

Neural Text Simplification in Low-Resource Conditions Using Weak Supervision

Alessio Palmero Aprosio
Fondazione Bruno Kessler
Trento, Italy
aprosio@fbk.eu

Sara Tonelli
Fondazione Bruno Kessler
Trento, Italy
satonelli@fbk.eu

Marco Turchi
Fondazione Bruno Kessler
Trento, Italy
turchi@fbk.eu

Matteo Negri
Fondazione Bruno Kessler
Trento, Italy
negri@fbk.eu

Mattia Di Gangi
Fondazione Bruno Kessler
Trento, Italy
digangi@fbk.eu

Abstract

Neural text simplification has gained increasing attention in the NLP community thanks to recent advancements in deep sequence-to-sequence learning. Most recent efforts with such a data-demanding paradigm have dealt with the English language, for which sizeable training datasets are currently available to deploy competitive models. Similar improvements on less resource-rich languages are conditioned either to intensive manual work to create training data, or to the design of effective automatic generation techniques to bypass the data acquisition bottleneck. Inspired by the machine translation field, in which synthetic parallel pairs generated from monolingual data yield significant improvements to neural models, in this paper we exploit large amounts of heterogeneous data to automatically select simple sentences, which are then used to create synthetic simplification pairs. We also evaluate other solutions, such as oversampling and the use of external word embeddings to be fed to the neural simplification system. Our approach is evaluated on Italian and Spanish, for which few thousand gold sentence pairs are available. The results show that these techniques yield performance improvements over a baseline sequence-to-sequence configuration.

1 Introduction

Text simplification aims at making a text more readable by reducing its lexical and structural complexity while preserving the meaning. (Chandrasekar and Bangalore, 1997; Carroll et al., 1998; Vickrey and Koller, 2008; Crossley et al., 2012; Shardlow, 2014). Neural approaches to the task have gained increasing attention in the NLP community thanks to recent advancements of deep,

sequence-to-sequence approaches. However, all recent improvements have dealt with English. The main reason is that such data-hungry approaches require large training sets (in the order of hundred thousand instances) and sizable datasets have been developed and made available only for this language. Indeed, the only available datasets composed of a complex and a simple version of the same document, which are large enough to experiment with deep neural systems, are Newsela (Xu et al., 2015) and the aligned version of simple and standard English Wikipedia (Zhu et al., 2010). These data have become the common benchmark for evaluating new approaches to neural text simplification. These methods rely on the use of deep reinforcement learning (Zhang and Lapata, 2017), memory-augmented neural networks (Vu et al., 2018), the combination of semantic parsing and neural approaches (Sulem et al., 2018) and the personalisation to specific grade levels (Scarton and Specia, 2018). Due to data paucity, none of them can be tested on other languages, for which less data-intensive, rule-based solutions have been proposed (Brouwers et al., 2012; Bott et al., 2012; Barlacchi and Tonelli, 2013). The main disadvantage of such solutions, however, is a reduced portability and scalability to new scenarios, which require the creation of new sets of rules each time a new language (or a new domain with specific idiosyncrasies) has to be covered.

To alleviate the data bottleneck issue, enabling the development of neural solutions also for languages other than English, we explore data augmentation techniques for creating task-specific training data. Our experiments range from simple oversampling techniques to weakly supervised data augmentation methods inspired by recent

works in other NLP tasks (Bérard et al., 2016; Ding and Balog, 2018), in particular Machine Translation (MT) (Sennrich et al., 2016b). In a nutshell, taking an opposite direction to simplification, we proceed by *i*) automatically selecting simple sentences from a large pool of monolingual data, and *ii*) synthetically creating complex sentences. These artificially created sentences will be then used as the “source” side of new *difficult–simple* training pairs fed into an MT-like encoder-decoder architecture.

Our hypothesis is that, though sub-optimal due to possible errors introduced in the automatic generation of complex sentences, these training pairs represent useful material for building our sequence-to-sequence text simplification models. Under this hypothesis, any noise in the source side of the pairs can still be treated as an approximation of text difficulty that, paired with its correct simplified counterpart, can contribute to model training.

We run our experiments on Italian and Spanish, two languages for which only small datasets of manually curated simplifications are available. The main contributions of this work are:

- We explore different approaches for augmenting training data for neural text simplification using weak supervision;
- We test them in under-resourced conditions on Italian and Spanish.

