Keep it Simple

Unsupervised Text Simplification

Philippe Laban, Tobias Schnabel, Paul Bennett, Marti A. Hearst

UC Berkeley, Microsoft

ACL2021 Recorded Presentation

Berkeley UNIVERSITY OF CALIFORNIA

Microsoft Research

Text Simplification (Example)

10th Grade

A rat named Magawa is being **honored** with one of the highest awards in the animal world. He has potentially saved **numerous** lives for clearing **landmines** from fields in Cambodia.

4th Grade

A rat in Cambodia has a very special job. He helps make **things safer** for people. The rat's name is Magawa. He received a **gold medal** for his work. It is the highest award in the animal world. He has possibly saved **many** lives.

Example credit: USA Today and Newsela (2020/08/11).

What is so difficult ....

with text simplification?

What is so difficult?

Existing datasets:

- Simple Wikipedia (large but noisy)

https://simple.wikipedia.org/
Xu, Wei, Chris Callison-Burch, and Courtney Napoles. "Problems in current text simplification research: New data can help." *Transactions of the Association for Computational Linguistics* 3 (2015): 283-297.

What is so difficult?

Existing datasets:

- Simple Wikipedia (large but noisy)
- Newsela (high-quality, small, hard to align)

https://newsela.com/about/resources/research/

What is so difficult?

Existing datasets:

- Simple Wikipedia (large but noisy)
- Newsela (high-quality, small, hard to align)

Supervised seq2seq is limited.

Do we really need data?

Keep it Simple

What if we approach Text Simplification in an **unsupervised** way:

