Predicting Sentence Deletions for Text Simplification Using a Functional Discourse Structure

Bohan Zhang[*]
University of Michigan
zbohan@umich.edu

Prafulla Kumar Choubey
Salesforce Research
pchoubey@salesforce.com

Ruihong Huang
Texas A&M University
huangrh@tamu.edu

Abstract

Document-level text simplification often deletes some sentences besides performing lexical, grammatical or structural simplification to reduce text complexity. In this work, we focus on sentence deletions for text simplification and use a news genre-specific functional discourse structure, which categorizes sentences based on their contents and their function roles in telling a news story, for predicting sentence deletion. We incorporate sentence categories into a neural net model in two ways for predicting sentence deletions, either as additional features or by jointly predicting sentence deletions and sentence categories. Experimental results using human-annotated data show that incorporating the functional structure improves the recall of sentence deletion prediction by 6.5% and 10.7% respectively using the two methods, and improves the overall F1-score by 3.6% and 4.3% respectively.

1 Introduction

Text simplification aims to rewrite complex texts in order to make them easier to read and understand. This task can benefit vast low literacy readers, including children, language learners and people with aphasia, and has recently attracted increasing attention from the research community (Xu et al., 2016; Zhao et al., 2018; Martin et al., 2019; Dong et al., 2019). However, most previous research has focused on sentence-level text simplification and aim to simplify one sentence at a time. As a result, few discourse-level phenomena have been examined or understood for achieving document-level text simplification.

Sentence deletion is a commonly used strategy to achieve intense simplification (Drndarevic and Saggion, 2012; Woodsend and Lapata, 2011), i.e., some less important sentences from an original article are simply deleted and ignored for simplification. While professional re-writers may consider many factors and use several measures of *importance* to decide if a sentence should be deleted, some discourse structures provide automated measures to derive *importance* for sentences in a document. In particular, functional discourse structures categorize text units (sentences or paragraphs) based on their contents and their function roles in serving the purpose of a specific text-genre, such as scientific papers (Teufel et al., 1999; Liakata et al., 2012) and news articles (Yarlott et al., 2018; Choubey et al., 2020), and are therefore, expected to directly reveal the *importance* of a sentence within a document.

In this work, we explore the use of news genre-specific functional structures for predicting sentence deletions in news documents. Specifically, we use news discourse profiling structure, which categorizes contents of news articles around the main news event, constructed through a publicly available system (Choubey et al., 2020)[1]. This system labels each sentence with one of eight content types reflecting common discourse roles of a sentence in telling a news story, including two content types for sentences describing the main news event and its immediate consequences (*main content*), two content types for sentences providing *context-informing contents* and four content types for sentences providing *further supportive information* in a news article.

We perform experiments using the Newsela corpus (Xu et al., 2015), a widely used dataset for text simplification research that contains 1492 English news articles and four simplified versions for each news article targeting audience of different reading levels (from elementary to high school students). Since we aim to achieve maximal simplification, we predict sentence deletions for tar-

[*]Most work was done while Bohan was a summer intern in the NLP lab at Texas A&M University.

[1]This system can be found here: https://github.com/prafulla77/Discourse_Profiling.

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Volume 2: Short Papers, pages 255 - 261

get reading level corresponding to the elementary school students. We first build a document-level neural network as the basic model for predicting sentence deletions. We then incorporate content types of sentences into the prediction system using two methods, 1) by using content type labels as additional features to enrich sentence representations, and 2) by jointly predicting both sentence deletion labels and discourse content type labels. Experimental results show that, with little to no drop on precision, both methods for incorporating sentence content type information improve the recall (F1 score) on the sentence deletion prediction task by 6.5% (3.6%) and 10.7% (4.3%) respectively. Analysis on the development set shows that the additional deletions correctly recognized by our system are all sentences providing context-informing or supportive contents.

2 Related Work

The previous research on text simplification has focused on word or phrase level simplification (Yatskar et al., 2010; Biran et al., 2011; Specia et al., 2012; Paetzold and Specia, 2017), or sentence-level simplification (Wubben et al., 2012; Sutskever et al., 2014; Nisioi et al., 2017; Zhao et al., 2018; Dong et al., 2019), few research has been conducted for document-level text simplification.