2 Related work

The lack of data for training sequence-to-sequence models is a problem that has been addressed in several NLP tasks. In MT, for instance, synthetic parallel data for low-resource settings have been generated by automatically translating sentences from the target language into the source language (Sennrich et al., 2016b,a). In speech translation, recent works (Bérard et al., 2016; Jia et al., 2018) have shown that end-to-end models can be successfully trained on artificial *source_audio–target_text* pairs built from synthesized speech data and/or machine-translated text.

For keyword-to-question generation, small training data have been first inverted to create a question-to-keyword dataset and then used to artificially generate keywords given a large quantity of questions (Ding and Balog, 2018).

In all these tasks, when added to the original data, the synthetic sets always result in significant improvements in performance. Even if sub-optimal due to variable noise introduced on the source side by automatic processing, large “silver” data provide a valuable additional complement to small “gold” training corpora.

Regarding neural text simplification, we are not aware of previous work on extending small training corpora with synthetic data. Indeed, the lack of training instances has been a major issue in the development of such applications for languages other than English.

3 Neural sentence simplification system

Our sentence simplification approach is based on the attentional encoder-decoder model (Bahdanau et al., 2014) initially proposed for MT. It takes as input a complex sentence and outputs its simplified version (Nisioi et al., 2017). Cast as a (monolingual) translation task, it provides a comprehensive solution to address both lexical and structural simplification, since the model does not only learn single term replacements, but also more complex structural changes. Initially, a sequence of words is fed to the encoder, which maps it into a sequence of continuous representations (the hidden states of the encoder) providing increasing levels of abstraction. At each time step, based on these continuous representations and the generated word in the previous time step, the decoder generates the next word. This process continues until the decoder generates the end-of-sentence symbol. This sequence-to-sequence model is extended by adding a pointer-generator network that allows both copying words via pointing to the source sentence, and generating words from a fixed vocabulary (See et al., 2017). At each time step, the network estimates the probability of generating a word and uses this probability as a gate to decide whether to generate or copy the word. To apply this pointer-generator network, a shared vocabulary containing all the words in the complex and simple training sentences is used. This architecture is implemented in the OpenNMT platform (Klein et al., 2017).

4 Data augmentation

Our experimentation starts from the availability of a limited quantity (few tens of thousand complex-to-simple sentence pairs) of high-quality

gold standard data that is used to train and evaluate our pointer-generator network baseline.

To satisfy the need of much larger training sets required to exploit the generalization capabilities of neural approaches,¹ we explore three different data augmentation strategies:

Oversampling: In line with the work in MT-related tasks like automatic post-editing (Chatterjee et al., 2017), we increase the size of the training set by multiplying the whole original training corpus (5 and 10 times) to maximize the use of the few “gold” sentence pairs available.

Simple-to-simple synthetic pairs creation: Starting from large monolingual corpora, we automatically extract the simplest sentences using different heuristics, and then duplicate them to create simple-to-simple pairs. These are then used as synthetic data to train the simplification system. The intuition behind this strategy is to add information that can be beneficial to the creation of better word embedding representations and to introduce a bias in the decoder towards producing simple outputs.

Simple-to-complex synthetic pairs creation: We convert the gold data into a set of simple-to-complex pairs inspired by the work in MT (Sennrich et al., 2016b) and in keyword-to-question (Ding and Balog, 2018), and then use the OpenNMT toolkit to train a “complexifier” system. Then, we run it on the set of simple sentences selected to create the simple-to-simple pairs (see above) to obtain additional simple-to-complex pairs. Finally, we revert the pairs again and use them as synthetic data to train the simplification system. The intuition behind this strategy is to maintain the human-generated simplified sentences in the target side of the parallel data to improve the generation of simplified sentences. This comes at the cost of accepting the low quality of automatically “complicated” source sentences. Due to the limited amount of training data available, we do not expect that complicating a sentence is an easier task than making it simpler, so the quality of the automatic complex sentences can be limited. With this method, however, we are interested in checking if the neural network approach is able to infer useful informa-

¹In speech recognition and MT, for instance, the impressive performance obtained by end-to-end systems is the result of resource-intensive training, respectively on thousands of hours of transcribed speech (Chiu et al., 2018) and tens of millions of parallel sentences (Hassan et al., 2018).

tion from low-quality data when dealing with few gold-standard sentence pairs. We expect that, similar to MT, a neural simplification model can be trained even if the source data is not of high quality, given that the sentences on the target side are correct.

