.

11:41:42 Participant:
(To Christian) Many thanks for a very clear and helpful presentation! A few Qs, please:
 1. When aligning the customer's data, and especially during the cleaning stage, do you have any mechanism to filter out localised (adapted) translations from literal translations, or do you use the Moses filtering features?
 2. Do you have experience using multiple translations for the same source segment, as well as multiple source segments for the same target segment in the training data? Any actual positives and negatives of using such data?
 3. Have you settled on a preferred CAT tool for error annotation and reporting, because you get slightly different info from each and they are slightly different in flexibility (am thinking of Studio v memoQ v Memsource, for instance…)
 4. Do you ever have MT evaluators work collaboratively rather than individually + inter-annotator agreement at the end?
 5. Interesting to see you include PE distance when Tracking Quality. Do you also track PE time?
 TY :)

11:41:51 Participant:
@Participant: Very interesting question too! I guess where that's where the inter-rater disagreement also enters the picture

11:42:12 Participant:
@Participant: That is exactly what I wanted to ask, too

11:42:38 Participant:
You are welcome, Participant ☺

11:47:08 Participant:
(To Christian) Is there a theoretically-grounded model of MQM or DQF or is it still largely based on practices that are not standardized yet?

11:49:43 Participant:
@Participant: is error annotation and reporting (of the kind performed with MT systems only) standard practice in CAT tools too? How standardized is the said practice from the LSP perespective, you reckon?

11:51:19 Organiser:
IATIS Training Event (afternoon sessions): Neural Machine Translation – Foundations, Applications and Implications
https://th-koeln.zoom.us/j/82692165063

Meeting-ID: 826 9216 5063
Kenncode: 578871

(Afternoon Session, second Zoom meeting)

(third talk by Jean Nitzke and Anke Tardel)

12:23:21 Participant:
neural bubble

12:35:57 Organiser:
A Short Guide to Post-Editing: https://langsci-press.org/catalog/book/319

12:51:26 Participant:
(To Jean and Anke) In the case of subtitling are these only potential job profiles or real ones based on industry adverts? Have you done any research on this or not yet?

12:51:36 Participant:
(To Jean and Anke) I'm interested in post-editing. I found an interesting role on a job site and I met most of the skills listed for the role except for the requirement for 2 years of experience. How important is experience when applying for roles? Do you have any tips for getting a first job in post-editing?

12:51:58 Participant:
(To Jean and Anke) A more general question: how common is the use of ASR in subtitling nowadays?

12:53:05 Participant:
(To Jean and Anke) Many thanks for this excellent talk!

12:53:15 Participant:
(To Jean and Anke) Thank you!

12:53:21 Participant:
(To Jean and Anke) Thank you guys. Good introduction to the topic!

12:53:26 Participant:
(To Jean and Anke) Thank you very much!

12:54:08 Participant:
(To Jean and Anke) Thanks for those interesting insights and the book recommendation especially

12:56:47 Participant:
(To Jean and Anke) Many thanks, very useful!

12:59:32 Participant:
(To Jean and Anke) Thanks!

12:59:50 Participant:
(To Jean and Anke) Many thanks!

13:00:04 Participant:
(To Jean and Anke) Thanks for your interesting talk and suggestions

13:00:17 Participant:
(To Jean and Anke) Thank you a lot!

(fourth talk by Joss Moorkens)

13:21:53 Participant (on behalf of Joss Moorkens):
https://app.sli.do/

13:22:29 Participant (on behalf of Joss Moorkens):
https://app.sli.do/event/aff13dot/live/polls

13:34:10 Participant:
It sounds like a new tag in the TMX standard would give freelancers a little bit more agency, Joss -

something like <no_mt>. Curious to see how many folks would seek it and use it when exporting their TMs. ;)

13:37:21 Participant:
(although we can immediately imagine exceptions popping up: <no_mt_from_DARPA>, <no_MT_from_X-domain-player>, <no_MT_from_Y-person-I-don't-like>, <no_MT_from_X-country>, etc. ;))

13:38:20 Participant:
@Participant: very good proposition! How do you envisage tag use across cloud-based platforms with TM export settings disabled, and the like?

13:40:52 Organiser:
On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?
https://dl.acm.org/doi/10.1145/3442188.3445922

13:41:20 Participant:
Thanks @Organiser!

13:42:02 Participant:
Thanks a lot, @Organiser!

13:43:15 Participant:
(to Joss) Brilliant as always, Joss! Thanks so much!!

