

KpopMT: Translation Dataset with Terminology for Kpop Fandom

JiWoo Kim
Sungkyunkwan University
Suwon, South Korea
wldn9705@skku.edu

Yunsu Kim
aiXplain Inc.
Los Gatos, USA
yunsu.kim@aixplain.com

JinYeong Bak
Sungkyunkwan University
Suwon, South Korea
jy.bak@skku.edu

Abstract

While machines learn from existing corpora, humans have the unique capability to establish and accept new language systems. This makes human form unique language systems within social groups. Aligning with this, we focus on a gap remaining in addressing translation challenges within social groups, where in-group members utilize unique terminologies. We propose KpopMT dataset, which aims to fill this gap by enabling precise terminology translation, choosing Kpop fandom as an initiative for social groups given its global popularity. Expert translators provide 1k English translations for Korean posts and comments, each annotated with specific terminology within social groups’ language systems. We evaluate existing translation systems including GPT models on KpopMT to identify their failure cases. Results show overall low scores, underscoring the challenges of reflecting group-specific terminologies and styles in translation. We make KpopMT publicly available.¹

1 Introduction

One of the most profound distinctions between humans and machines lies in their ability to form a new language system. While machines learn from and rely on existing corpora, humans have the unique capability to establish new social conventions, such as agreeing to call an apple “apple”. This inventive ability is deeply intertwined with the social nature of language, wherein words gain influence and become integrated into the vernacular of social groups – termed as social dialect – through social interactions and contexts.

In light of this, our research places a particular emphasis on specific phenomena: Social groups often develop their unique linguistic systems, replete with specific terminologies and jargon, as in Table 1 (Wolfram, 2004; Peterson, 2014). The global

¹<https://github.com/skswldndi/KpopMT>

Kpop Fandom Group’s Language System	
Korean	원래 덕질 할 때 갠팬 수준으로 최애 위주 로만 파는 타입이었음
English	Originally, when I stan , I tended to focus on only my bias almost it seems like solo stan .

Table 1: Example of how specialized terminologies (colorboxes) are used in a global social group.

surge in social media usage has brought individuals from diverse countries and continents together, forming cohesive communities (Sawyer and Chen, 2012). An example is the Kpop fandom across various regions (Choi et al., 2014); these fans seek to connect and establish interpersonal relationships that transcend language barriers (Malik and Haidar, 2021).

However, the unique linguistic systems of social groups are currently not adequately represented in machine translation (MT) systems. Despite proposals from researchers to consider language diversity in MT (Kumar et al., 2021; Lakew et al., 2018), including cross-domain standard languages (Hu et al., 2019; Currey et al., 2017), local dialects (Abe et al., 2018; Hassan et al., 2017), and related languages (Pourdamghani and Knight, 2017), the specialized language systems of social groups remain underexplored in MT research. For example, Google Translator translates the example sentence in Table 1 as *Originally, when I was a fan, I focused on my favorites. It was the type that only sells.*, which differs from the specialized language system of the group.

In order to promote research in this direction, we argue for the necessity of a benchmark dataset that encompasses the language systems of social groups. This dataset is also crucial for demonstrating the capabilities of the current translation systems specifically tailored to handle the nuances of social group languages.

In our study, we reference the field of terminology-based MT (Xie and Way, 2020; Knowles et al., 2023), which focuses on accurately translating specialized terminology in machine-generated output. Notably, the medical domain has seen significant advancements in this area (Alam et al., 2021; Anastasopoulos et al., 2020), highlighted by the release of datasets that aid the NLP community, particularly those interested in terminology translation and constrained neural MT tasks (Zhang et al., 2023; Wang et al., 2022).

We propose a terminology-tagged MT dataset tailored for social groups, named **KpopMT**. Through human survey, we demonstrate participants in a social group show a strong preference for reference translations in KpopMT compared to translations without terminology. We choose the Kpop fandom as the social group for this initiative, given its global popularity and the widespread sharing of content on social media platforms among fans from diverse countries (Ringland et al., 2022).

KpopMT is collected from Korean posts and comments found on fan-related websites, including X (Twitter). Expert translators fluent in the terminologies used within this social group provide English translations, resulting in 1k sentence pairs, each annotated with specific terminologies. We evaluate existing translation systems, including state-of-the-art models such as GPTs, on KpopMT to identify areas of improvement and establish baseline performance for future research.