Additionally, we explore also whether large scale **pre-trained embeddings** can improve text simplification models. A similar setting was evaluated on English (Nisioi et al., 2017) and did not yield remarkable improvements. However, our intuition is that pre-trained embeddings may be more beneficial in low-resource conditions, providing additional information that cannot be extracted from small training corpora.

5 Experimental Setup

We run our experiments on two languages, Italian and Spanish. Below, we describe for each language the gold standard and the simple monolingual data extraction process to augment our training data.

5.1 Italian

To obtain the Italian gold standard, we merge three available data sets, namely:

- The SIMPITIKI corpus (Tonelli et al., 2016), a manually curated corpus with 1,166 complex-simple pairs extracted from Italian Wikipedia and from documents in the administrative domain;
- The corpus presented in (Brunato et al., 2015), another manually curated corpus comprising 1,690 sentence pairs from the educational domain;
- A subset of the PaCCSS-it corpus (Brunato et al., 2016), which contains 63,000 complex-to-simple sentence pairs automatically extracted from the Web. In order to extract only the pairs of higher quality, we pre-processed the corpus by discarding sentence pairs with special characters, misspellings, non-matching numerals or dates, and a cosine similarity below 0.5.

The final gold standard contains 32,210 complex-to-simple pairs.

The set of simple sentences used to create the synthetic pairs is obtained from a large monolingual corpus covering both formal and infor-

Pretr. Emb.	Copied	Complic.	ITA x1	ITA x5	ITA x10	SPA x1	SPA x5	SPA x10
-	-	-	44.6	48.5	48.1	28.4	27.4	27.6
✓	-	-	47.3	48.8	49.5	29.1	28.2	28.1
-	✓	-	44.4	49.4	49.1	23.1	24.3	28.6
✓	✓	-	44.2	49.2	49.2	24.6	25.0	27.2
-	-	✓	48.0	49.9	49.8	28.6	28.6	30.6
✓	-	✓	47.9	49.9	50.0	28.6	28.7	30.8
-	✓	✓	45.6	49.3	49.5	29.0	29.1	26.2
✓	✓	✓	45.2	49.9	49.7	24.9	25.0	26.2

Table 1: Results of neural simplification experiments on Italian and Spanish data (SARI)

mal language, including Italian Opensubtitles,² the Paisà corpus (Lyding et al., 2014), Wikipedia and the collection of Italian laws.³ This merging process results in around 1.3B words and 125M sentences. We rank all sentences by readability level according to the best features described in (Dell’Orletta et al., 2014) and keep the 500,000 most readable (i.e. simplest) sentences to create the synthetic pairs. This process is needed due to the lack of an Italian equivalent of the Simple English Wikipedia,⁴ which is widely used as a source of simple monolingual data when dealing with English text simplification (Zhu et al., 2010; Woodsend and Lapata, 2011). From the large corpus described above, before filtering only simple sentences, we also create word embeddings with 300 dimensions using word2vec (Mikolov et al., 2013).

5.2 Spanish

The Spanish gold standard is obtained from the Spanish Newsela corpus,⁵ containing 1,221 documents manually annotated by professionals for different proficiency levels. We align complex-simple pairs using the CATS-Align⁶ tool (Štajner et al., 2018) and discard the pairs coupled with an alignment accuracy below 0.5. The gold standard contains 55,890 complex-to-simple pairs.

The set of simple sentences used to create the synthetic pairs is extracted from a large monolin-

gual corpus covering different domains, obtained from websites written in simple Spanish for language learners.⁷ The documents are then ranked based on the Flesch-Szigriszt readability score for Spanish (Szigriszt, 1993)⁸ and all sentences belonging to the most readable ones are included in the set of simple monolingual data (484,325 simple sentences in total, from a set of about 1.2M sentences). For Spanish, we do not rank directly the sentences because there is no specific study to identify metrics at sentence level similar to the one for Italian presented in (Dell’Orletta et al., 2014).

The Spanish embeddings used in the simplification process are those obtained from the Spanish Billion Word Corpus, that is widely used in NLP experiments on Spanish (Zea et al., 2016; Quirós et al., 2016).⁹ To favour the extraction of word embeddings from simple texts, we increase the Spanish Billion Word Corpus by adding our extracted simple Spanish texts. In total, Spanish word embeddings are extracted from a corpus of nearly 1.5B words.