1. Define a “text simplification” reward

2. Train a pre-trained language model to optimize the reward.

Talk Outline

1. Reward Design

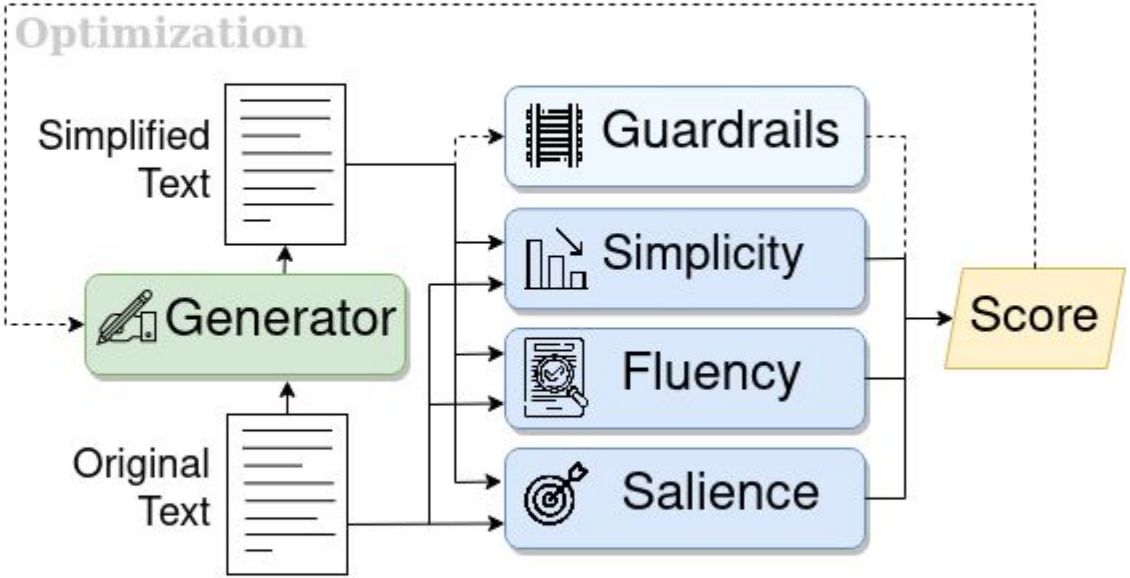
2. Optimization

3. Evaluation

1
Reward

Keep it Simple

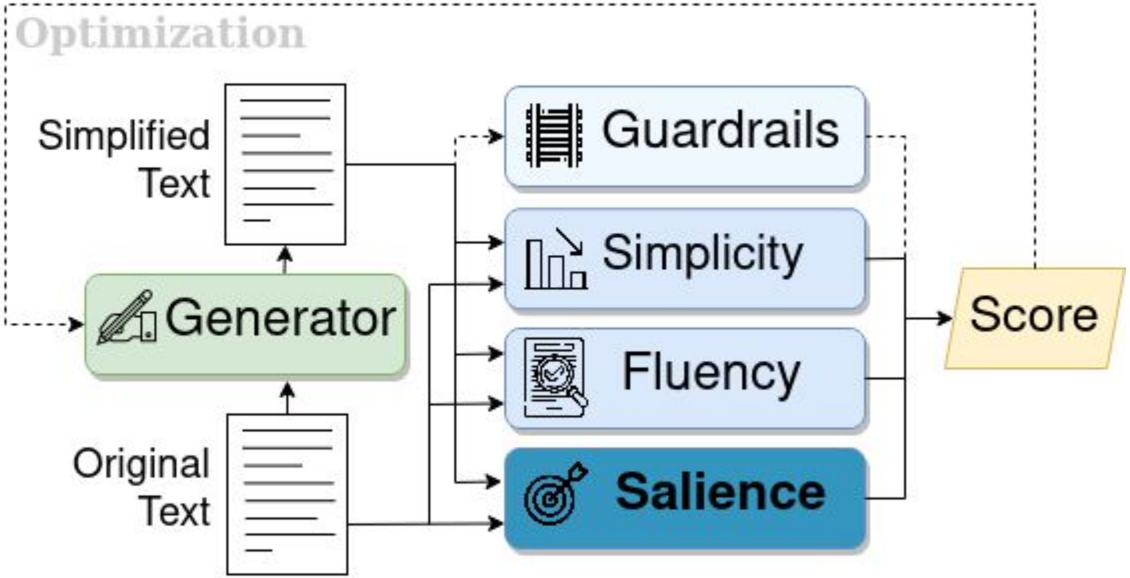
Adapting the Summary Loop to the domain of ***Text Simplification***



Laban, Philippe et al. "The Summary Loop: Learning to Write Abstractive Summaries Without Examples." ACL2020

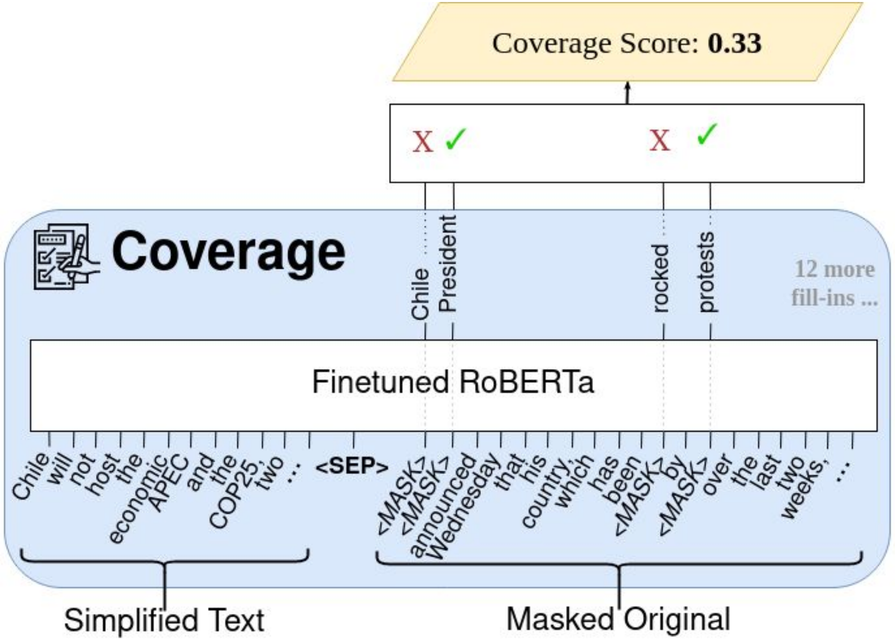
Keep it Simple: Salience

Objective: The generated text should contain the same information as the original text.