2 KpopMT

KpopMT consists of three parts: 1) the parallel sentence tagged with terminologies, 2) termbase which contains parallel glossary, and 3) fandom monolingual dataset. In this section, we illustrate how the parallel sentence and termbase are constructed. We address fandom monolingual dataset in Section 3.1.

In Table 2, we show an example of KpopMT. Capturing the most distinctive part of social groups’ language system is terminology, we also provide terminology information. The parallel sentence has fandom-related terms included in both the source and target side, which are annotated as tags.

To ensure translation reliability, we get confirmation from five native English Kpop fans who know both Korean fandom terms and English fandom terms.

2.1 Construction

KpopMT is constructed in two phases. First, we construct parallel sentences. Second, we annotate parallel terminology information in the sentences.

Sentence Phase To obtain sentences that include fandom-related terminology, we make a query list derived from crowd-sourced dictionaries.² Using the query list, we manually collect Korean monolingual data from the fan community sites^{3,4} and Twitter. Then we hire ten human translators who pass the qualification test, by asking for English translations of ten Korean fandom terms and their meanings. We only hire translators who answer correctly for at least eight terms. Then we ask them to translate Korean sentences into English sentences, resulting in 1,000 sentence pairs.

Please note that we ask them to translate with the inclusion of English fandom terms, such as ‘stan.’ If there are no specific English fandom terms, we ask the translators to include internet slang in the sentence as a variation of standard language. We also forbid the usage of translation services such as Google Translate.

Terminology Phase In this phase, we mark terms included in the parallel sentences first. After extracting all marked terms, we create a parallel glossary by matching Korean terms with their English counterparts. If there are other possible or morphological variations in English terms, we include them on the target of the glossary, separating them with ‘|’. Subsequently, experts on the terminology used within fandom and internet slang confirm the consistency of parallel glossary.

Since we separate the sentence phase and the terminology phase, KpopMT is not structured according to the terminologies in this glossary. This complicates the creation of a one-to-one dictionary. For instance, ‘떡볶이’ is not a direct equivalent of only ‘local’. Therefore, we opt to begin with the Korean side as it aligns with the actual translation direction in which the data was generated, following Alam et al. (2021).

We categorize the glossary terms into three: **Group-Lexicon**, **Group-NE**, and **Slang**. Group-Lexicon refers to the lexicon unique to the fandom, which may not be understood by those outside the fandom. Group-NE represents elements related to

²Naver Open-Dict <http://naver.me/xUyZv8N>

³Theqoo <http://theqoo.net>

⁴Instiz <http://instiz.net>

1. Sentence Phase	
Source	트친 이랑 푸드코트에서 밥 먹다가 용수 만났어요. 포카 보여주고 팬이라고 싸인도 받았어요. 덕계못 인줄 알았는데
Reference	I met Yongsoo at food court while eating with my moot. We showed him pc and got an autograph saying that we're fans of him. I thought <i>stan</i> can't get luck to see <i>faves</i>
2. Terminology Phase	
Tag 1	<term id="202" type="slang" source="트친" target="twitter mootslmootslmoot"> moot </term>
Tag 2	<term id="146" type="group-NE" source="용수" target="Yongsoo"> Yongsoo </term>
Tag 3	<term id="8" type="group-lexicon" source="포카" target="pocaslpocalpcslpc"> pc </term>

Table 2: Example of KpopMT.

the fandom’s named entities, such as idol group names or nicknames. Slang encompasses internet slang, which is a variation of standard language. We only apply truecasing to Group-Lexicon and not to Group-NE and Slang, as capitalization is crucial for names and we believe there may be different meanings in truecased slang (e.g., ‘ig’ and ‘IG’).

In the end, we tag the sentence pairs on both source and target side with possible terminology translations from the glossary. We tag the sentence pairs with terminology translations only if both source terminology and a corresponding target terminology exist in the reference translation, following Alam et al. (2021). So, we do not tag ‘stan’ and ‘faves’ terms in Table 2.