5.3 System configuration

OpenNMT is run on a Nvidia Tesla K80 GPU using stochastic gradient descent (Robbins and Monro, 1951) optimization with learning rate 1. Each run is repeated three times with different seeds, then the average value is considered. Since the source and target languages are the same, in

²www.opensubtitles.org

³www.gazzettaufficiale.it

⁴simple.wikipedia.org/wiki/Simple_English_Wikipedia.

⁵newsela.com/data

⁶The tool (github.com/neosyon/SimpTextAlign) includes several lexical and semantic text similarity methods and alignment strategies for simplified text alignment at different text representation levels (paragraph, sentence, and sentence with paragraph pre-alignment).

⁷For example www.cuentosinfantiles.net or www.mundoprimaria.com

⁸This score is an adaptation of the Flesch Index (Flesch, 1946), which provides a readability measure combining word and sentence length in a 1-100 scale (the closer the score is to 100, the easier the text is to read). The Flesch-Szigriszt adaptation refines the original Flesch equation by also considering the number of syllables and phrases in the text.

⁹github.com/uchile-nlp/spanish-word-embeddings

the preprocessing phase their vocabulary is shared. We split data into train/dev/test with ratio 90/5/5 respectively. For Italian, this results in a split of 29, 260/1, 475/1, 475 sentence pairs, while for Spanish it is 50, 301/2, 794/2, 795.

6 Evaluation

We report in Table 1 the results of Italian and Spanish text simplification using different settings and data augmentation techniques. For each language, we evaluate the results using the gold training set as is, and expanding it through oversampling (i.e. repetition of the same sentence pairs 5 and 10 times). In addition, we evaluate the impact of: *i*) adding pre-trained word embeddings built on large monolingual corpora (*Pretr.Emb*), *ii*) using the simple-to-simple pairs for data augmentation (*Copied*), and *iii*) using, for the same purpose, the simple-to-complex synthetic pairs (*Complic.*). We also explore the addition of different combinations of the aforementioned resources. The evaluation is performed by computing the SARI score (Xu et al., 2016) on the test set.

Our results show that adding only pre-trained word embeddings trained on large monolingual corpora achieves, in general, better performance than the baseline (max: +2.73 SARI points for Italian, +0.8 for Spanish). Our experiments show also that the usefulness of simple-to-simple pairs cannot be generalised: they are beneficial for all results on Italian and *SPAx10*, while they are harmful for *SPAx1* and *SPAx5*. Our intuition is that the copied data pushed the system in the direction of learning to copy the source sentence in the output instead of simplifying it, which can create some instability in the model during training. The addition of simple-to-simple pairs and of pre-trained word embeddings does not yield large improvements, confirming the idea that the copied pairs mainly affect the quality of the word embedding representations instead of the relation between complex and simple sentences (i.e. attention network).

The largest gains in performance are obtained when using the simple-to-complex synthetic pairs. Both in isolation and when paired with pre-trained embeddings, they make the neural model able to outperform the baseline up to +3.4 SARI points. The best results for both languages are obtained by multiplying the training data by 10 and adding the simple-to-complex synthetic data. These config-

urations outperform the standard settings (*ITAx1* and *SPAx1*) by +5.4 SARI points for Italian and +2.4 for Spanish.

When concatenating all the synthetic and real data, and the pre-trained embeddings are used, the performance is comparable with the one obtained using the simple-to-complex synthetic pairs, but at the cost of using a larger quantity of training data.

Although we cannot make a direct comparison of the SARI scores across different languages, Italian and Spanish are typologically very similar, and therefore we can argue that our models for neural simplification in Italian works better than the Spanish ones. This may depend on several reasons. For Italian, the selection of 500,000 simple sentences is based on sentence-specific features correlated with high readability, emerged from the analysis in (Dell’Orletta et al., 2014). On the contrary, extracting simple monolingual sentences based on the readability score at document level, as we did for Spanish, is more prone to inconsistencies. Other differences may be due to the quality of gold standard data: although the Spanish gold standard is bigger than the Italian one (55,890 complex-simple sentence pairs vs. 32,210 pairs respectively), its language is generally more complex, since it contains news articles, while the Italian gold standard includes to a large extent stories for children and textbooks. Besides, while some of the Italian sentences were manually aligned, the Spanish gold data were obtained by automatically extracting complex-to-simple pairs from the Newsela corpus, in which the alignment had been done at document level.