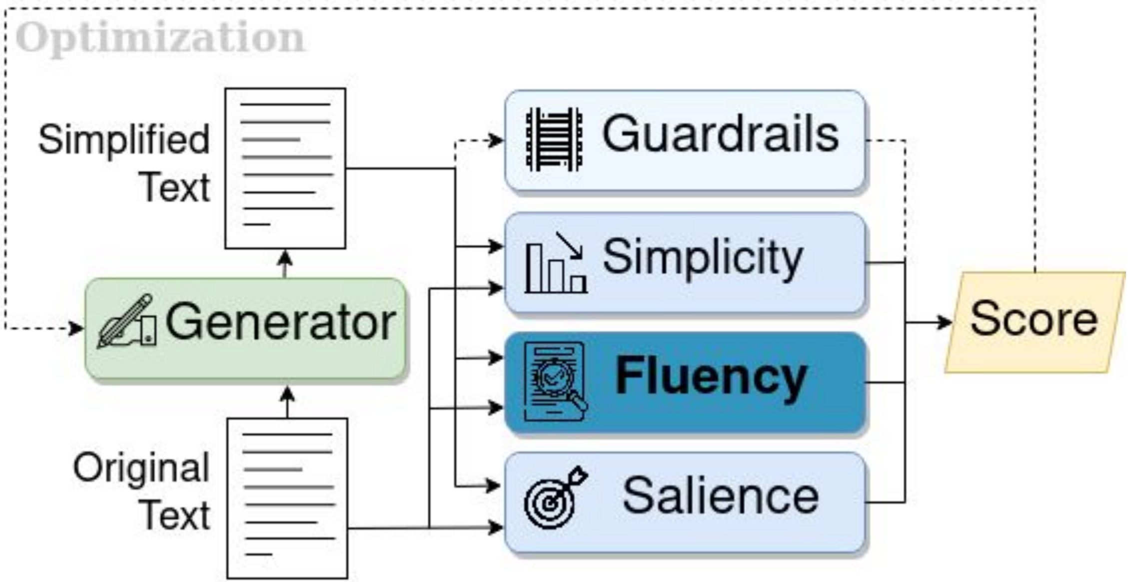
Keep it Simple: Salience

<u>Key Idea:</u> Adapting the Coverage model from the Summary Loop (ACL 2020) to text simplification.



Keep it Simple: Fluency

Objective: The generated text should be grammatical and be in fluent English.



Keep it Simple: Fluency

Fluency Component 1: Pretrained-Language Model

```
GPT2Score("A rat in Cambodia has a very special job...") =    0.98

GPT2Score("landmines honor Cambodia Magawa rat saving...") = 0.01
```

Keyword soup
gets low fluency score

The language model checks that word sequences *feel like* sentences.

Keep it Simple: Fluency

Fluency Component 1: Pretrained-Language Model

Problem: the fluency model is static (it does not change during training).
The generator can learn *common patterns* to artificially score high on the fluency metric.

Keep it Simple: Fluency

Fluency Component 1: Pretrained-Language Model

Fluency Component 2: Dynamic Discriminator (Adversarial)

"Real" Corpus Sentences
(label = 1)

Generated Sentences
(label = 0)

Discriminator

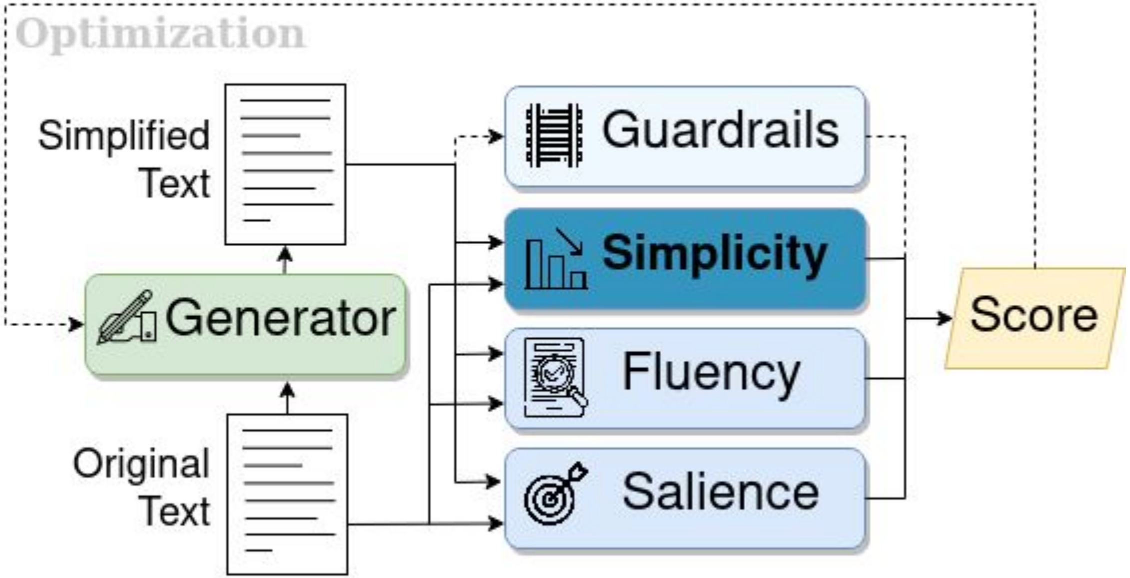
Score(text) = P(Y=real|text)

Train dynamically
(every 2,000 generated samples)

Details in paper

Keep it Simple: Simplicity

Objective: The generated text should be *simpler* than the original text, both *syntactically* and *lexically*.