Sentence deletion, as an interesting phenomenon for document-level text simplification, has been studied in several pilot studies. (Petersen and Ostendorf, 2007) conducted a corpus analysis and showed that sentence position and content influence sentence deletion or retention. The recent pilot research for sentence deletion prediction (Zhong et al., 2019) considers sentence position in a document, document length and topic, as well as exploits rhetorical discourse structures that capture text coherence in general and can be used to derive the *salience* of a sentence in a discourse. However, while sentence position and the two document characteristics are shown useful for sentence deletion prediction, discourse features based on rhetorical discourse structures are shown to have little impact for this task. Compared to general rhetorical discourse structures that do not consider genre specialties, the genre-specific functional structure we examine in this paper can more directly reveal the importance of a sentence within a document.

3 The News Discourse Structure and Sentence Types

News discourse profiling (Choubey et al., 2020) categorizes sentences in news articles into eight schematic categories that describe the common discourse roles of sentences in telling a news story, following the news content schemata proposed by Van Dijk (Teun A, 1986; Van Dijk, 1988a,b). These eight sentence categories fall into three groups.

Main Contents: are the most relevant information of news articles, including sentences that introduce the main event as the major subjects of a news article (**Main Event**), and sentences that describe consequence events immediately triggered by the main event (**Consequence**).

Context Informing Contents: provide information of the actual situation in which main event occurred, including sentences that describe the recent events that act as possible causes or preconditions for the main event (**Previous Events**), and sentences that describe ongoing situation and other context informing contents (**Current Context**).

Additional Supportive Contents: contain the least relevant information, including sentences that describe past events that precede the main events in months and years (**Historical Event**), sentences that describe unverifiable situations, fictional or personal account of incidents of an unknown person (**Anecdotal Event**), opinionated contents that describe reactions from immediate participants, experts, known personalities as well as journalist or news source (**Evaluation**), and speculations on the possible consequences of the main or contextual events (**Expectation**)

3.1 Analysis of Deletions w.r.t Sentence Types

We conducted an analysis on deletion rate for each sentence category using the development set (Section 5.1) which was manually annotated with sentence deletion labels. The discourse content type labels of sentences were predicted by the news discourse profiling system (Choubey et al., 2020). Table 1 shows the results. We can see that **Main Event** sentences have the lowest deletion rate of 14.7%, much lower than other types of sentences. **Previous Event** sentences, as one type of context informing contents, have a relatively low deletion rate as well to provide necessary context, i.e., possible causes or preconditions, to understand the main news events. While additional supportive contents overall have a high deletion rate, **Anecdotal**

The NFL delivered that message in a resounding way Monday, suspending the New England Patriots star without pay for the first four games of next season for "conduct detrimental to the integrity of the NFL." (Main Event)

The punishment comes days after the league announced results of an investigation that found Brady was "likely generally aware" that equipment assistants employed by the team had conspired to deflate the Patriots' footballs for last season's AFC championship game, making the balls easier to throw and catch. (Previous Event)

The Patriots also were fined $1 million — equaling the largest in league history — and stripped of their first-round draft pick next year and a fourth-round selection in 2017. (Consequence)

The Pariots have been accused of cheating in the past, and in 2007 were caught breaking league rules by videotaping the sideline hand signals of New York Jets coaches. (Historical Event)

That incident, nicknamed Spygate, cost New England coach Bill Belichick $500,000 and the league docked the Patriots a first-round draft pick. (Historical Event)

By every indication, the incident has not dimmed Brady's star one iota among Patriots' fans. (Current Context)

He was cheered enthusiastically last week, one day after the Wells report was released, when he spoke at an event at Salem State University in Massachusetts. (Current Context)

Figure 1: An example article: Brady Deflategate. Sentences in purple were deleted for text simplification.

	Main Event	Consequence	Previous Event	Current Context	Historical Event	Anecdotal Event	Evaluation	Expectation
Deleted	5 (14.7)	0 (NA)	7 (31.8)	128 (37.5)	36 (46.2)	11 (27.5)	206 (41.2)	35 (33.7)
Retained	29 (85.3)	0 (NA)	15 (68.2)	213 (62.5)	42 (53.8)	29 (72.5)	294 (58.8)	69 (66.3)

Table 1: The number (percentage) of sentences in each type that are deleted or retained, on the development set. The news discourse profiling system did not label any sentence in the development set as Consequence, which is a minority class as revealed by (Choubey et al., 2020)

Event sentences have a low deletion rate, possibly because personal account of incidents present especially interesting contents for elementary students, the target group of our chosen simplification level.

Figure 1 shows an example document where both deleted sentences (colored in purple) are of one additional supportive content type, **Historical Event**.

