Charles Locock, Lowcock or Lockhart? Offline Speech Translation: Test Suite for Named Entities

Maximilian Awiszus[1], **Jan Niehues**[2], **Marco Turchi**[1], **Sebastian Stüker**[1], **Alex Waibel**[3]
[1]Zoom Video Communications, [2]Karlsruhe Institute of Technology, [3]Carnegie Mellon University
{maximilian.awiszus, marco.turchi, sebastian.stueker}@zoom.us, jan.niehues@kit.edu, waibel@cs.cmu.edu

Abstract

Generating rare words is a challenging task for natural language processing in general and in speech translation (ST) specifically. This paper introduces a test suite prepared for the Offline ST shared task at IWSLT. In the test suite, corresponding rare words (i.e. named entities) were annotated on TED-Talks for English and German and the English side was made available to the participants together with some distractors (irrelevant named entities). Our evaluation checks the capabilities of ST systems to leverage the information in the contextual list of named entities and improve translation quality. Systems are ranked based on the recall and precision of named entities (separately on person, location, and organization names) in the translated texts. Our evaluation shows that using contextual information improves translation quality as well as the recall and precision of NEs. The recall of organization names in all submissions is the lowest of all categories with a maximum of 87.5% confirming the difficulties of ST systems in dealing with names.

1 Introduction

Generating rare words is a big challenge for several natural language processing (NLP) tasks such as machine and speech translation and speech recognition. Rare words are those terms that have a low frequency in the training data and include, among others, named entities (NE), i.e. names of persons, organizations, and locations, acronyms and abbreviations, and domain-specific terms. These words carry a huge amount of the information of a sentence (Li et al., 2013) and their wrong realization in a text can significantly impact the user's understanding and experience.

In machine translation (MT), there has been a significant effort in making the translation system able to translate better the rare words (Sennrich et al., 2016; Koehn and Knowles, 2017; Niehues and Cho, 2017). This becomes crucial when translations serve as a base for upstream tasks like summarization, errors in those named entities can introduce wrong attributions or overall misleading information. To improve the accuracy of translating named entities correctly one could either integrate a knowledge graph (Mota et al., 2022; Xie et al., 2022) or use NE tags in the source sentence to make the NMT system aware of the NEs (Ugawa et al., 2018; Dinu et al., 2019; Zhou et al., 2020).

For automatic speech recognition (ASR) the problem with rare words and NEs is even harder since with speech the system has to handle an additional modality. Similar to NMT there is a lack of training data for those entities and in addition to that, the pronunciation of named entities is often different compared to other words. Current state-of-the-art approaches tackle this problem using contextual biasing (Sathyendra et al., 2022) where the ASR system is provided with contextual information which can be a list of named entities. The work is usually distinguished in a shallow-fusion, where the actual ASR model is untouched and only modifications are added at inference time (Wang et al., 2023) and a deep-fusion approach, where a context mechanism is trained and later used as a black-box (Munkhdalai et al., 2023; Zhou et al., 2023; Huber et al., 2021; Sathyendra et al., 2022; Bruguier et al., 2019).

In speech translation (ST), addressing the modality problems encountered in ASR and the lack of alternative translations for NEs in neural machine translation simultaneously increases the complexity of the problem. There is already existing work exploring the capability of ST system handling NEs (Gaido et al., 2021). Similar to their work also this test suite concentrates on evaluating the accuracy of translating named entities for person, organization and location names. Additional to Gaido et al. also the precision in translating named entities of the systems is evaluated. Furthermore contextual information is given per talk as a list of named entities

Proceedings of the 21st International Conference on Spoken Language Translation (IWSLT 2024), pages 339–345
August 15-16, 2024 ©2024 Association for Computational Linguistics

to evaluate if a system can utilize this information for the translation task.

It has been shown that the main factors for a cascaded system might be the frequency of words occurring in the training and foreign words with different pronunciations (Gaido et al., 2022). They suggest tackling the first factor by using more data, synthetic data, or fine-tuning on in-domain data. The second factor is tackled by using multilingual speech data to increase the variety seen for phoneme-to-grapheme mappings during training. Additionally, there has been work incorporating a list of named entities into a direct ST model to improve the accuracy for NE translation (Gaido et al., 2023) based on the CLAS approach for contextual ASR (Pundak et al., 2018).