2.2 Key Characteristics

The parallel corpus of KpopMT contains 6k Korean tokens and 12k English tokens, excluding the tags. It has a substantial number of tagged sentences, totaling 1,035, among which the tags categorized as Group-Lexicon are the most prevalent, accounting for 858 (82.8%), followed by Group-NE with 92 (8.8%), and Slang with 85 (8.4%). This indicates that KpopMT contains substantial information on a specific social group, which is Kpop fandom.

In addition, to assess its suitability for evaluating terminology-based translation, we compare KpopMT with TICO-19, widely recognized as the most common terminology machine translation dataset (Table 3) (Anastasopoulos et al., 2020; Odermatt et al., 2023). It is important to note that significant portions of TICO-19 lack any terminological content. KpopMT encompasses more extensive meaningful portions of terms compared to TICO-19.

Dataset	Number of Terms				
	0	1	2	>=3	max
Ours	27.5%	47.7%	19.8%	5.0%	5
TICO-19	40.2%	22.6%	6.6%	3.9%	9

Table 3: Comparison with TICO-19 regarding number of terminologies in each line.

Moreover, to demonstrate the necessity of KpopMT, we conduct a human survey involving five native English-speaking Kpop fans. None of them are involved in data construction process. They are given with 80 English sentences, each of which has two versions: samples from KpopMT with fandom-specific terms and expressed in standard language. Fans strongly prefer translations with fandom terms (89.75%) when asked which ones make them feel more connected to fandom members. This highlights the importance of considering cultural factors in communication (Gudykunst, 2003), making KpopMT valuable for both translation accuracy and social connectedness.

3 Experiments

We evaluate existing machine translation systems on KpopMT to assess its difficulty.

3.1 Experimental Setting

Data In our experiments, we utilize two types of data: general standard language data and fandom language data. To acquire the general standard language data, we download a Korean-English dataset (800k) from AI Hub, which is a platform releasing AI data by the Korean government.⁵ To obtain the fandom language data, we scrape 40k monolingual

⁵<https://www.aihub.or.kr>

Systems	Fandom Data	General Data	Finetuned from
M2M		✓	
mBART w/o Fandom		✓	mBART
mBART w/ Fandom	✓	✓	mBART
SL-MT		✓	
Domain Adaptation	✓	✓	SL-MT

Table 4: Features of open-source baselines.

data samples for each language from fan-related websites, employing the query list specified in 2.1. Korean monolingual data has 287k tokens with an average sentence length of 29.79, while English monolingual data has 376k tokens with an average length of 49.61. For evaluation purposes, we employ a test split comprising 500 sentences from our parallel KpopMT.

Baselines We implement two kinds of baselines: open-source machine translation models and proprietary machine translation systems. Our baseline choice considers three factors: general standard language data trained with, fandom monolingual data trained with (Table 4), and state-of-the-art systems.

Regarding general standard language data, we set baselines of M2M, mBART w/o Fandom and Standard Language MT (SL-MT). M2M is a multilingual translation model that can translate between any pair from 100 languages (Fan et al., 2021). mBART w/o Fandom is a finetuned translation model from multilingual Bart (mBART) using general data (Liu et al., 2020). mBART is the state-of-the-art model on IWSLT-17 (Cettolo et al., 2017; Park et al., 2021). SL-MT is a Korean-English state-of-the-art translation model on general standard language data (Park et al., 2020).

Regarding fandom monolingual data, we set baselines of mBART w/ Fandom and Domain Adaptation. mBART w/ Fandom is a finetuned translation model from mBART with an injection of both general parallel data and fandom monolingual data. We use the back-translation technique to make pseudo-parallel data for fandom monolingual data (Sennrich et al., 2016). Domain Adaptation is a finetuned translation model from Standard Language MT, with the technique of iterative back-translation of fandom monolingual data (Hu et al., 2019; Dou et al., 2019). The differences from mBART w/ Fandom are the pretrained model and iteration of back-translation. In a preliminary study,

we find that mBART w/ Fandom’s performance drops with iterative back-translation.

For proprietary machine translation systems, we experiment with OpenAI’s GPT models (GPT-3.5 and GPT-4) and Google Translator. For OpenAI’s GPT models (OpenAI, 2023; Eloundou et al., 2023), we assign the role of a Kpop fan to the model. This fan is familiar with terminologies used within the Kpop fandom and internet slang, which has shown empirically the best performance. We provide the prompt in Korean.