4 Models

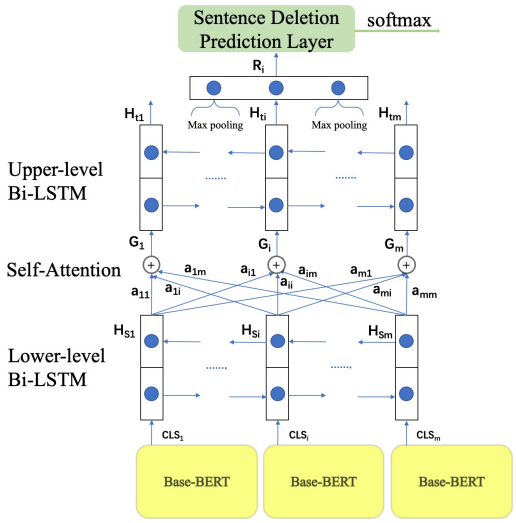

Figure 2: The Baseline Model

As a baseline model, (shown in Figure 2), we built a document-level neural network model to learn context aware sentence representations for predicting sentence deletions. Similar architectures have been shown useful for several other discourse-level tasks (Nallapati et al., 2016; Choubey et al., 2020).

Specifically, the model takes a document as input and has two document-level BiLSTM layers (Hochreiter and Schmidhuber, 1997) stacked up with a self-attention layer between them, to sufficiently exploit document wide contexts for building sentence representations. In addition, for each sentence, we further concatenate its sentence representation with two vectors obtained by max pooling over representations of its surrounding sentences (two sentences to each side), to obtain the final sentence representation R_i, that is better aware of the local context. We use a feed forward neural network with 1024-2 units to predict a binary label (deleted or not) for each sentence[2] based on its final sentence representation. We apply base BERT (Devlin et al., 2019) to obtain the initial sentence representations of 768 dimensions. Both BiLSTMs

[2]We also tried to add a CRF layer to capture deletion label dependencies between sentences, and predict labels for a sequence of sentences in a document, however, it did not improve the sentence deletion prediction performance.

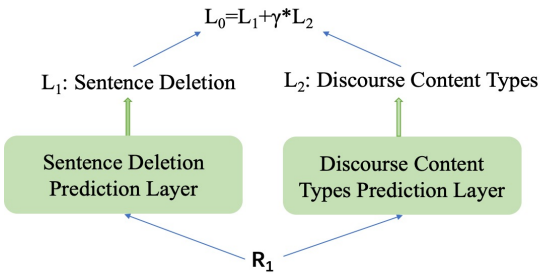

Figure 3: Jointly Predicting Two Types of Labels

have the hidden dimension size of 512.

Next, we present two methods to utilize the functional structure for sentence deletion prediction.

4.1 Feature Concatenation

For each sentence, we create a feature vector F_i with eight dimensions corresponding to the eight discourse content types[3], and values in the vector are probabilities of content types for the target sentence as output by the news discourse profiling system. We concatenate the feature vector F_i with the final sentence representation R_i and feed the concatenated vector to the sentence deletion prediction layer.

4.2 Joint Learning

Instead of creating features, we learn to jointly predict both sentence deletion labels and discourse content type labels (system predicted) using shared sentence representations (Figure 3). Specifically, we add a new prediction layer with 1024-9[4] units to predict discourse content types for sentences, and learn to jointly predict both types of labels by minimizing the aggregated loss of two tasks: $L_0 = L_1 + \gamma * L_2$, where L_1 is the cross-entropy loss for the sentence deletion prediction task and L_2 is the mean squared loss for the discourse content type prediction task[5].

[3]Document length and sentence position in a document have been shown useful for sentence deletion prediction in the previous work when used in a feature based approach (Zhong et al., 2019). We also concatenated these features with the final sentence representations. However, these features hurt the performance a little in our system, so we removed them. We suspect that document length and sentence position have been captured by the document-level neural net model and adding the features cause redundancies.

[4]Eight discourse content types plus one "Other" category.

[5]The mean squared loss is calculated against probabilities of content types for the target sentence as output by the news discourse profiling system.

5 Evaluation

5.1 Dataset

We conduct experiments using the Newsela corpus for text simplification (Xu et al., 2015). This corpus contains 1492 English news articles and four simplified versions for each article targeting students ranging from grade 2 to grade 12. In our study, we focus on predicting sentence deletions to achieve the relatively aggressive level of simplification that targets elementary school students (grades 2 to 5).

Test and Development Data: We created a new annotated dataset. The annotated dataset of 50 documents used in Zhong et al. (2019) was not released yet when we started to work on this project. Our code and the method to obtain our annotated dataset can be found on github[6].