We proposed a test suite for the Offline ST shared task at IWSLT to draw attention to the problem of NE translation in speech translation. The test suite was used to evaluate the ability of ST systems to translate NEs in the English-German TED test set accurately. The test suite provides contextual information in the form of a list of source language NEs that may or may not be present in the source spoken audio. The aim is to assist the ST system in improving translation quality. This paper introduces the test suite and examines the performance of different submitted ST systems on our test. Our findings indicate that ST systems encounter difficulties when translating NEs, but the list of NEs can help enhance the performance when utilized.

2 Test Suite

2.1 Task

This test suite has been developed to check the capability of a speech-translation system to leverage source language textual knowledge to improve the translation of specific aspects (i.e. named entities), and properly translate named entities.

For this reason, in addition to the classic test audio for the English into German translation direction, contextual information is available in textual form. This information might be used to mitigate translation errors on these contextual terms.

The context information was given as a list of entities per English audio file. To emulate real scenarios, where large lists can be used without any adaptation to specific audio, some entities that were not present in that audio were added as distractors. The goal of each participant and system is to distill the correct information from the list and use it to improve translation quality.

2.2 Data

As a test corpus we use 27 English TED talks with translations into German used as one of the evaluation sets in the Offline task.

A state-of-the-art multilingual fined-tuned named-entity-recognition (NER) model based on BERT (Kenton and Toutanova, 2019)[1] is used to annotate NEs in our test corpus for English as well as for German. The NER tagger outputs different name entity classes – in the following, we will concentrate on the most frequent classes which are person names, locations, and organization names.

Additionally, in the first post-processing step, some miss-classified words were manually removed and statistics of tagged words were calculated to get a consistent tagging of all words. In the second step, the correspondence for the named entities from English and German is estimated since we are only interested in named entities which occur in the reference as well as in the target. As an heuristic we construct a graph where each named entity is represented by a node. In the graph, there is only an edge between two nodes if the character edit distance of two named entities of the two different languages was below a specific threshold. To finally estimate the correspondence a maximum bipartite matching (Hopcroft and Karp, 1973) is calculated between the named entities of German and English per segment.

Finally, the lists for each segment were merged per talk resulting in a list of named entities with corresponding entities in English resp. German.

Exemplary excerpts of a talk can be examined in table 1.

Table 1: Exemplary corresponding *named entity* in the test corpus tagged by a NER model.

English Transcription
a. The Company and *Jan Pieterszoon Coen*, its Governor-General
b. In 1971, *East Pakistan* seceded
German Translation
a. Das Unternehmen und *Jan Pieterszoon Coen*, sein Generalgouverneur
b. 1971 spaltete sich *Ostpakistan* ab

[1]The cased version of BERT is used because also the transcripts resp. translations are provided cased.

The same procedure as described above was applied to nearly 400,000 sentences from other TED talks to extract named entities. In the final step for each English audio in the test set distractors were sampled from these entities to add at least one distractor per audio but a maximum reach of 20% distractors per audio (c.f. table 2). This results in a final named entity list containing 153 words in total (including distractors).

Table 2: Excerpt of the final context list containing named entities. One line corresponds to one whole audio of the utterance in table 1. The list was artificially augmented by adding **distractors (bold)**.

a. Banda, Banda Islands, Bandanese, Coen, **David Brin**, Europe, Jan Pieterszoon Coen, Verenigde Oostindische Compagnie
b. **Alex Kipman**, Assam, Bangladesh, Bengal, Calcutta, Delhi, Dhaka, East Pakistan, Hindus, India, Jawaharlal Nehru, Karachi, Kashmir, Lahore, Mohandas Gandhi, Muhammad Ali Jinnah, Pakistan, Punjab, **Shree Bose**

2.3 Metric

The submitted hypotheses were automatically resegmented based on the reference translation.

Since the hypothesis-reference sentence alignment might not always be correct in the following evaluation the named entity measurements are calculated per audio. A named entity in the hypotheses translations is considered a hit if an exact case-sensitive match in the reference is found and a miss otherwise. Those hits and misses per audio are then used to calculate the recall.