13:43:35 Participant:
(to Joss) Thank you very much!

13:44:30 Participant:
yeah, settings are commonly disabled, right?

13:45:04 Participant:
(to Joss) Great ! Thanks

13:45:17 Participant:
@Participant, so true! We were chatting in the office about how freelancers at least used to have TMs to show for their troubles, but with some cloud-based tools, they don't even have that...

13:45:34 Participant:
(to Joss) I'm interested in the idea of end-users as stakeholders, how much power do they have?

13:46:01 Participant:
Yep, loss of agency and autonomy, yet again, Participant...

13:46:05 Participant:
But then the tag will be rendered a little bit useless when the translation is published and then crawled

13:47:14 Participant:
Thanks, Organiser

13:47:44 Participant:
But if they use MT it's often out of need not desire

13:48:05 Participant:
(to Joss) Misinformation is a big topic at the moment. You mentioned translations of user generated content on Twitter and how it doesn't make sense to use human translators. Do you think that social media platforms have a responsibility to use MT engines that avoid catastrophic translations errors that lead to the spread of misinformation? I ask because I've seen some pretty problematic and incorrect MT translations of user content.

13:49:52 Participant:
Great point, @Participant. The AI Incident Database has a few examples of translation issues, too, among other AI mishaps https://incidentdatabase.ai/?lang=en

13:50:27 Participant:
(to Joss) Thank you! It was really interesting.

13:50:29 Participant:
Just like Universaldoctor app for refugee-seeker communication

(fifth talk by Pilar Sánchez-Gijón and Olga Torres-Hostench)

14:14:18 Participant:
Can you please paste the links on chat?

14:14:28 Orgniser:
http://ntradumatica.uab.cat
https://cutt.ly/MutNMT_IATIS

14:14:42 Participant: Thank you

14:15:53 Olga Torres-Hostench:
cutt.ly/MutNMT_IATIS

14:17:25 Olga Torres-Hostench:
ntradumatica.uab.cat

14:18:54 Participant:
(To Pilar and Olga) Am I the only one not seeing the Upload option?

14:19:14 Participant:
(To Pilar and Olga) Same here (I can see public corpora but can't find an option to upload)

14:19:18 Participant:
You are possibly seeing the public corpora only, right?

14:19:34 Participant:
Demo version-related

14:19:59 Participant:
Yes, I can see only the public corpora.

14:20:17 Participant: User-related

14:20:39 Participant:
Do I have to be an expert user to upload?

14:20:52 Participant: It seems so

14:21:29 Olga Torres-Hostench:
Dear colleagues, yes, you need to be expert to upload. We suggest Teachers to be experts and students Beginners (if you only show MutNMT in a demo class)

14:22:38 Olga Torres-Hostench:
Write us to become Experts (pilar.sanchez.gijon@uab.cat)

14:23:20 Participant:
(To Pilar and Olga) It looks great. I am wondering how easy it is to install this at University? Installing new

software at our University is a pain, which is why we usually go for cloud-based options (we're using KantanMT right now).

14:23:38 Participant:
(To Pilar and Olga) Is the source code available to integrate other engines besides JoeyNMT

14:36:49 Olga Torres-Hostench:
Hi Participant (so nice to see you again ;-) I do confirm this is no a usual CAT Tool. For demo purposes in class, we suggest to use the UAB online platform. It is free and open for the Erasmus+ Project. Grenoble university has also an open platform (we'll shere the address in the website), but if you have engineers, funding for hardware and research purposes you can instal it in your university. Please feel free to use the UAB online free platform your clases!!!

14:37:13 Olga Torres-Hostench:
Ignore typos: install ---for your clases, etc. Thanks.

14:37:39 Participant:
(To Pilar and Olga) It's curious that source text is optional when comparing the translations.☺

14:38:30 Joss Moorkens:
I think only neural AEMs consider the source!

14:39:01 Organiser:
More Information on BLEU, chrF3, TER, etc. (and options for computing these scores):
https://colab.research.google.com/drive/19Mub2IJV6SoyuvvsyjSeaGiSQ1q433jD?usp=sharing

14:40:28 Olga Torres-Hostench:
You can download training materials Pilar suggested at the beginning and prepare a training class. You can use the recording of this session to guide you to prepare the class.

14:41:11 Olga Torres-Hostench:
This is an educational platform to give an overview of the possibilities of NMT

14:41:23 Participant:
(To Pilar and Olga) Is it possible to adapt existing engines, or only create a new engine (NN models) from scratch?