As a comparison, we evaluate on the same test set also the MUSST syntactic simplifier (Scarton et al., 2017), a freely available system implementing a set of simplification rules for Italian and Spanish. We obtain 20.16 SARI for Italian and 21.24 for Spanish. Our results show that, despite some issues described before, low-resource neural simplification is still a promising research direction to pursue, especially with data augmentation. This is particularly true for Spanish MUSST, which includes a richer set of rules than the Italian version, but that achieves nevertheless -9.56 SARI points than the best neural model for Spanish.

7 Conclusions

We presented several techniques to augment the amount of training data for neural text simplifica-

tion through weak supervision. Our solutions were evaluated on Italian and Spanish using a sequence-to-sequence approach. Our results show that using external embeddings is generally beneficial in a low-resource setting, since they provide additional information that cannot be extracted from a limited amount of training pairs. Another gain in performance is achieved using complex-to-simple synthetic pairs created with a ‘complexifier’ system.

In the future, we plan to extend both the languages of the experiments and the data augmentation techniques, for example by applying machine translation to increase the amount of gold sentence pairs across languages, or by using bootstrapping techniques.

Acknowledgments

This work has been partially supported by the European Commission project SIMPATICO (H2020-EURO-6-2015, grant number 692819).

References

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Gianni Barlacchi and Sara Tonelli. 2013. **ERNESTA: A Sentence Simplification Tool for Children’s Stories in Italian**. In *Computational Linguistics and Intelligent Text Processing: 14th International Conference, CICLing 2013, Samos, Greece, March 24-30, 2013, Proceedings, Part II*, pages 476–487, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Alexandre Bérard, Olivier Pietquin, Laurent Besacier, and Christophe Servan. 2016. Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation. In *NIPS Workshop on end-to-end learning for speech and audio processing*, Barcelona, Spain.
- Stefan Bott, Horacio Saggion, and David Figueroa. 2012. A Hybrid System for Spanish Text Simplification. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, pages 75–84.
- Laetitia Brouwers, Delphine Bernhard, Anne-Laure Ligozat, and Thomas François. 2012. **Simplification syntaxique de phrases pour le français (Syntactic Simplification for French Sentences) [in French]**. In *Proceedings of the Joint Conference JEP-TALN-RECITAL 2012, volume 2: TALN*, pages 211–224. ATALA/AFCP.
- Dominique Brunato, Andrea Cimino, Felice Dell’Orletta, and Giulia Venturi. 2016. **PaCCSS-IT: A Parallel Corpus of Complex-Simple Sentences for Automatic Text Simplification**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 351–361. Association for Computational Linguistics.
- Dominique Brunato, Felice Dell’Orletta, Giulia Venturi, and Simonetta Montemagni. 2015. **Design and Annotation of the First Italian Corpus for Text Simplification**. In *Proceedings of The 9th Linguistic Annotation Workshop*, pages 31–41, Denver, Colorado, USA. Association for Computational Linguistics.
- John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical Simplification of English Newspaper Text to Assist Aphasic Readers. In *In Proc. of AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10.
- Raman Chandrasekar and Srinivas Bangalore. 1997. Automatic induction of rules for text simplification. *Knowledge-Based Systems*, 10(3):183–190.
- Rajen Chatterjee, M. Amin Farajian, Matteo Negri, Marco Turchi, Ankit Srivastava, and Santanu Pal. 2017. **Multi-source Neural Automatic Post-Editing: FBK’s Participation in the WMT 2017 APE Shared Task**. In *Proceedings of the Second Conference on Machine Translation*, pages 630–638. Association for Computational Linguistics.
- Chung-Cheng Chiu, Tara Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J. Weiss, Kanishka Rao, Katya Gonina, Navdeep Jaitly, Bo Li, Jan Chorowski, and Michiel Bacchiani. 2018. State-of-the-art Speech Recognition With Sequence-to-Sequence Models. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP) 2018*, pages 4774–4778, Calgary, Alberta, Canada.
- Scott A. Crossley, David Allen, and Danielle S. McNamara. 2012. Text simplification and comprehensible input: A case for an intuitive approach. *Language Teaching Research*, 16(1):89–108.
- Felice Dell’Orletta, Martijn Wieling, Giulia Venturi, Andrea Cimino, and Simonetta Montemagni. 2014. Assessing the Readability of Sentences: Which Corpora and Features? In *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*.
- Heng Ding and Krisztian Balog. 2018. **Generating Synthetic Data for Neural Keyword-to-Question Models**. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval, ICTIR*, pages 51–58.
- Rudolf Flesch. 1946. *The Art of plain talk*. Harper.

- Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou. 2018. Achieving Human Parity on Automatic Chinese to English News Translation. *ArXiv e-prints arXiv:1803.05567v2*.
- Ye Jia, Melvin Johnson, Wolfgang Macherey, Ron J. Weiss, Yuan Cao, Chung-Cheng Chiu, Stella-Lorenzo Ari, and Yonghui Wu. 2018. Leveraging Weakly Supervised Data to Improve End-to-End Speech-to-Text Translation. *ArXiv e-prints arXiv:1811.02050*.
- Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. 2017. **OpenNMT: Open-Source Toolkit for Neural Machine Translation**. In *Proceedings of ACL*.
- Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell’Orletta, Henrik Dittmann, Alessandro Lenci, and Vito Pirrelli. 2014. **The PAISÀ Corpus of Italian Web Texts**. In *Proceedings of the 9th Web as Corpus Workshop, WaC@EACL 2014, Gothenburg, Sweden, April 26, 2014*, pages 36–43. Association for Computational Linguistics.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS’13*, pages 3111–3119, USA. Curran Associates Inc.
- Sergiu Nisioi, Sanja Stajner, 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, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, pages 85–91. Association for Computational Linguistics.
- Antonio Quirós, Isabel Segura-Bedmar, and Paloma Martínez. 2016. LABDA at the 2016 TASS Challenge Task: Using Word Embeddings for the Sentiment Analysis Task. In *TASS@ SEPLN*, pages 29–33.
- Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407.
- Carolina Scarton, Alessio Palmero Aprosio, Sara Tonelli, Tamara Martín Wanton, and Lucia Specia. 2017. **MUSST: A Multilingual Syntactic Simplification Tool**. In *Proceedings of the IJCNLP 2017, System Demonstrations*, pages 25–28. Association for Computational Linguistics.
- 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)*, pages 712–718. Association for Computational Linguistics.
- Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. **Get To The Point: Summarization with Pointer-Generator Networks**. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083. Association for Computational Linguistics.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. **Edinburgh Neural Machine Translation Systems for WMT 16**. In *Proceedings of the First Conference on Machine Translation*, pages 371–376, Berlin, Germany. Association for Computational Linguistics.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. **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. Association for Computational Linguistics.
- Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications(IJACSA), Special Issue on Natural Language Processing 2014*, 4(1).
- Elior Sulem, Omri Abend, and Ari Rappoport. 2018. **Simple and Effective Text Simplification Using Semantic and Neural Methods**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 162–173. Association for Computational Linguistics.
- Francisco Szigriszt. 1993. *Sistemas predictivos de legibilidad del lenguaje escrito*. Ph.D. thesis, Tesis doctoral. Madrid: Universidad Complutense de Madrid.
- Sara Tonelli, Alessio Palmero Aprosio, and Francesca Saltori. 2016. **SIMPITIKI: a Simplification Corpus for Italian**. In *Proceedings of the 3rd Italian Conference on Computational Linguistics (CLiC-it)*, volume 1749 of *CEUR Workshop Proceedings*.
- David Vickrey and Daphne Koller. 2008. Applying Sentence Simplification to the CoNLL-2008 Shared Task. In *Proceedings of the Twelfth Conference on Computational Natural Language Learning, CoNLL*, pages 268–272, Manchester, UK.
- 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).

Tu Vu, Baotian Hu, Tsendsuren Munkhdalai, and Hong Yu. 2018. [Sentence Simplification with Memory-Augmented Neural Networks](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 79–85. Association for Computational Linguistics.

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.

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.

Jenny Linet Copara Zea, Jose Eduardo Ochoa Luna, Camilo Thorne, and Goran Glavaš. 2016. Spanish NER with Word Representations and Conditional Random Fields. In *Proceedings of the Sixth Named Entity Workshop*, pages 34–40.

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, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*, pages 584–594. Association for Computational Linguistics.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. [A Monolingual Tree-based Translation Model for Sentence Simplification](#). In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361.