Linguistically Motivated Neural Machine Translation

Haiyue Song and **Hour Kaing** and **Raj Dabre**
National Institute of Information and Communications Technology (NICT)
Hikaridai 3-5, Seika-cho, Soraku-gun, Kyoto, Japan
{haiyue.song,hour_kaing,raj.dabre}@nict.go.jp

Abstract

In this tutorial, we focus on a niche area of neural machine translation (NMT) that aims to incorporate linguistics into different stages in the NMT pipeline, from pre-processing to model training to evaluation. We first introduce the background of NMT and fundamental analysis tools, such as word segmenters, part-of-speech taggers, and dependency parsers. We then cover topics including 1) word/subword segmentation, and character decomposition during MT data pre-processing, 2) incorporating direct and indirect linguistic features into NMT models, and 3) fine-grained linguistic evaluation for MT systems. We reveal the impact of orthography, syntax, and semantics information on translation performance. This tutorial is mainly aimed at researchers interested in the intersection of linguistics and low-resource machine translation. We hope this tutorial inspires and encourages them to develop linguistically motivated high-quality MT systems and evaluation benchmarks.

1 Relevence to the MT community

For machine translation (MT) tasks, purely data-driven approaches have been dominant in recent years, and in turn language knowledge-related approaches are being neglected. However, data is not always sufficient for all $7,000+$ languages worldwide. For NMT, a large number of parallel sentences are required to supervise a system to learn how to translate. In contrast, systems with limited training data show very limited performance, where leveraging external knowledge, such as linguistic knowledge, becomes essential.

Language is a structural system that consists of grammar and vocabulary. Grammar governs units in vocabulary to convey meanings, which humans use to communicate. Many natural language processing researchers believe that models with the ability to imitate human behavior would produce natural outputs to communicate with humans. This tutorial aims to cover the efforts that leverage linguistic knowledge to improve NMT, which emerged from 2016. Our tutorial intends to answer the question of

How to incorporate various linguistic knowledge into the development and evaluation of MT systems?

To answer this, we will dive deep into three areas: 1) the role of word segmentation, subword segmentation, and character decomposition during pre-processing, 2) the impact of direct and indirect linguistic features on MT models, and 3) fine-grained linguistic evaluation for NMT systems. This tutorial should benefit researchers who are focusing on low-resource MT where the parallel data is limited but linguistic analysis tools exist for the source or/and target language, which is often the case. Therefore, most methods we will introduce in this tutorial are highly generalizable. In addition, this tutorial could be a good starting point for increasing researchers' interest and awareness about linguistic methods in the neural era, building linguistic analysis tools for low-resource languages, and exploring more effective linguistic knowledge assisting methods even for high-resource language pairs.

2 Tutorial Overview

This tutorial covers techniques incorporating linguistic knowledge into NMT systems. We begin with a brief introduction to MT and the connection to linguistics and the NMT architecture (Vaswani et al., 2017). We then cover how linguistic knowledge can help NMT in different stages of the pipeline, including pre-processing, training, and evaluation.

In the pre-processing stage, we introduce how to leverage linguistic information from word segmentation (Tolmachev et al., 2018), subword segmentation (Song et al., 2022), and character decomposition (Zhang and Komachi, 2018) into input and output data instead of purely compression-based tokenization (Kudo and Richardson, 2018). We also cover how to leverage data from related languages (Amrhein and Sennrich, 2020).

For the model training stage, we discuss how to integrate linguistic features such as morphology and syntax information into the encoder and decoder of the NMT models. First, we introduce tools to generate linguistic features (Manning et al., 2014). We then introduce how to utilize them such as turning them into additional input embeddings (Sennrich and Haddow, 2016) and modifying the model architecture to leverage hierarchical sentence structure during encoding (Eriguchi et al., 2016) and decoding (Eriguchi et al., 2017).

Lastly, we cover works that evaluate or analyze the performance of linguistic phenomenons (Avramidis and Macketanz, 2022; Voita et al., 2019) for both the traditional NMT systems and large language models (LLMs).

3 Tutorial Outline

Below we list an outline of the general structure of the tutorial and only the most representative works under each section for brevity.

1. Introduction to Neural Machine Translation (20 minutes)
    - Brief introduction to MT and its historical connection to linguistics.
    - Overview of the basic NMT architectures (Bahdanau et al., 2016; Vaswani et al., 2017).