Keep it Simple: Simplicity

Syntactic Simplicity: S_{Score}

We use the standard Flesch-Kincaid Grade Level (FKGL)

```
FKGL("A rat named Magawa is being honored...")       = 9.5
FKGL("A rat in Cambodia has a very special job...") = 4.1
```

During training: target a fixed amount of drop (e.g., 2 grade levels).
Ramp score based on how close the model gets to the target.

Flesch, Rudolf, and Alan J. Gould. *The art of readable writing*. Vol. 8. New York: Harper, 1949.

Keep it Simple: Simplicity

Syntactic Simplicity: S_{Score}

Lexical Simplicity: L_{Score}

Words added in the generated text should be *more common* than **words removed** from the original text.

Words Removed		Words Added	
commonness(*honored*)	= 4.0	commonness(*very*)	= 6.0
commonness(*potentially*)	= 4.3	commonness(*possibly*)	= 4.7
commonness(*numerous*)	= 4.7	commonness(*many*)	= 5.9
commonness(*landmines*)	= 2.7	commonness(*things*)	= 5.7

Details in paper

2
Optimization

Combining Reward Components

1. Salience

Coverage

2. Fluency

Language Model

Discriminator

3. Simplicity

Lexical

Syntactic

Each of the 5 components is **normalized** between [0,1]

The total score is the **product** of component scores.

Self-Critical Sequence Training Recap

1. Generate 2 candidates (1 argmax and 1 sampled)
2. Compute each candidate's total reward
3. Train model to increase likelihood of highest-reward candidate

$$L = (\hat{R} - R^S) \sum_{i=0}^{N} \log p(w_i^S | w_1^S ... w_{i-1}^S, P)$$

Rennie, Steven J., et al. "Self-critical sequence training for image captioning." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

Self-Critical Sequence Training Recap

1. Generate 2 candidates (1 argmax and 1 sampled)
2. Compute each candidate's total reward
3. Train model to increase likelihood of highest-reward candidate

$$L = (\hat{R} - R^S) \sum_{i=0}^{N} \log p(w_i^S | w_1^S \ldots w_{i-1}^S, P)$$

If both candidates have similar scores, loss is close to zero, no learning in the sample.
In practice can happen with >30% of samples.

Rennie, Steven J., et al. "Self-critical sequence training for image captioning." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

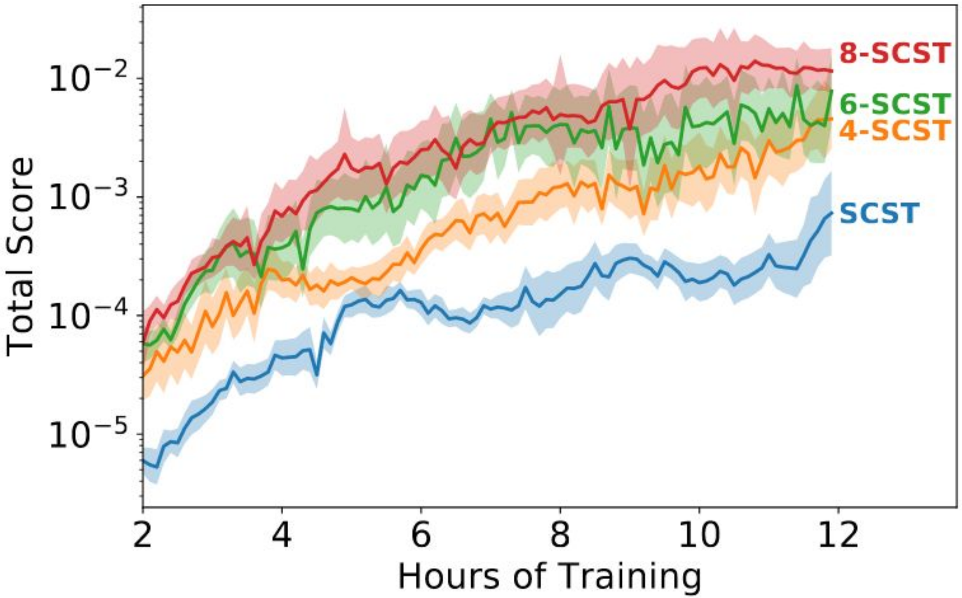
Contribution: k-SCST modification

1. Generate k candidates (all sampled, for example k=6)
2. Compute each candidate's total reward
3. Train model to increase likelihood of candidates with **higher than average** total reward

$$L = \boxed{\sum_{j=1}^{k}(\bar{R}^{S} - R^{Sj})} \sum_{i=0}^{N} \log p(w_i^{Sj} | w_1^{Sj} \ldots w_{i-1}^{Sj}, P)$$

As k increases, likelihood of good performing candidates increases.

k-SCST modification



Training the KiS model, each configuration run with **6 runs**.
Increasing k leads to faster, more stable training.