Different from the crowd-sourcing based annotation method of Zhong et al. (2019) that decomposes the document-level sentence alignment task to a paragraph alignment task followed by a paragraph-level sentence alignment task, we ask our two annotators to read through a whole news article and its simplified article before annotating alignment sentence by sentence, which enables thorough annotations. Then, for each sentence in an original article, we instruct our annotators to align it with all the sentences in the simplified article that contain part or all of its contents (or paraphrases), one sentence in an original article will be labeled as "deleted" if *no* sentence in its simplified article is aligned with this sentence.

We annotated 95 (containing 4,334 sentences) randomly selected news articles. The two annotators first annotated five news articles (228 sentences) in common and achieved a high kappa agreement (Artstein and Poesio, 2008) of 0.911. Then, each of them annotated 45 more articles. We randomly selected 25 annotated articles and use them as the development set, and use the other 70 articles as the test set. 48% and 38% of sentences are annotated as deleted in the test and development sets respectively. We will publish our annotations.

Training Data: We create noisy supervision to train the systems by applying an automatic sentence alignment tool CATS[7] (Štajner et al., 2018) to the remaining 1397 unlabeled news articles and quickly obtained alignments between these news

[6]https://github.com/XMUBQ/SentenceDeletion

[7]CATS is a lexical similarity based sentence/paragraph alignment tool specifically designed for text simplification, and has been shown to perform well on the Newsela corpus.

	Current Context	Historical Event	Anecdotal Event	Evaluation	Expectation
Feature Concatenation	24	7	6	20	3
Joint Learning	20	3	3	21	1

Table 2: Numbers of additional deleted sentences from each content type that were correctly predicted. None of the correctly deleted sentences are from main event, consequence, and previous event content types.

Models	Dev Set	Test Set
FNN (Zhong et al., 2019)	44.6/60.4/51.3	56.7/57.2/57.0
Our Baseline	52.0/62.2/56.6	63.4/60.8/62.0
Feature Concatenation	**52.7**/64.8/58.1	**64.0**/67.3/65.6
Joint Learning	50.7/**69.8**/**58.7**	61.8/**71.5**/**66.3**

Table 3: Sentence deletion prediction results (P/R/F) (our dataset). Statistical significance tests show that compared with our baseline, both methods achieved significant improvements ($p<0.01$) in F1 measure.

Models	Dev Set	Test Set
FNN (Zhong et al., 2019)	61.7/60.7/61.0	56.8/60.6/58.6
Our Baseline	63.8/67.2/65.4	59.2/63.3/61.2
Feature Concatenation	69.7/**70.2**/70.0	**61.8**/66.1/63.9
Joint Learning	**70.9**/69.8/**70.4**	59.9/**68.6**/**63.9**

Table 4: Sentence deletion prediction results (P/R/F) (on the dataset from Zhong et al. (2019). Note that the results are not directly comparable with those in Zhong et al. (2019), as the training datasets are different. We used the Newsela corpus of a newer version and different automatic alignment tools to build our training dataset.

articles and their simplified articles. 82.11% of sentence alignments produced by CATS are correct when evaluated on our development set.

5.2 Experimental Settings

For regularization, we use dropout of 0.5 on the output of both BiLSTMs and the self-attention layer. We apply Adam optimizer (Kingma and Ba, 2014) for training, and the learning rate is set to 3e-4. All the neural models are trained for 15 epochs and we use the epoch yielding the best validation performance. We searched the hyper-parameter γ value over the range [0, 3] with a step size of 0.5, and its best value equals to 1.5.

5.3 Results and Analysis

In Table 3, we report the performance of our baseline and the two news discourse profiling structure-aware models. For better positioning of our work, we also re-implemented the model proposed in a recent work by Zhong et al. (2019), a feedforward neural network (FNN) model with sparse features[8]. First, our baseline system performs better than the feature based FNN model with 5.3% and 5.0% higher F1 score on validation and test datasets respectively. Then, both methods for incorporating discourse information have noticeably improved the performance on sentence deletion prediction. We also evaluate the models on the dataset from Zhong et al. (2019). As shown in Table 4, similar trends were observed on this dataset as well.

Since the performance gains of both discourse-aware models are mainly on recall, we analyze the distribution of additional deleted sentences correctly predicted by the two models. As shown in Table 2, the additional deleted sentences are either context informing contents or additional supportive contents, but none is main content. This observation corroborates our analysis in section 3.1.