Evaluation We evaluate systems on both translation accuracy and terminological accuracy. Translation accuracy is evaluated with standard reference-based MT metrics: SacreBLEU (Post, 2018; Papineni et al., 2002), COMET (Rei et al., 2020), chrF++ (Popović, 2017). Terminology accuracy is evaluated with exact-match term accuracy (EMA) and 1-TERm score (Anastasopoulos et al., 2021). EMA is an accuracy score that searches for exact term translation matches (of the terminology required output) over the original hypothesis. The 1-TERm score is a modification of the TER metric, biased to assign higher edit cost weights for words belonging to a term (and then simply reversed so that a higher score is better). When computing terminology targeted evaluation, we consider synonyms which are split by ‘|’ in the tags. We rank systems according to EMA.

3.2 Results

Table 5 shows overall low scores, underscoring the considerable challenges posed by KpopMT in terms of both terminology-focused content and translation quality.

GPTs Overall, the GPT models demonstrate higher scores than other systems. When examining the words they succeed in generating, they excel in producing certain words compared to other models. ‘bias,’ ‘ult,’ and ‘stan’ are frequently generated, contributing to GPTs’ higher EMA scores.

However, Group-Lexicon type faces challenges, especially when excluding those basic words. We conduct an analysis to determine which types of terms GPT-4 could generate or not. We calculate each type’s EMA score, resulting in Group-Lexicon 16.3%, Group-NE 21.54%, and Slang 20%. As seen in overall low scores in success translation rate for all types, GPT-4 struggles with generating terminologies, particularly those related to Group-Lexicon. Given the importance of ‘Group-Lexicon’

	Systems	Terminology-focused		Translation Quality		
		EMA	1-TERm	BLEU	COMET	chrF++
Open Source	M2M	2.4	10.5	4.1	51.2	19.9
	mBART w/o Fandom	6.7	13.7	10.4	60.5	32.3
	mBART w/ Fandom	6.4	12.3	9.1	56.6	29.0
	SL-MT	3.8	6.5	8.7	56.4	30.3
	Domain Adaptation	13.9	13.3	9.7	57.7	31.2
Proprietary	GPT-3.5-turbo-0613	19.3	1.1	8.7	65.8	32.9
	GPT-4-0613	26.4	10.3	9.9	65.8	35.0
	Google Translator	10.4	16.3	13.9	65.4	35.9

Table 5: Result on test split set using existing machine translation systems.

in Kpop social groups, this deficiency shows a clear need for better translation in specific social contexts.

General Data We ascertain whether a translation model trained on general standard language data could grasp specific language systems within social groups. Although mBART w/o Fandom displays better performance in translation quality evaluations, its EMA score remains low, indicating a lack of specific knowledge pertaining to social groups.

Fandom Data Our findings suggest that incorporating fandom data into the training process does not consistently yield improved results. The possible explanation is fandom monolingual data during the back-translation process is noisy, resulting in pseudo-parallel data that significantly deviated from general data, which seems to have hindered proper comprehension. Following [Michel and Neubig \(2018\)](#) and calculating the perplexity score of fandom monolingual data using a language model trained with general data, we get 1395.4 for Korean and 713.8 for English. This limitation might prevent effective inference of the meaning and context of relevant terminology.

Translation Quality As highlighted by previous research ([Pascual et al., 2020](#)), there exists a tension between fluency of the translation and terminology accuracy in our results. Based on this, we intend to explore in future work whether it is possible to enhance the overall translation quality while preserving social groups’ distinctive terminology.

4 Conclusions

Our approach acknowledges the dynamic and evolving nature of language, especially in digitally mediated communities, which are often underrepresented in traditional linguistic resources. We propose a benchmark dataset for social groups’

language systems, named KpopMT. 1k parallel dataset contains not only Korean-English parallel sentences but also terminology information of Kpop fandom group. We also make termbase and monolingual data publicly available. We evaluate existing translation approaches on KpopMT to identify their failure cases. Our future plan includes expanding KpopMT to encompass other social groups, such as sports and global movie communities.