2. Linguistically Motivated Tokenization and Transfer Learning (30 minutes)
    - Word segmentation for languages without spaces as word boundaries (Tolmachev et al., 2018).
    - Linguistically motivated subword segmentation (He et al., 2020; Song et al., 2022; Batsuren et al., 2021; Ataman et al., 2017).
    - Character decomposition (Zhang and Komachi, 2018).
    - Noisy tokenization for related languages (Maurya et al., 2024; Brahma et al., 2023).
    - Transfer learning from related languages (Amrhein and Sennrich, 2020; Husain et al., 2024; Gala et al., 2023; Dabre et al., 2021; Song et al., 2020; Joshi et al., 2024).

3. Augmenting NMT Architectures with Linguistic Features (60 minutes)
    - Linguistic Analysis and Tools (Manning et al., 2014; Qi et al., 2020; Kondratyuk and Straka, 2019; Dyer et al., 2016; Kitaev and Klein, 2018).
    - Augmented input feature (Sennrich and Haddow, 2016; Chakrabarty et al., 2020; Chakrabarty et al., 2022; Chakrabarty et al., 2023; Currey and Heafield, 2018; Currey and Heafield, 2019).
    - Tree encoder that encode sentence in hierarchical manner (Eriguchi et al., 2016; Chen et al., 2017; Li et al., 2017).
    - Syntax-aware representation (Niehues and Cho, 2017; Zhang et al., 2019).
    - Syntax-aware self-attention (Hao et al., 2019; Bugliarello and Okazaki, 2020; Pu and Sima'an, 2022).

4. Linguistically Aware Decoding (20 minutes)
    - Tree decoder where output are generated hierarchically (Eriguchi et al., 2017; Wang et al., 2018; Wu et al., 2017).
    - Linearized trees (Aharoni and Goldberg, 2017; Nădejde et al., 2017).
    - Structural template prediction (Yang et al., 2020; Li et al., 2023).

5. Linguistically Motivated Evaluation (20 minutes)

- A fine-grained benchmark covering more than 100 linguistic phenomena (Macketanz et al., 2021; Avramidis and Macketanz, 2022).
- Analysis of specific linguistic phenomena (Müller et al., 2018; Voita et al., 2018; Voita et al., 2019; Adebara et al., 2022).
- Linguistic analysis of LLMs (GPT-4, BLOOM, LlaMa) (Manakhimova et al., 2023) .

6. Limitations and Future Directions (10 minutes)
   - Languages without proper linguistic analysis tools.
   - Application to high-resource languages in the era of LLMs.
7. Summary and Conclusion (5 minutes)
8. Discussion and Q/A (15 minutes)

Total time 180 minutes (excluding break)

Type of the Tutorial Cutting-edge

Target Audience and Size MT researchers and engineers, especially those interested in low-resource MT. 20–40 people.

Prerequisites This tutorial is primarily aimed at researchers who have a basic understanding of MT.

Reading List

- NMT architecture (Vaswani et al., 2017).
- Linguistic knowledge as input features (Sennrich and Haddow, 2016).

Diversity Considerations This tutorial covers improving MT for low-resource language pairs. Presenters have diverse backgrounds with different native languages, some of which are low-resourced ones. Our instructor will promote this tutorial on social media to diversify our audience participation.

Special Requirements N/A

Ethical Considerations We do not anticipate any ethical issues particularly regarding the topic of the tutorial. Nevertheless, training data and MT models may contain biases.

4 Tutorial Instructors

Haiyue Song (haiyue.song@nict.go.jp) is a Ph.D. student at Kyoto University and a technical researcher at the Advanced Translation Technology Laboratory, National Institute of Information and Communications Technology (NICT), Japan. He received his B.S. from Shanghai Jiao Tong University and his M.S. from Kyoto University. His research interests include machine translation, large language models, subword segmentation, and decoding algorithms. He has MT and LLMs related publications in TALLIP, AACL, LREC, ACL, and EMNLP. Additional information is available at `https://shyyhs.github.io/`.