14:41:34 Participant:
(To Pilar and Olga) Impressive!! Thanks! I am wondering if MutNMT has an API to connect with a CAT tool? It could be interesting to show the Connection of CAT to MT as well

14:41:52 Joss Moorkens:
Thanks Pilar and Olga!

14:42:03 Participant:
(To Pilar and Olga) Great, many thanks, Olga! And very nice to "see" you here too!

14:42:07 Participant:
(To Pilar and Olga) Yes, thank you, very useful!

14:42:09 Olga Torres-Hostench:
Not yet an API (no funding to that now I am afraid;-(

14:42:12 Participant:
(To Pilar and Olga) Brilliant, thanks!

14:42:14 Participant:
(To Pilar and Olga) Many thanks!

14:42:19 Participant:
(To Pilar and Olga) Great job! Thanks a lot!

14:42:56 Participant:
(To Pilar and Olga) YEs, very accesible for both trainers and students!

14:44:03 Participant:
Thanks a lot, Pilar & Olga :)

14:44:31 Olga Torres-Hostench:
Please feel free to use https://ntradumatica.uab.cat/

14:44:35 Josef van Genabith:
(To Pilar and Olga) Nice! Is there a container-based version of the system? Do you have teaching material (slides etc.) for this?

14:47:08 Olga Torres-Hostench:
Teaching material: demo videos in the platform (see More on MUTNMT on the home page) and PowerPoint too (I am thinking now it is not uploaded it on the website. We'll do that soon! Please ask us the PowerPoint (olga.torres.hostench@uab.cat). We'll upload the PowerPoint soon!

14:49:01 Participant:
(To Pilar and Olga) Thanks a lot!

14:50:47 Participant:
(To Pilar and Olga) That's great

14:50:52 Participant:
(To Pilar and Olga) Thanks a lot

14:51:57 Participant:
(To Pilar and Olga) Great !!

(sixth talk by Claudia Wiesinger, Raluca Chereji and Justus Brockmann)

15:05:11 Josef van Genabith:
(To Claudia, Raluca and Justus) Great! Did you integrate the ASR in MateCAT and TTS into Trados?

15:16:05 Participant on behalf of Claudia, Raluca and Justus:
https://translatingintimesofcrisis.limesurvey.net/688765?lang=en

15:16:15 Claudia Wiesinger:
Thank you @Participant!

15:26:49 Participant:
(To Claudia, Raluca and Justus) good insights

15:26:52 Participant:
(To Claudia, Raluca and Justus) Thank you!

15:27:00 Participant on behalf of Claudia, Raluca and Justus:
By the way, the survey mentioned by Claudia is available in English
…https://translatingintimesofcrisis.limesurvey.net/688765?lang=en , as well as in German:
https://translatingintimesofcrisis.limesurvey.net/688765?lang=de

15:27:42 Participant:
(To Claudia, Raluca and Justus) Very interesting projects! Keep up the good work! Thank you all for the presentations. Hope to catch up with the morning sessions when recordings are made available.

15:29:00 Claudia Wiesinger:
Here is the TTS plug-in: https://community.rws.com/product-groups/trados-portfolio/trados-studio/w/customer-experience/5078/trados-tts

15:31:25 Josef van Genabith :
:)

(Closing Discussion)

15:34:21 Participants:
(To Pilar and Olga) Many thanks, Pilar and colleagues!

15:34:35 Participant:
(To Pilar and Olga) Yes, wonderful resources

15:36:07 Pilar Sánchez-Gijón:
Activities: http://tinyurl.com/nmtactivities

15:36:16 Participant:
A question to Joss Moorkens: Thank you for the inspiring talk. You have described a variety of ethical issues regarding NMT – how do you think should we react on them both in academia with special regard to teaching and in the profession? In particular, what kind of institutionalized responses to the issues you raised could you imagine?

15:36:30 Participant:
D-MT

15:36:37 Pilar Sánchez-Gijón:
Book outline and sample: http://tinyurl.com/hownmtworks

15:36:46 Participant:
Yes, a few days ago, it was published

15:37:13 Organiser:
https://tausdata.medium.com/machine-translation-redefined-with-demt-no-human-in-the-loop-3d4b7af0b3e3

15:37:48 Participant:
A question for later: Are there use cases today where you would advise against the use of PEMT and prefer a from scratch approach to translation? And since post-editors need to have a high level of expertise in their translation craft, would you say it's justified to reduce remuneration on PEMT tasks compared to from scratch translation?