Furthermore the same procedure as described in section 2.2 was applied to all submitted translations. By finding a match of the detected named entities in the reference, the precision of translated named entities can be calculated which is reported as NE-Precision.[2]

The translation quality is computed using the COMET score (Rei et al., 2020).

3 ST Models

All tested systems are cascaded systems that first transcribe the audio by an ASR system and translate the transcript with an NMT system. That might be due to the fact that cascaded systems performed better than end-to-end systems for Offline ST in the last years' evaluations (Agarwal et al., 2023; Anastasopoulos et al., 2022, 2021). There exist three different data conditions[3]: Firstly constrained, where the systems are only allowed to be trained on a fixed amount of data, secondly constrained + LLM where in addition a list of allowed large language model (LLM) can be used and thirdly unconstrained to allow training the system and a large amount of training data.

The only system incorporating the contextual information is the submission of the Karlsruhe Institute of Technology (KIT). Their cascaded system uses a LoRA (Hu et al., 2021) fine-tuned LLM to 1) post-edit the ASR transcript incorporating the N-best list and 2) to post-edit the MT output on document-level. Only their primary (prm) submission injects contextual information in the second step by including the words into their LLM prompt. The first contrastive submission (ctr1) only applies the ASR post-edit step and for ctr2 both LLM corrections are used but without injecting the contextual information.

All unconstrained systems use a multilingual ASR model - namely Whipser-large-v3 (Radford et al., 2023) - for transcription.

As stated above also the Huawei Translation Service Center (HW-TSC) and Carnegie Mellon University (CMU) submitted a cascaded approach.

4 Results

All systems' results are reported in table 3 grouped by the aforementioned data condition (c.f. section 3). It can be observed that unconstrained systems are performing better on the general ST metric, COMET, as well as on the named entity recall and precision. Because the unconstrained systems are trained on more data, also the number of named entities might be higher, which directly is related to predicting named entities correctly (Gaido et al., 2022). Additionally the multilingual ASR component of the unconstrained cascaded ST systems might be beneficial for the translation of named entities because often names originate from different languages than the actual source language (English in our case). This observation is also au-pair with other investigations (Gaido et al., 2022). Also, we

[2]We want to note that this metric depends on the performance of the NER model used for extracting NEs on the different translation submissions.

[3]For more details visit the webpage of IWSLT-2024 offline track: `https://iwslt.org/2024/offline`

Table 3: Systems evaluated using general MT metric COMET as well as recall (NE-Recall) and precision (NE-Precision) of named entities per category person (per), location (loc) and organization (org) evaluated in the target language (German) and number of predicted distractors (DT).

System	**COMET**	**NE-Recall [%]**				**NE-Precision [%]**				**DT**
		ALL	per	loc	org	ALL	per	loc	org	
Data Condition: Unconstrained										
NYA (prm)	0.8339	88.68	**84.44**	97.78	75.00	75.15	76.36	**82.05**	57.89	-
NYA (ctr1)	0.8329	**91.51**	**84.44**	**100.00**	**87.50**	**74.56**	**78.18**	78.75	58.33	-
NYA (ctr2)	0.8330	**91.51**	**84.44**	**100.00**	**87.50**	74.55	77.36	78.48	57.14	-
NYA (ctr3)	0.8332	**91.51**	**84.44**	**100.00**	**87.50**	73.10	**78.18**	77.78	54.05	-
CMU (prm)	**0.8596**	83.96	80.00	93.33	68.75	64.61	65.08	72.15	47.50	-
CMU (ctr1)	0.8542	83.02	80.00	91.11	68.75	61.96	65.08	71.43	42.55	-
CMU (ctr2)	0.8358	83.96	80.00	93.33	68.75	63.74	65.57	75.64	42.55	-
HW-TSC (prm)	0.8461	88.68	**84.44**	95.56	81.25	71.76	75.41	76.71	54.05	-
HW-TSC (ctr)	0.8472	88.68	**84.44**	95.56	81.25	73.21	76.67	55.56	**78.08**	-
Data Condition: Constrained										
HW-TSC	0.8376	87.74	**84.44**	93.33	**81.25**	**73.91**	76.27	76.06	**60.61**	-
Data Condition: Constrained + LLM										
KIT (prm)	0.8283	87.74	**86.67**	93.33	75.00	68.75	73.68	78.08	42.42	0
KIT (ctr1)	0.8245	83.96	80.00	93.33	68.75	64.85	60.32	**79.45**	40.62	-
KIT (ctr2)	0.8260	85.85	84.44	93.33	68.75	66.47	67.80	78.38	40.54	-
HW-TSC	**0.8490**	**89.62**	**86.67**	**95.56**	**81.25**	**73.78**	**79.63**	74.36	**60.61**	-

might suspect a data leakage problem since the Whisper model was released in November 2023 and some TED talks from the test set are publicly available since 2013.