Hour Kaing (hour_kaing@nict.go.jp) is a researcher at the Advanced Translation Technology Laboratory, National Institute of Information and Communications Technology (NICT), Japan. He received his B.S. from Institute of Technology of Cambodia, Cambodia, his M.Sc from University of Grenoble 1, France, and his Ph.D. from NARA Institute of Science and Technology, Japan. He is interested in linguistic analysis, low-resource machine translation, language modeling, and speech processing. He has publications in TALLIP, EACL, PACLIC, LREC, and IWSLT. Additional information is available at `https://hour.github.io/`.

Raj Dabre (raj.dabre@nict.go.jp) is a senior researcher at the Advanced Translation Technology Laboratory, National Institute of Information and Communications Technology (NICT), Japan and an Adjunct Faculty at IIT Madras, India. He received his Ph.D. from Kyoto University and Masters from IIT Bombay. His primary interests are in low-resource NLP, language modeling and efficiency. He has published in ACL, EMNLP, NAACL, TMLR, AAAI, AACL, IJCNLP and CSUR. Additional information is available at `https://prajdabre.github.io/`.

References

[Adebara et al.2022] Adebara, Ife, Muhammad Abdul-Mageed, and Miikka Silfverberg. 2022. Linguistically-motivated Yorùbá-English machine translation. In Calzolari, Nicoletta, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia

Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5066–5075, Gyeongju, Republic of Korea, October. International Committee on Computational Linguistics.

[Aharoni and Goldberg2017] Aharoni, Roee and Yoav Goldberg. 2017. Towards string-to-tree neural machine translation. In Barzilay, Regina and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 132–140, Vancouver, Canada, July. Association for Computational Linguistics.

[Amrhein and Sennrich2020] Amrhein, Chantal and Rico Sennrich. 2020. On Romanization for model transfer between scripts in neural machine translation. In Cohn, Trevor, Yulan He, and Yang Liu, editors, *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2461–2469, Online, November. Association for Computational Linguistics.

[Ataman et al.2017] Ataman, Duygu, Matteo Negri, Marco Turchi, and Marcello Federico. 2017. Linguistically motivated vocabulary reduction for neural machine translation from turkish to english.

[Avramidis and Macketanz2022] Avramidis, Eleftherios and Vivien Macketanz. 2022. Linguistically motivated evaluation of machine translation metrics based on a challenge set. In Koehn, Philipp, Loïc Barrault, Ondřej Bojar, Fethi Bougares, Rajen Chatterjee, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Alexander Fraser, Markus Freitag, Yvette Graham, Roman Grundkiewicz, Paco Guzman, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Tom Kocmi, André Martins, Makoto Morishita, Christof Monz, Masaaki Nagata, Toshiaki Nakazawa, Matteo Negri, Aurélie Névéol, Mariana Neves, Martin Popel, Marco Turchi, and Marcos Zampieri, editors, *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 514–529, Abu Dhabi, United Arab Emirates (Hybrid), December. Association for Computational Linguistics.

[Bahdanau et al.2016] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2016. Neural machine translation by jointly learning to align and translate. May.

[Batsuren et al.2021] Batsuren, Khuyagbaatar, Gábor Bella, and Fausto Giunchiglia. 2021. MorphyNet: a large multilingual database of derivational and inflectional morphology. In Nicolai, Garrett, Kyle Gorman, and Ryan Cotterell, editors, *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 39–48, Online, August. Association for Computational Linguistics.

[Brahma et al.2023] Brahma, Maharaj, Kaushal Maurya, and Maunendra Desarkar. 2023. SelectNoise: Unsupervised noise injection to enable zero-shot machine translation for extremely low-resource languages. In Bouamor, Houda, Juan Pino, and Kalika Bali, editors, *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1615–1629, Singapore, December. Association for Computational Linguistics.

[Bugliarello and Okazaki2020] Bugliarello, Emanuele and Naoaki Okazaki. 2020. Enhancing machine translation with dependency-aware self-attention. In Jurafsky, Dan, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1618–1627, Online, July. Association for Computational Linguistics.

[Chakrabarty et al.2020] Chakrabarty, Abhisek, Raj Dabre, Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2020. Improving low-resource NMT through relevance based linguistic features incorporation. In Scott, Donia, Nuria Bel, and Chengqing Zong, editors, *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4263–4274, Barcelona, Spain (Online), December. International Committee on Computational Linguistics.