3
Evaluation

Automatic Results

Keep it Simple achieves state-of-the-art on news simplification (even outperforming supervised models)

	Method	SARI	% Lexile
	Reference	-	**79**
Supervised	Finetune Baseline	0.470	52
Supervised	ACCESS (Martin et al. 2020)	0.666	63
Supervised	ACCESS 90	0.674	64
Unsup.	Unsup. NTS (Surya et al. 2019)	0.677	57
Unsup.	Keep It Simple (ours)	**0.709**	72

SARI is an n-gram overlap measure with hand-written simplified references.

% Lexile is the percentage of generated texts that achieve higher readability than the input, according to the Lexile measure (gold standard)

Human Evaluation of Simplification

Designing human evaluation for text simplification.

What does success mean for simplification?

Human Evaluation of Simplification

Designing human evaluation for text simplification.

Hypothesis 1: Increasing Accessibility

Can a broader audience understand the simplified text than the original text?

Human Evaluation of Simplification

Designing human evaluation for text simplification.

<u>Hypothesis 1: Increasing Accessibility</u>
Can a broader audience understand the simplified text than the original text?

⚠ *Hypothesis 1 would require a study with a large and diverse population.*

Human Evaluation of Simplification

Designing human evaluation for text simplification.

<u>Hypothesis 2: Decrease Cognitive Load</u>
Does the simplified text lead to a similar level of understanding in less time?

Human Evaluation of Simplification

Designing human evaluation for text simplification.

Hypothesis 2: Decrease Cognitive Load
Does the simplified text lead to a similar level of understanding in less time?

*We design a **Human Comprehension Study** to investigate Hypothesis 2.*

Human Comprehension Study

Original
Short News Article (~200 words)

Human Comprehension Study

Original
Short News Article (~200 words)

1. **Who manages.... ?** A/ B/ C/ D/
2. **Why is.... ?** A/ B/ C/
3. **How did ... ?** A/ B/ C/
4. **What is ... ?** A/ B/ C/ D/
5. **When did ... ?** A/ B/

Generate 5 comprehension questions.

Human Comprehension Study

4 candidate simplifications (1 human, 3 systems)

Original Newsela ACCESS Finetune KiS

1. Who manages.... ? A/ B/ C/ D/
2. Why is.... ? A/ B/ C/
3. How did ... ? A/ B/ C/
4. What is ... ? A/ B/ C/ D/
5. When did ... ? A/ B/

Generate 4 candidate simplifications

Human Comprehension Study

Original Newsela ACCESS Finetune KiS

1. **Who manages.... ?** A/ B/ C/ D/
2. **Why is.... ?** A/ B/ C/
3. **How did ... ?** A/ B/ C/
4. **What is ... ?** A/ B/ C/ D/
5. **When did ... ?** A/ B/

Participants are randomly assigned to a text version

Human Comprehension Study

Original | Newsela | ACCESS | Finetune | KiS

1. Who manages.... ? A/ B/ C/ D/
2. Why is.... ? A/ B/ C/
3. How did ... ? A/ B/ C/
4. What is ... ? A/ B/ C/ D/
5. When did ... ? A/ B/

Participants re-submit questionnaire until correct

Human Comprehension Study

Original | Newsela | ACCESS | Finetune | KiS

1. **Who manages.... ?** A/ B/ C/ D/
2. **Why is.... ?** A/ B/ C/
3. **How did ... ?** A/ B/ C/
4. **What is ... ?** A/ B/ C/ D/
5. **When did ... ?** A/ B/

90 participants; 4 documents; 244 total submissions

Human Comprehension Study

Hypothesis 2: Confirmed. Simplified text leads to faster completion.

Method	Completion Time (sec.)
Original Text	174.0
Newsela References	163.3 ✓
ACCESS (Martin et al. 2020)	188.5
Finetune Baseline	161.0 ✓
Keep It Simple (ours)	**142.6** * ✓

3 / 4 methods lead to drop in completion time
KiS stat. significant drop ($p < 0.05$) compared to original

Thanks

See you at the Q&A!

Keep It Simple:Unsupervised Text Simplification
Laban, Schnabel, Bennet, Hearst
Code on Github:
https://github.com/tingofurro/keep_it_simple

Contact:
phillab@berkeley.edu

Berkeley UNIVERSITY OF CALIFORNIA

Microsoft Research