Furthermore the recall and precision for locations archives the highest score, followed by persons and then organization names. That might be related to the main factor of frequency of words occurring in the training which likely is higher for location names compared to person and organization names.

Looking closer at the unconstrained submissions one can observe that CMU's primary submission is the best-performing submission for COMET, but NYA's contrastive submissions achieve a better NE-Recall as well as NE-Precision.

Comparing HW-TSC's primary submission on the constrained data to the condition with LLM, it achieves the highest precision for named entities in general and also has a competitive performance for the recall.

From the results for KIT's primary (prm) and second contrastive (ctr2) submission, it can be seen that the overall recall and precision of NEs as well as the scores for person and organization names increased. This indicates that the provided context information can be useful to not only increase the general COMET score but also the translation for NEs.

Additionally, the number of appearing distractors (DT) in the translations was measured. Only KIT's primary submission used the provided context information and is therefore prone to copying a wrong-named entity from the provided list. Nevertheless, 0 distractors were copied from the provided context list.

Table 4: Exemplary misses for the person *named entity* (*Charles Locock*) as well as one correct translation of four German hypotheses translations of unconstrained - NYA (prm) and CMU (prm) - and constrained systems with a LLM - HW-TSC and KIT (ctr2).

Reference	
Mediziner wie Sir *Charles Locock*	
Hypotheses	
NYA (prm)	Ärzte wie Sir *Charles Lowcock*
CMU (prm)	Ärzte wie Sir *Charles Lockhart*
HW-TSC	Ärzte wie Sir *Charles Lowcock*
KIT (ctr2)	Ärzte wie Sir Charles Locock

In table 4 an example of a person-named entity that was mistranslated by most of the tested systems can be examined. In that example, only KIT's submissions translated the name *Charles Lo-*

cock correctly. Other systems translated the last name as *Lockhart*, *Lowcock*, or *Lowcock*. All miss-translations are close to the actual name *Locock* but might raise confusion when reading the translation without having access to the original audio.

Table 5: Two exemplary misses for the organizational *named entity* (*WARIF*) as well as two correct translations in four German hypotheses translations of unconstrained - NYA (prm) - and constrained systems - KIT (ctr2), KIT (prm) and HW-TSC.

Reference	
Internationale Stiftung für Frauen in Gefahr, *WARIF*, gegründet	
Hypotheses	
NYA (prm)	Women at Risk International Foundation, *WAR*
KIT (ctr2)	Women at Risk International Foundation (*WRIF*) gegründet
KIT (prm)	Women at Risk International Foundation (*WARIF*) gegründet
HW-TSC	Women at Risk International Foundation, *WARIF*

Additionally translations of an abbreviation resp. an organizational named entity, namely *WARIF* which is short for *Women At Risk International Foundation*, are reported in table 5. The NYA's primary resp. KIT's second contrastive system is missing the NE and translates it with only *WAR* resp. *WRIF*. Also, it's worth noting that when injecting the contextual information the KIT's primary system is translating this organizational NE correctly. For completes: also the HW-TSC's primary submission was translating this NE correctly without using any contextual information. Especially for organization terms, it's important to translate them correctly. In this example, it can be seen that a hallucinated abbreviation also introduces confusion and makes it hard to understand the meaning of the translation.

