

# Multitask Models for Controlling the Complexity of Neural Machine Translation

**Sweta Agrawal**

Department of Computer Science  
University of Maryland  
sweagraw@cs.umd.edu

**Marine Carpuat**

Department of Computer Science  
University of Maryland  
marine@cs.umd.edu

## Abstract

We introduce a machine translation task where the output is aimed at audiences of different levels of target language proficiency. We collect a novel dataset of news articles available in English and Spanish and written for diverse reading grade levels. We leverage this dataset to train multitask sequence to sequence models that translate Spanish into English targeted at an easier reading grade level than the original Spanish. We show that multitask models outperform pipeline approaches that translate and simplify text independently.<sup>1</sup>

## 1 Introduction

Generating text at the right level of complexity is important to make machine translation (MT) more accessible to non-native speakers, language learners (Petersen and Ostendorf, 2007; Allen, 2009) or people who suffer from language impairments (Carroll et al., 1999; Canning et al., 2000; Inui et al., 2003). Simplification has been used to improve MT by restructuring complex sentences into shorter and simpler segments that are easier to translate (Gerber and Hovy, 1998; Štajner and Popovic, 2016; Hasler et al., 2017). Closest to our goal, Marchisio et al. (2019) address the task of producing either simple or complex translations of the same input, using automatic readability scoring of parallel corpora. Our work shares their goal of controlling translation complexity, but considers a broader range of reading grade levels and simplification operations grounded in professionally edited text simplification corpora.

We collect examples of Spanish sentences paired with several English translations that span a range of complexity levels from the Newsela website, which provides professionally edited simplifications and translations. While Newsela dataset has been used to build English text simplification systems (Xu et al., 2016; Zhang and Lapata, 2017; Scarton and Specia, 2018; Nishihara et al., 2019; Zhong et al., 2020) and Spanish simplification systems (Štajner et al., 2018), we exploit the document level alignment between English and Spanish articles to construct evaluation and training samples for complexity controlled MT. By contrast with MT parallel corpora, the English and Spanish translations at different grade levels are only comparable. We adopt a multitask approach that trains a single encoder-decoder model to perform the two distinct tasks of machine translation and text simplification and evaluate it on Spanish-English complexity controlled MT. Our empirical study shows that multitask models produce better and simpler translations than pipelines of independent translation and simplification models. Scripts to replicate our model configurations and our cross-lingual segment aligner are available at <https://github.com/sweta20/ComplexityControlledMT>.

## 2 A Multitask Approach to Complexity Controlled MT

We define **complexity controlled MT** as a task that takes two inputs: an input language segment  $s_i$  and a target complexity  $c$  representing the desired reading grade level of the output. The goal is to generate a translation  $s_o$  in the output language with complexity  $c$ .

<sup>1</sup>This paper is an abridged version of our work (Agrawal and Carpuat, 2019).

We model  $P(s_o|s_i, c; \theta)$  as a neural encoder-decoder with attention (Bahdanau et al., 2015). Target complexity  $c$  is incorporated as a special token appended to the beginning of the input sequence, which acts as a side constraint. Our multitask training configuration lets us exploit different types of training examples to train shared encoder-decoder parameters  $\theta$ . We use the following samples/tasks:

- Complexity controlled MT samples  $(s_i, c_o, s_o)$ : These are the closest samples to the task at hand, but are hard to obtain. They are used to define the complexity-controlled MT loss,  $\mathcal{L}_{CMT} = \sum_{(s_i, c_o, s_o)} \log P(s_o|s_i, c_o; \theta)$ .
- MT samples  $(s_i, s_o)$ : These are sentence pairs drawn from parallel corpora. They are available in large quantities for many language pairs (Tiedemann, 2012) and are used to define the MT loss,  $\mathcal{L}_{MT} = \sum_{(s_i, s_o)} \log P(s_o|s_i; \theta)$ .
- Text simplification samples in the MT target language  $(s_o, c_{s'_o}, s'_o)$  where  $s'_o$  is a simplified version of complexity  $c_{s'_o}$  for input  $s_o$ , which are likely to be available in much smaller quantities than MT samples and are used to define the simplification loss,  $\mathcal{L}_{Simplify} = \sum_{(s_o, c_{s'_o}, s'_o)} \log P(s'_o|s_o, c_{s'_o}; \theta)$ .

The multitask loss is simply obtained by summing the losses from individual tasks:  $\mathcal{L}_{CMT} + \mathcal{L}_{MT} + \mathcal{L}_{Simplify}$ .

### 3 Evaluation and Discussion

We extract segment pairs from the Newsela corpus which consists of English articles in their original form, 4 or 5 different versions re-written by professionals to suit different grade levels as well as optional translations of original and/or simplified English articles into Spanish. English and Spanish segments are aligned across complexity levels to create a bilingual dataset for training and evaluating complexity controlled MT systems. We report results using both translation and simplification metrics: BLEU (Papineni et al., 2002), SARI (Xu et al., 2016) and the Pearson Correlation between the complexity of NMT output and of reference translations (Heilman et al., 2008), where the reading grade level complexity of MT outputs and reference translations is estimated using the Automatic Readability Index (ARI).<sup>2</sup>