[Chakrabarty et al.2022] Chakrabarty, Abhisek, Raj Dabre, Chenchen Ding, Hideki Tanaka, Masao Utiyama, and Eiichiro Sumita. 2022. FeatureBART: Feature based sequence-to-sequence pre-training for low-resource NMT. In Calzolari, Nicoletta, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics*, pages 5014–5020, Gyeongju, Republic of Korea, October. International Committee on Computational Linguistics.

[Chakrabarty et al.2023] Chakrabarty, Abhisek, Raj Dabre, Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2023. Low-resource multilingual neural translation using linguistic feature based relevance mechanisms. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[Chen et al.2017] Chen, Huadong, Shujian Huang, David Chiang, and Jiajun Chen. 2017. Improved neural machine translation with a syntax-aware encoder and decoder. In Barzilay, Regina and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1936–1945, Vancouver, Canada, July. Association for Computational Linguistics.

[Currey and Heafield2018] Currey, Anna and Kenneth Heafield. 2018. Multi-source syntactic neural machine translation. In Riloff, Ellen, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2961–2966, Brussels, Belgium, October-November. Association for Computational Linguistics.

[Currey and Heafield2019] Currey, Anna and Kenneth Heafield. 2019. Incorporating source syntax into transformer-based neural machine translation. In Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, André Martins, Christof Monz, Matteo Negri, Aurélie Névéol, Mariana Neves, Matt Post, Marco Turchi, and Karin Verspoor, editors, *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 24–33, Florence, Italy, August. Association for Computational Linguistics.

[Dabre et al.2021] Dabre, Raj, Himani Shrotriya, Anoop Kunchukuttan, Ratish Puduppully, Mitesh M. Khapra, and Pratyush Kumar. 2021. Indicbart: A pre-trained model for natural language generation of indic languages.

[Dyer et al.2016] Dyer, Chris, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. 2016. Recurrent neural network grammars. In Knight, Kevin, Ani Nenkova, and Owen Rambow, editors, *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 199–209, San Diego, California, June. Association for Computational Linguistics.

[Eriguchi et al.2016] Eriguchi, Akiko, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2016. Tree-to-sequence attentional neural machine translation. In Erk, Katrin and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 823–833, Berlin, Germany, August. Association for Computational Linguistics.

[Eriguchi et al.2017] Eriguchi, Akiko, Yoshimasa Tsuruoka, and Kyunghyun Cho. 2017. Learning to parse and translate improves neural machine translation. In Barzilay, Regina and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 72–78, Vancouver, Canada, July. Association for Computational Linguistics.

[Gala et al.2023] Gala, Jay, Pranjal A Chitale, A K Raghavan, Varun Gumma, Sumanth Doddapaneni, Aswanth Kumar M, Janki Atul Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M Khapra, Raj Dabre, and Anoop Kunchukuttan. 2023. Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages. *Transactions on Machine Learning Research*.

[Hao et al.2019] Hao, Jie, Xing Wang, Shuming Shi, Jinfeng Zhang, and Zhaopeng Tu. 2019. Multi-granularity self-attention for neural machine translation. In Inui, Kentaro, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *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 887–897, Hong Kong, China, November. Association for Computational Linguistics.

[He et al.2020] He, Xuanli, Gholamreza Haffari, and Mohammad Norouzi. 2020. Dynamic programming encoding for subword segmentation in neural machine translation. In Jurafsky, Dan, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3042–3051, Online, July. Association for Computational Linguistics.

[Husain et al.2024] Husain, Jaavid Aktar, Raj Dabre, Aswanth Kumar, Jay Gala, Thanmay Jayakumar, Ratish Puduppully, and Anoop Kunchukuttan. 2024. Romansetu: Efficiently unlocking multilingual capabilities of large language models models via romanization.

[Joshi et al.2024] Joshi, Aditya, Raj Dabre, Diptesh Kanojia, Zhuang Li, Haolan Zhan, Gholamreza Haffari, and Doris Dippold. 2024. Natural language processing for dialects of a language: A survey.

[Kitaev and Klein2018] Kitaev, Nikita and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia, July. Association for Computational Linguistics.

[Kondratyuk and Straka2019] Kondratyuk, Dan and Milan Straka. 2019. 75 languages, 1 model: Parsing Universal Dependencies universally. In Inui, Kentaro, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *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 2779–2795, Hong Kong, China, November. Association for Computational Linguistics.