5 Conclusions

In our test suite, we explored the translation of named entities for English-German ST. Named entities are translated correctly with a recall of approx. 92% and a precision of approx. 75% in an unconstrained, approx. 88% resp. 74 % in a constrained data condition without LLMs and approx. 90% resp. 81% in a constrained data condition with using a LLM. Firstly this indicates that LLMs comprise contextual knowledge about named entities which is useful to translate named entities. But secondly that also suggests that there is still a gap in translating named entities correctly, especially looking at the category of organization names where when additionally using a LLM the precision and recall was not improved. Furthermore that might indicate that the capabilities of LLM of improving the quality of named entity translation is limited due to the fact that some mis-reconized named entities can not be corrected without the access to audio information in a cascaded system.

The given contextual information (list of named entities) improved the overall COMET score as well as the recall and precision of NE translation. We are looking forward having more systems using a context list for ST to see more benefits from using provided contextual information or LLMs using audio information for translation directly.

References

Milind Agarwal, Sweta Agrawal, Antonios Anastasopoulos, Luisa Bentivogli, Ondřej Bojar, Claudia Borg, Marine Carpuat, Roldano Cattoni, Mauro Cettolo, Mingda Chen, William Chen, Khalid Choukri, Alexandra Chronopoulou, Anna Currey, Thierry Declerck, Qianqian Dong, Kevin Duh, Yannick Estève, Marcello Federico, Souhir Gahbiche, Barry Haddow, Benjamin Hsu, Phu Mon Htut, Hirofumi Inaguma, Dávid Javorský, John Judge, Yasumasa Kano, Tom Ko, Rishu Kumar, Pengwei Li, Xutai Ma, Prashant Mathur, Evgeny Matusov, Paul McNamee, John P. McCrae, Kenton Murray, Maria Nadejde, Satoshi Nakamura, Matteo Negri, Ha Nguyen, Jan Niehues, Xing Niu, Atul Kr. Ojha, John E. Ortega, Proyag Pal, Juan Pino, Lonneke van der Plas, Peter Polák, Elijah Rippeth, Elizabeth Salesky, Jiatong Shi, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Yun Tang, Brian Thompson, Kevin Tran, Marco Turchi, Alex Waibel, Mingxuan Wang, Shinji Watanabe, and Rodolfo Zevallos. 2023. FINDINGS OF THE IWSLT 2023 EVALUATION CAMPAIGN. In *Proceedings of the 20th International Conference on Spoken Language Translation (IWSLT 2023)*, pages 1–61, Toronto, Canada (in-person and online). Association for Computational Linguistics.

Antonios Anastasopoulos, Loïc Barrault, Luisa Bentivogli, Marcely Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad, Clara Emmanuel, Yannick Estève, Marcello Federico, Christian Federmann, Souhir Gahbiche, Hongyu Gong, Roman Grundkiewicz, Barry Haddow, Benjamin Hsu, Dávid Javorský, Věra Kloudová, Surafel Lakew, Xutai Ma, Prashant Mathur, Paul McNamee, Kenton Murray, Maria Nădejde, Satoshi Nakamura, Matteo Negri, Jan Niehues, Xing Niu, John Ortega, Juan Pino, Elizabeth Salesky, Jiatong Shi, Matthias Sperber, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Yogesh Virkar, Alexander Waibel, Changhan Wang, and Shinji Watanabe. 2022. Findings of the IWSLT 2022 evaluation campaign. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 98–157, Dublin, Ireland (in-person and online). Association for Computational Linguistics.

Antonios Anastasopoulos, Ondřej Bojar, Jacob Bremerman, Roldano Cattoni, Maha Elbayad, Marcello Federico, Xutai Ma, Satoshi Nakamura, Matteo Negri, Jan Niehues, Juan Pino, Elizabeth Salesky, Sebastian Stüker, Katsuhito Sudoh, Marco Turchi, Alexander Waibel, Changhan Wang, and Matthew Wiesner. 2021. FINDINGS OF THE IWSLT 2021 EVALUATION CAMPAIGN. In *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, pages 1–29, Bangkok, Thailand (online). Association for Computational Linguistics.

Antoine Bruguier, Rohit Prabhavalkar, Golan Pundak, and Tara N Sainath. 2019. Phoebe: Pronunciation-aware contextualization for end-to-end speech recognition. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6171–6175. IEEE.

Georgiana Dinu, Prashant Mathur, Marcello Federico, and Yaser Al-Onaizan. 2019. Training neural machine translation to apply terminology constraints. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3063–3068, Florence, Italy. Association for Computational Linguistics.