6 Conclusion

We study sentence deletion prediction to achieve document-level text simplification. We have showed that a genre-specific functional discourse structure improves the prediction performance by large margins, when incorporated into a neural net model either as new features or for joint learning. For future work, we will study other useful discourse-level factors for sentence deletion prediction, we will also investigate multi-task learning to benefit both sentence deletion prediction and discourse parsing tasks.

7 Acknowledgements

We gratefully acknowledge support from National Science Foundation via the awards IIS-1942918. We would also like to thank the anonymous reviewers for their feedback.

[8]Some features have little impact to the performance, we only implemented the ones that have been shown useful in their ablation study, specifically, the document length and sentence position features. The model parameters and training settings were identical to the paper.

References

Ron Artstein and Massimo Poesio. 2008. Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Or Biran, Samuel Brody, and Noémie Elhadad. 2011. Putting it simply: a context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 496–501, Portland, Oregon, USA. Association for Computational Linguistics.

Prafulla Kumar Choubey, Aaron Lee, Ruihong Huang, and Lu Wang. 2020. Discourse as a function of event: Profiling discourse structure in news articles around the main event. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5374–5386, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yue Dong, Zichao Li, Mehdi Rezagholizadeh, and Jackie Chi Kit Cheung. 2019. EditNTS: An neural programmer-interpreter model for sentence simplification through explicit editing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3393–3402, Florence, Italy. Association for Computational Linguistics.

Biljana Drndarevic and Horacio Saggion. 2012. Reducing text complexity through automatic lexical simplification: An empirical study for spanish. *Procesamiento del lenguaje natural*, 49:13–20.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Maria Liakata, Shyamasree Saha, Simon Dobnik, Colin Batchelor, and Dietrich Rebholz-Schuhmann. 2012. Automatic recognition of conceptualization zones in scientific articles and two life science applications. *Bioinformatics*, 28(7):991–1000.

Louis Martin, Benoît Sagot, Éric de la Clergerie, and Antoine Bordes. 2019. Controllable sentence simplification. *arXiv preprint arXiv:1910.02677*.

Ramesh Nallapati, Bowen Zhou, Caglar Gulcehre, Bing Xiang, et al. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond. *arXiv preprint arXiv:1602.06023*.

Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P. Dinu. 2017. Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 85–91, Vancouver, Canada. Association for Computational Linguistics.

Gustavo Paetzold and Lucia Specia. 2017. Lexical simplification with neural ranking. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 34–40, Valencia, Spain. Association for Computational Linguistics.

Sarah E Petersen and Mari Ostendorf. 2007. Text simplification for language learners: a corpus analysis. In *Workshop on Speech and Language Technology in Education*.

Lucia Specia, Sujay Kumar Jauhar, and Rada Mihalcea. 2012. SemEval-2012 task 1: English lexical simplification. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 347–355, Montréal, Canada. Association for Computational Linguistics.

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)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3104–3112. Curran Associates, Inc.

Simone Teufel, Jean Carletta, and Marc Moens. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, Bergen, Norway. Association for Computational Linguistics.

Van Dijk Teun A. 1986. News schemata. *Studying writing: linguistic approaches*, 1:155–186.

Teun A Van Dijk. 1988a. News analysis. *Case Studies of International and National News in the Press. New Jersey: Lawrence*.

Teun A Van Dijk. 1988b. News as discourse. Hillsdale,NJ, US: Lawrence Erlbaum Associates, Inc.

Kristian Woodsend and Mirella Lapata. 2011. Learning to simplify sentences with quasi-synchronous grammar and integer programming. In *Proceedings of the*

2011 Conference on Empirical Methods in Natural Language Processing, pages 409–420, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Sander Wubben, Antal van den Bosch, and Emiel Krahmer. 2012. Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1015–1024, Jeju Island, Korea. Association for Computational Linguistics.

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

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.

W. Victor Yarlott, Cristina Cornelio, Tian Gao, and Mark Finlayson. 2018. Identifying the discourse function of news article paragraphs. In *Proceedings of the Workshop Events and Stories in the News 2018*, pages 25–33, Santa Fe, New Mexico, U.S.A. Association for Computational Linguistics.

Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2010. For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 365–368, Los Angeles, California. Association for Computational Linguistics.

Sanqiang Zhao, Rui Meng, Daqing He, Andi Saptono, and Bambang Parmanto. 2018. Integrating transformer and paraphrase rules for sentence simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3164–3173, Brussels, Belgium. Association for Computational Linguistics.

Yang Zhong, Chao Jiang, Wei Xu, and Junyi Jessy Li. 2019. Discourse level factors for sentence deletion in text simplification. *arXiv preprint arXiv:1911.10384*.