[Kudo and Richardson2018] Kudo, Taku and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In Blanco, Eduardo and Wei Lu, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium, November. Association for Computational Linguistics.

[Li et al.2017] Li, Junhui, Deyi Xiong, Zhaopeng Tu, Muhua Zhu, Min Zhang, and Guodong Zhou. 2017.

Modeling source syntax for neural machine translation. In Barzilay, Regina and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 688–697, Vancouver, Canada, July. Association for Computational Linguistics.

[Li et al.2023] Li, Yafu, Leyang Cui, Jianhao Yan, Yongjing Yin, Wei Bi, Shuming Shi, and Yue Zhang. 2023. Explicit syntactic guidance for neural text generation. In Rogers, Anna, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14095–14112, Toronto, Canada, July. Association for Computational Linguistics.

[Macketanz et al.2021] Macketanz, Vivien, Eleftherios Avramidis, Shushen Manakhimova, and Sebastian Möller. 2021. Linguistic evaluation for the 2021 state-of-the-art machine translation systems for German to English and English to German. In Barrault, Loic, Ondrej Bojar, Fethi Bougares, Rajen Chatterjee, Marta R. Costa-jussa, Christian Federmann, Mark Fishel, Alexander Fraser, Markus Freitag, Yvette Graham, Roman Grundkiewicz, Paco Guzman, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Tom Kocmi, Andre Martins, Makoto Morishita, and Christof Monz, editors, *Proceedings of the Sixth Conference on Machine Translation*, pages 1059–1073, Online, November. Association for Computational Linguistics.

[Manakhimova et al.2023] Manakhimova, Shushen, Eleftherios Avramidis, Vivien Macketanz, Ekaterina Lapshinova-Koltunski, Sergei Bagdasarov, and Sebastian Möller. 2023. Linguistically motivated evaluation of the 2023 state-of-the-art machine translation: Can ChatGPT outperform NMT? In Koehn, Philipp, Barry Haddow, Tom Kocmi, and Christof Monz, editors, *Proceedings of the Eighth Conference on Machine Translation*, pages 224–245, Singapore, December. Association for Computational Linguistics.

[Manning et al.2014] Manning, Christopher, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In Bontcheva, Kalina and Jingbo Zhu, editors, *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland, June. Association for Computational Linguistics.

[Maurya et al.2024] Maurya, Kaushal Kumar, Rahul Kejriwal, Maunendra Sankar Desarkar, and Anoop Kunchukuttan. 2024. Charspan: Utilizing lexical similarity to enable zero-shot machine translation for extremely low-resource languages.

[Müller et al.2018] Müller, Mathias, Annette Rios, Elena Voita, and Rico Sennrich. 2018. A large-scale test set for the evaluation of context-aware pronoun translation in neural machine translation. In Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Christof Monz, Matteo Negri, Aurélie Névéol, Mariana Neves, Matt Post, Lucia Specia, Marco Turchi, and Karin Verspoor, editors, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 61–72, Brussels, Belgium, October. Association for Computational Linguistics.

[Niehues and Cho2017] Niehues, Jan and Eunah Cho. 2017. Exploiting linguistic resources for neural machine translation using multi-task learning. In Bojar, Ondřej, Christian Buck, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, and Julia Kreutzer, editors, *Proceedings of the Second Conference on Machine Translation*, pages 80–89, Copenhagen, Denmark, September. Association for Computational Linguistics.

[Nădejde et al.2017] Nădejde, Maria, Siva Reddy, Rico Sennrich, Tomasz Dwojak, Marcin Junczys-Dowmunt, Philipp Koehn, and Alexandra Birch. 2017. Predicting target language ccg supertags improves neural machine translation. In *Proceedings of the Second Conference on Machine Translation*, pages 68–79.

[Pu and Sima'an2022] Pu, Dongqi and Khalil Sima'an. 2022. Passing parser uncertainty to the transformer: Labeled dependency distributions for neural machine translation. In Moniz, Helena, Lieve Macken, Andrew Rufener, Loïc Barrault, Marta R. Costa-jussà, Christophe Declercq, Maarit Koponen, Ellie Kemp, Spyridon Pilos, Mikel L. Forcada, Carolina Scarton, Joachim Van den Bogaert, Joke Daems, Arda Tezcan, Bram Vanroy, and Margot Fonteyne, editors, *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 41–50, Ghent, Belgium, June. European Association for Machine Translation.