Limitations

In the study, we view Kpop fandom as a broad spectrum and do not focus on a specific fandom of any one idol group. There are numerous Kpop Idols, such as BTS, Twice, Seventeen, etc. and each terminology of their fans is distinct from one another. For instance, BTS fans use *Borahae* to mean *I love you*, while Seventeen fans use *horanghae* to mean the same thing.

Ethical Considerations

The created dataset will be released under the terms of the Twitter API. Aside from Tweets, which are sourced from publicly accessible websites, we ensure that there is no violation of copyright or invasion of privacy. To prevent user tracking, we remove any information related to users. All of our dataset is made publicly available through a Creative Commons CC BY-SA 4.0 license.

We compensate volunteer translators with more than the minimum wage in Korea. They are fully aware of their tasks, and we provide them with detailed annotation instructions. After compensation, we anonymize and remove their personal information. Additionally, we carefully filter out data containing hate speech about celebrities to ensure that human translators do not face any risks or harm associated with their participation.

Acknowledgments

We would like to thank the anonymous reviewers for their helpful questions and comments. We further would like to express our gratitude to Jonas Belouadi for discussions, proofreading, and comments on our work. This research was partly supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Korean government (MSIT) (NRF-2021M3A9E4080780, RS-2023-00252083), Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00421, AI Graduate School Support Program (Sungkyunkwan University)).

References

- Kaori Abe, Yuichiroh Matsubayashi, Naoaki Okazaki, and Kentaro Inui. 2018. [Multi-dialect neural machine translation and dialectometry](#). In *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation*, Hong Kong. Association for Computational Linguistics.
- Md Mahfuz Ibn Alam, Ivana Kvpilíková, Antonios Anastasopoulos, Laurent Besacier, Georgiana Dinu, Marcello Federico, Matthias Gallé, Kweonwoo Jung, Philipp Koehn, and Vassilina Nikoulina. 2021. Findings of the wmt shared task on machine translation using terminologies. In *Proceedings of the Sixth Conference on Machine Translation*, pages 652–663.
- Antonios Anastasopoulos, Laurent Besacier, James Cross, Matthias Gallé, Philipp Koehn, Vassilina Nikoulina, et al. 2021. On the evaluation of machine translation for terminology consistency. *arXiv preprint arXiv:2106.11891*.
- Antonios Anastasopoulos, Alessandro Cattelan, Zi-Yi Dou, Marcello Federico, Christian Federmann, Dmitriy Genzel, Francisco Guzmán, Junjie Hu, Macduff Hughes, Philipp Koehn, et al. 2020. Tico-19: the translation initiative for covid-19. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.
- Mauro Cettolo, Marcello Federico, Luisa Bentivogli, Jan Niehues, Sebastian Stüker, Katsuhito Sudoh, Koichiro Yoshino, and Christian Federmann. 2017. [Overview of the IWSLT 2017 evaluation campaign](#). In *Proceedings of the 14th International Conference on Spoken Language Translation*, pages 2–14, Tokyo, Japan. International Workshop on Spoken Language Translation.
- Seong Cheol Choi, Xanat Vargas Meza, and Han Woo Park. 2014. South korean culture goes latin america: Social network analysis of kpop tweets in mexico. *International Journal of Contents*, 10(1):36–42.
- Anna Currey, Antonio Valerio Miceli-Barone, and Kenneth Heafield. 2017. Copied monolingual data improves low-resource neural machine translation. In *Proceedings of the second conference on machine translation*, pages 148–156.
- Zi-Yi Dou, Junjie Hu, Antonios Anastasopoulos, and Graham Neubig. 2019. [Unsupervised domain adaptation for neural machine translation with domain-aware feature embeddings](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1417–1422, Hong Kong, China. Association for Computational Linguistics.
- Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. [Gpts are gpts: An early look at the labor market impact potential of large language models](#). *Preprint*, arXiv:2303.10130.
- Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, et al. 2021. Beyond english-centric multilingual machine translation. *J. Mach. Learn. Res.*, 22(107):1–48.
- William B Gudykunst. 2003. *Cross-cultural and intercultural communication*. Sage.
- Hany Hassan, Mostafa Elaraby, and Ahmed Y. Tawfik. 2017. [Synthetic data for neural machine translation of spoken-dialects](#). In *Proceedings of the 14th International Conference on Spoken Language Translation*, pages 82–89, Tokyo, Japan. International Workshop on Spoken Language Translation.
- Junjie Hu, Mengzhou Xia, Graham Neubig, and Jaime G Carbonell. 2019. Domain adaptation of neural machine translation by lexicon induction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2989–3001.
- Rebecca Knowles, Samuel Larkin, Marc Tessier, and Michel Simard. 2023. Terminology in neural machine translation: A case study of the canadian hansard. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 481–488.
- Sachin Kumar, Antonios Anastasopoulos, Shuly Winter, and Yulia Tsvetkov. 2021. [Machine translation into low-resource language varieties](#). In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 110–121, Online. Association for Computational Linguistics.
- Surafel Melaku Lakew, Aliia Erofeeva, and Marcello Federico. 2018. [Neural machine translation into language varieties](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 156–164, Brussels, Belgium. Association for Computational Linguistics.

- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Zunera Malik and Sham Haidar. 2021. [English language learning and social media: Schematic learning on kpop stan twitter](#). *E-Learning and Digital Media*, 18(4):361–382.
- Paul Michel and Graham Neubig. 2018. [Mtn: A testbed for machine translation of noisy text](#). *Preprint*, arXiv:1809.00388.
- Frédéric Oudermatt, Béni Egressy, and Roger Wattenhofer. 2023. Cascaded beam search: Plug-and-play terminology-forcing for neural machine translation. *arXiv preprint arXiv:2305.14538*.
- OpenAI. 2023. [Gpt-4 technical report](#). *Preprint*, arXiv:2303.08774.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Chanjun Park, Sugyeong Eo, Hyeonseok Moon, and Heui-Seok Lim. 2021. Should we find another model?: Improving neural machine translation performance with one-piece tokenization method without model modification. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Papers*, pages 97–104.
- Kyubyong Park, Joohong Lee, Seongbo Jang, and Da-woon Jung. 2020. [An empirical study of tokenization strategies for various Korean NLP tasks](#). In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 133–142, Suzhou, China. Association for Computational Linguistics.
- Damian Pascual, Beni Egressy, Florian Bolli, and Roger Wattenhofer. 2020. Directed beam search: Plug-and-play lexically constrained language generation. *arXiv preprint arXiv:2012.15416*.
- Britt Peterson. 2014. The linguistics of lol. what internet vernacular reveals about the evolution of language. *The Atlantic*, 10:2014.
- Maja Popović. 2017. chrf++: words helping character n-grams. In *Proceedings of the second conference on machine translation*, pages 612–618.
- Matt Post. 2018. A call for clarity in reporting bleu scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191.
- Nima Pourdamghani and Kevin Knight. 2017. [Deciphering related languages](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2513–2518, Copenhagen, Denmark. Association for Computational Linguistics.
- 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.
- Kathryn E Ringland, Arpita Bhattacharya, Kevin Weatherwax, Tessa Eagle, and Christine T Wolf. 2022. Army’s magic shop: Understanding the collaborative construction of playful places in online communities. In *CHI Conference on Human Factors in Computing Systems*, pages 1–19.
- Rebecca Sawyer and Guo-Ming Chen. 2012. The impact of social media on intercultural adaptation.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. [Improving neural machine translation models with monolingual data](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.
- Shuo Wang, Peng Li, Zhixing Tan, Zhaopeng Tu, Maosong Sun, and Yang Liu. 2022. A template-based method for constrained neural machine translation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3665–3679.
- Walt Wolfram. 2004. Social varieties of american english. *Language in the USA: Themes for the twenty-first century*, pages 58–75.
- Guodong Xie and Andy Way. 2020. Constraining the transformer nmt model with heuristic grid beam search. In *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 36–49.
- Jinpeng Zhang, Nini Xiao, Ke Wang, Chuanqi Dong, Xiangyu Duan, Yuqi Zhang, and Min Zhang. 2023. [Disambiguated lexically constrained neural machine translation](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 10583–10596, Toronto, Canada. Association for Computational Linguistics.