15:38:04 Participant:
I think it is about selling data :)

15:38:17 Participant:
https://datamarketplace.taus.net/enhance-mt

15:39:39 Participant:
Apologies, this is the link I meant to share: https://tausdata.medium.com/machine-translation-redefined-with-demt-no-human-in-the-loop-3d4b7af0b3e3

15:40:02 Participant:
https://tausdata.medium.com/machine-translation-redefined-with-demt-no-human-in-the-loop-3d4b7af0b3e3

15:40:30 Participant:
Thank you all for the truly inspiring event, it was great!

15:42:19 Participant:
https://blog.taus.net/natural-language-technologies-nlt-for-ai-solutions

15:43:10 Participant:
Thank you, Joss, this is exactly what I do in my classes :-)

15:43:31 Joss Moorkens:
@Participant 👍

15:50:59 Participant:
Great points, Jean. ELIS 2022 were also so surprised that they released this before their full report:
https://euatc.org/industry-surveys/mt-hype-or-fact-elis-2022-early-trends-emerging/

15:51:32 Participant:
LCS: 20% of projects. Freelancers: 13% daily (MT hype or fact? Use rates)

15:53:18 Participant:
On the other hand, though, we hear some of the largest LSPs saying quite openly that they MT pretty much everything (but, of course, they don't share which clients for, what kind of content, etc...) The picture is still fuzzy, despite the apparently clear statements ;)

16:01:11 Participant:
That's a super interesting Q, Organiser. The thing is, though, that, despite on the one hand regretting the lower lexical density and variation in PEMT content, on the other hand some MT system trainers are quite active in separating localised segments from more literal translations. That's why I was wondering earlier if BLC did this, as well...

16:01:16 Joss Moorkens:
Wasn't that an example from your recent article Organiser?

16:02:34 Participant:
Question: is anyone aware of any MT developments that, beyond domain-specific training data, also incorporate text genre-related (e.g., e-mails vs. letters vs. chats vs. FAQs vs. fora, etc.) training sets?

16:03:32 Participant:
I am thinking of Polish, for instance

16:04:29 Participant:
Formal vs. informal uses across different text types and genres, e.g., the stuff MT struggles so much with

16:04:52 Participant:
Omniscien do a bit of user-generated content normalisation, or one of their former guests to their webinar series... or was it a GALA webinar?... uh... age... :/

16:05:12 Participant:
Yes, fully agree, Pilar. Thanks!

16:05:53 Participant:
Absolutely! Coudn´t agree more

16:06:18 Participant:
Many thanks to all the speakers!

16:06:22 Participant:
THANK YOU EVERYONE!

16:06:25 Participant:
Thank you so much for the entire event!

16:06:33 Participant:
Thank you very much for this interesting event, the organisation and all the talks and discussions! ☺

16:06:36 Participant:
It was a great event! Thanks! :-)

16:06:39 Participant:
Thank you very much!

16:06:40 Participant:
Fantastic conference! Thanks so much

16:06:41 Participant:
Thank you so much for the presenters and for the organizers

16:06:41 Participant:
Many thanks to the speakers and to the organizers!

16:06:41 Participant:
Thank you so much, Organiser and team! Super useful event! Many thanks to all the presenters!!! :)

16:06:42 Participant:
Very interesting event! Indeed thought-provoking!

16:06:42 Participant:
Thank You!!!

16:06:43 Participant:
Thank you very much!

16:06:44 Participant:
Thanks a lot for your great presentations!

16:06:44 Participant:
thanks a lot

16:06:45 Participant:
Many thanks for this great event!

16:06:46 Participant:
Thnk you!

16:06:48 Participant:
Many thanks!

16:06:48 Participant:
thanks, great event!!!

16:06:50 Participant:
Thank you!

16:06:52 Participant:
thank you for this excellent event to the organizers and the presenters

16:06:52 Participant:
Thank you, bye! :-)

16:06:53 Participant:
aDIOS

16:06:55 Participant:
thanks

16:06:58 Participant:
Thank you very much, great event!

16:06:59 Participant:
Thank you !

16:07:03 Participant:
Please send us the link of the recording? Thanks a lot

16:07:04 Participant:
Thank you!

16:07:06 Participant:
Thanks and waiting for more events like that

16:07:16 Participant:
Thank you

16:07:16 Participant:
Thanks to the organizers!

16:07:21 Participant:
Excellent event !!!

16:07:29 Participant:
Thank you all!

16:07:37 Participant:
Congratulations!!!