[Qi et al.2020] Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In Celikyilmaz, Asli and Tsung-Hsien Wen, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online, July. Association for Computational Linguistics.

[Sennrich and Haddow2016] Sennrich, Rico and Barry Haddow. 2016. Linguistic input features improve neural machine translation. In Bojar, Ondřej, Christian Buck, Rajen Chatterjee, Christian Federmann, Liane Guillou, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Aurélie Névéol, Mariana Neves, Pavel Pecina, Martin Popel, Philipp Koehn, Christof Monz, Matteo Negri, Matt Post, Lucia Specia, Karin

Verspoor, Jörg Tiedemann, and Marco Turchi, editors, *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, pages 83–91, Berlin, Germany, August. Association for Computational Linguistics.

[Song et al.2020] Song, Haiyue, Raj Dabre, Zhuoyuan Mao, Fei Cheng, Sadao Kurohashi, and Eiichiro Sumita. 2020. Pre-training via leveraging assisting languages for neural machine translation. In Rijhwani, Shruti, Jiangming Liu, Yizhong Wang, and Rotem Dror, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 279–285, Online, July. Association for Computational Linguistics.

[Song et al.2022] Song, Haiyue, Raj Dabre, Zhuoyuan Mao, Chenhui Chu, and Sadao Kurohashi. 2022. BERTSeg: BERT based unsupervised subword segmentation for neural machine translation. In He, Yulan, Heng Ji, Sujian Li, Yang Liu, and Chua-Hui Chang, editors, *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 85–94, Online only, November. Association for Computational Linguistics.

[Tolmachev et al.2018] Tolmachev, Arseny, Daisuke Kawahara, and Sadao Kurohashi. 2018. Juman++: A morphological analysis toolkit for scriptio continua. In Blanco, Eduardo and Wei Lu, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 54–59, Brussels, Belgium, November. Association for Computational Linguistics.

[Vaswani et al.2017] Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In Guyon, I., U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

[Voita et al.2018] Voita, Elena, Pavel Serdyukov, Rico Sennrich, and Ivan Titov. 2018. Context-aware neural machine translation learns anaphora resolution. In Gurevych, Iryna and Yusuke Miyao, editors, *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1264–1274, Melbourne, Australia, July. Association for Computational Linguistics.

[Voita et al.2019] Voita, Elena, Rico Sennrich, and Ivan Titov. 2019. When a good translation is wrong in context: Context-aware machine translation improves on deixis, ellipsis, and lexical cohesion. In Korhonen, Anna, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1198–1212, Florence, Italy, July. Association for Computational Linguistics.

[Wang et al.2018] Wang, Xinyi, Hieu Pham, Pengcheng Yin, and Graham Neubig. 2018. A tree-based decoder for neural machine translation. In Riloff, Ellen, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4772–4777, Brussels, Belgium, October-November. Association for Computational Linguistics.

[Wu et al.2017] Wu, Shuangzhi, Dongdong Zhang, Nan Yang, Mu Li, and Ming Zhou. 2017. Sequence-to-dependency neural machine translation. In Barzilay, Regina and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 698–707, Vancouver, Canada, July. Association for Computational Linguistics.

[Yang et al.2020] Yang, Jian, Shuming Ma, Dongdong Zhang, Zhoujun Li, and Ming Zhou. 2020. Improving neural machine translation with soft template prediction. In Jurafsky, Dan, Joyce Chai, Natalie Schluter, and Joel Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5979–5989, Online, July. Association for Computational Linguistics.

[Zhang and Komachi2018] Zhang, Longtu and Mamoru Komachi. 2018. Neural machine translation of logographic language using sub-character level information. In Bojar, Ondřej, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Christof Monz, Matteo Negri, Aurélie Névéol, Mariana Neves, Matt Post, Lucia Specia, Marco Turchi, and Karin Verspoor, editors, *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 17–25, Brussels, Belgium, October. Association for Computational Linguistics.

[Zhang et al.2019] Zhang, Meishan, Zhenghua Li, Guohong Fu, and Min Zhang. 2019. Syntax-enhanced neural machine translation with syntax-aware word representations. In Burstein, Jill, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1151–1161, Minneapolis, Minnesota, June. Association for Computational Linguistics.