Marco Gaido, Matteo Negri, Marco Turchi, et al. 2022. Who are we talking about? handling person names in speech translation. In *Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022)*, pages 62–73. Association for Computational Linguistics (ACL).

Marco Gaido, Rodríguez Susana, Matteo Negri, Luisa Bentivogli, and Marco Turchi. 2021. Is" moby dick" a whale or a bird? named entities and terminology in speech translation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1707–1716. Association for Computational Linguistics.

Marco Gaido, Yun Tang, Ilia Kulikov, Rongqing Huang, Hongyu Gong, and Hirofumi Inaguma. 2023. Named entity detection and injection for direct speech translation. In *Proceedings of ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

John E Hopcroft and Richard M Karp. 1973. An n^5/2 algorithm for maximum matchings in bipartite graphs. *SIAM Journal on computing*, 2(4):225–231.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.

Christian Huber, Juan Hussain, Sebastian Stüker, and Alexander Waibel. 2021. Instant one-shot word-learning for context-specific neural sequence-to-sequence speech recognition. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 1–7. IEEE.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Philipp Koehn and Rebecca Knowles. 2017. Six challenges for neural machine translation. In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver. Association for Computational Linguistics.

Haibo Li, Jing Zheng, Heng Ji, Qi Li, and Wen Wang. 2013. Name-aware machine translation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 604–614, Sofia, Bulgaria. Association for Computational Linguistics.

Pedro Mota, Vera Cabarrao, and Eduardo Farah. 2022. Fast-paced improvements to named entity handling for neural machine translation. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 141–149, Ghent, Belgium. European Association for Machine Translation.

Tsendsuren Munkhdalai, Zelin Wu, Golan Pundak, Khe Chai Sim, Jiayang Li, Pat Rondon, and Tara N Sainath. 2023. Nam+: Towards scalable end-to-end contextual biasing for adaptive asr. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 190–196. IEEE.

Jan Niehues and Eunah Cho. 2017. Exploiting linguistic resources for neural machine translation using multi-task learning. In *Proceedings of the Second Conference on Machine Translation*, pages 80–89.

Golan Pundak, Tara N Sainath, Rohit Prabhavalkar, Anjuli Kannan, and Ding Zhao. 2018. Deep context: end-to-end contextual speech recognition. In *2018 IEEE spoken language technology workshop (SLT)*, pages 418–425. IEEE.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

Ricardo Rei, Craig Stewart, Ana C Farinha, and Alon Lavie. 2020. Comet: A neural framework for mt evaluation. In *Proceedings of the 2020 Conference on*

Empirical Methods in Natural Language Processing (EMNLP), pages 2685–2702.

Kanthashree Mysore Sathyendra, Thejaswi Muniyappa, Feng-Ju Chang, Jing Liu, Jinru Su, Grant P Strimel, Athanasios Mouchtaris, and Siegfried Kunzmann. 2022. Contextual adapters for personalized speech recognition in neural transducers. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8537–8541. IEEE.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Arata Ugawa, Akihiro Tamura, Takashi Ninomiya, Hiroya Takamura, and Manabu Okumura. 2018. Neural machine translation incorporating named entity. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3240–3250, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Weiran Wang, Zelin Wu, Diamantino Caseiro, Tsendsuren Munkhdalai, Khe Chai Sim, Pat Rondon, Golan Pundak, Gan Song, Rohit Prabhavalkar, Zhong Meng, et al. 2023. Contextual biasing with the knuth-morris-pratt matching algorithm. *arXiv preprint arXiv:2310.00178*.

Shufang Xie, Yingce Xia, Lijun Wu, Yiqing Huang, Yang Fan, and Tao Qin. 2022. End-to-end entity-aware neural machine translation. *Machine Learning*, 111(3):1181–1203.

Leiying Zhou, Wenjie Lu, Jie Zhou, Kui Meng, and Gongshen Liu. 2020. Incorporating named entity information into neural machine translation. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part I 9*, pages 391–402. Springer.

Shilin Zhou, Zhenghua Li, Yu Hong, Min Zhang, Zhefeng Wang, and Baoxing Huai. 2023. Copyne: Better contextual asr by copying named entities. *arXiv preprint arXiv:2305.12839*.