We contrast the multitask system with pipeline based approaches, where translation and simplification are treated as independent consecutive steps. In the first pipeline setup, the output from the translation model is fed as input to an English simplification model while in the second, the output from the Spanish simplification model is fed as input to a translation model. Table 1 shows that compared to pipeline models, multitask models generate translations that better match human references according to BLEU. SARI suggests that multitask translations are simpler than baseline translations, and their resulting complexity correlates better with reference grade levels according to PCC. The ‘‘All tasks’’ model highlights the strengths of the multitask approach: combining training samples from many tasks yields improvements over the ‘‘Translate and Simplify’’ multitask model which is trained on the exact same data as the pipelines. However, even without additional training data, the multitask ‘‘Translate and Simplify’’ model improves over baselines mainly by simplifying the output more, which suggests that the simplification component of the multitask model benefits from the additional MT training data. Table 2 illustrates simplification operations observed for a fixed grade 12 Spanish input into English with target grade levels ranging from 9 to 3. For lower grade levels such as 7 and 5, paraphrasing (e.g. ‘‘inaugurar’’

Complexity cont. MT	BLEU	SARI	PCC
<i>Pipeline Baselines</i>			
Translate then Simplify	21.98	30.4	0.436
Simplify then Translate	17.09	37.4	0.275
<i>Multitask Models</i>			
Translate and Simplify	22.51	44.8	0.572
All Tasks	<b>22.75</b>	<b>45.0</b>	<b>0.608</b>

Table 1: Compared to pipeline models, multitask models produce complexity controlled translations that better match human references (BLEU), that are simpler (SARI), and whose resulting complexity correlates better with the target grade level (PCC).

<sup>2</sup><https://github.com/mmautner/readability>

is translated as “set to open”) and sentence splitting is observed. For the simplest grade level, the model deletes additional content such as “authoritations” and “historical”.

---

12	Ahora el museo Mauritshuis está por inaugurar una exposición dedicada a los autorretratos del siglo XVII, que destaca las similitudes y diferencias entre las fotos modernas y las obras de arte históricas.
9	Now the museum Mauritois is launching an exhibition dedicated to the 18th century authoritations, highlighting the similarities and differences between modern photos and historical artworks.
7	The museum is <b>now set to open</b> an exhibition dedicated to the 18th century authoritations, highlighting the similarities and differences between modern photos and historical artworks.
5	The museum is now set to open an exhibit dedicated to the 18th century <u>authoritations</u> . <b>It highlights</b> the similarities and differences between modern photos and historical artworks.
3	The museum is now set to open an exhibit dedicated to the 18th century. It <b>shows</b> the similarities and differences between modern photos and <del>historical</del> art works.

---

Table 2: Example of multi-task model outputs when translating grade 12 Spanish into increasingly simpler English: the multi-task model performs a wide range of simplification operations including sentence splitting, **paraphrasing** and deletion.

However, even when simplifying translations, multitask models are not yet able to exactly match the desired complexity level, and the gap between the complexity achieved and the target complexity increases with the amount of simplification required. Our datasets and models thus provide a foundation to investigate strategies for a tighter control on output complexity in future work using training objectives that explicitly addresses this gap or via modelling the type of lexical and syntactic operations performed when simplifying to a grade level.

## References

- Sweta Agrawal and Marine Carpuat. 2019. Controlling text complexity in neural machine translation. 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 1549–1564.
- David Allen. 2009. A study of the role of relative clauses in the simplification of news texts for learners of english. *System*, 37(4):585–599.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *International Conference on Learning Representations (ICLR)*.
- Yvonne Canning, John Tait, Jackie Archibald, and Ros Crawley. 2000. Cohesive generation of syntactically simplified newspaper text. In *International Workshop on Text, Speech and Dialogue*, pages 145–150. Springer.
- John Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999. Simplifying text for language-impaired readers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*.
- Laurie Gerber and Eduard Hovy. 1998. Improving Translation Quality by Manipulating Sentence Length. In *Conference of the Association for Machine Translation in the Americas (AMTA)*.
- Eva Hasler, Adrià de Gispert, Felix Stahlberg, Aurelien Waite, and Bill Byrne. 2017. Source sentence simplification for statistical machine translation. *Computer Speech & Language*, 45(Supplement C):221–235, September.
- Michael Heilman, Kevyn Collins-Thompson, and Maxine Eskenazi. 2008. An analysis of statistical models and features for reading difficulty prediction. In *Proceedings of the third workshop on innovative use of NLP for building educational applications*, pages 71–79. Association for Computational Linguistics.
- Kentaro Inui, Atsushi Fujita, Tetsuro Takahashi, Ryu Iida, and Tomoya Iwakura. 2003. Text simplification for reading assistance: A project note. In *Proceedings of the Second International Workshop on Paraphrasing*, pages 9–16, Sapporo, Japan, July. Association for Computational Linguistics.

- Kelly Marchisio, Jialiang Guo, Cheng-I Lai, and Philipp Koehn. 2019. Controlling the reading level of machine translation output. In *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*, pages 193–203.
- Daiki Nishihara, Tomoyuki Kajiwara, and Yuki Arase. 2019. Controllable text simplification with lexical constraint loss. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 260–266.
- 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*, Philadelphia, PA, July.
- Sarah E Petersen and Mari Ostendorf. 2007. Text simplification for language learners: a corpus analysis. In *Workshop on Speech and Language Technology in Education*.
- Carolina Scarton and Lucia Specia. 2018. Learning Simplifications for Specific Target Audiences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 712–718.
- Sanja Štajner and Maja Popovic. 2016. Can text simplification help machine translation? In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*, pages 230–242.
- Sanja Štajner, Marc Franco-Salvador, Paolo Rosso, and Simone Paolo Ponzetto. 2018. Cats: A tool for customized alignment of text simplification corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*.
- Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Ugur Dogan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.
- Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark, September. Association for Computational Linguistics.
- Yang Zhong, Chao Jiang, Wei Xu, and Junyi Jessy Li. 2020. Discourse level factors for sentence deletion in text simplification. *ArXiv*, abs/1911.10384.