Elaborative Simplification: Content Addition and Explanation Generation in Text Simplification

Neha Srikanth[1] , **Junyi Jessy Li**[2]

[1] Department of Computer Science, University of Texas at Austin
[2] Department of Linguistics, University of Texas at Austin

nehasrik@utexas.edu, jessy@austin.utexas.edu

Motivation

When simplifying text, lexical or syntactic transformations may be **insufficient**: there may be **additional background information** that readers need to understand content in the original text.

Original

Results, she said, "could help the team better understand ancient Egyptian health" and, correspondingly, modern-day health.

For instance, some mummies still have arteries in their mummified remains, Miller-Thomas said.

And, sometimes, scientists can tell if those arteries had hardened.

Simplified

The scans could help the team understand about ancient Egyptians' health.

For example, some mummies still have arteries.

An artery is a tube that moves blood through the body.

The artery could show if the person had been healthy or not.

Define "artery"

Explain

Motivation

When simplifying text, lexical or syntactic transformations may be **insufficient**: there may be **additional background information** that readers need to understand content in the original text.

Original

Results, she said, "could help the team better understand ancient Egyptian health" and, correspondingly, modern-day health.

For instance, some mummies still have arteries in their mummified remains, Miller-Thomas said.

Simplified

The scans could help the team understand about ancient Egyptians' health.

For example, some mummies still have arteries.

We understand this well! There is lots of prior research on sentence simplification.

Motivation

When simplifying text, lexical or syntactic transformations may be **insufficient**: there may be **additional background information** that readers need to understand content in the original text.

We have yet to understand **insertion of content** as a phenomenon during simplification.

Past work is limited:

- Retrieving & inserting definitions[1]

- Generating post-modifiers[2]

Simplified

The scans could help the team understand about ancient Egyptians' health.

For example, some mummies still have arteries.

| Define "artery" | An artery is a tube that moves blood through the body. |

| Explain | The artery could show if the person had been healthy or not. |

[1]Damay et al., 2006; Kandula et al., 2010; Eom et al., 2012; Paetzold and Specia, 2016)
[2]Kang et al., 2019

Elaborative Simplification

We introduce *elaborative simplification*, the insertion of content during the simplification process, meant to **aid document understanding.**

Our work attempts to:

Characterize naturally occurring elaborations in the Newsela corpus
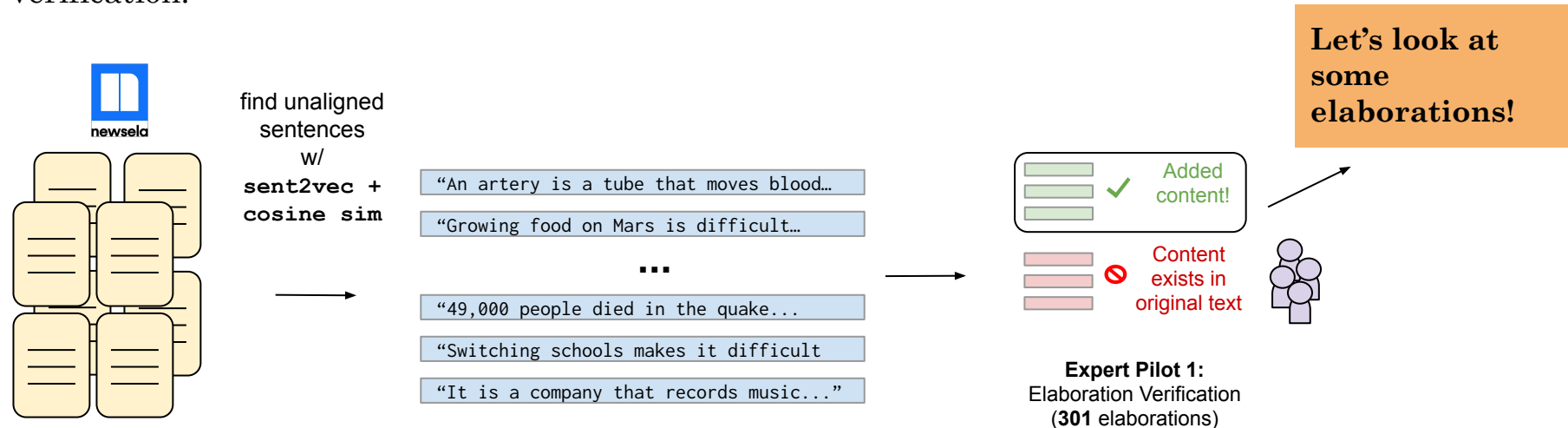
Provide a corpus of elaborations for exploration and modeling

Present a framework for collecting elaborations at scale

Establish baseline models for generating elaborations

Data Collection

We collected a corpus of elaborations using a combination of automatic extraction and human verification.

find unaligned sentences w/ **sent2vec + cosine sim**

"An artery is a tube that moves blood…"

"Growing food on Mars is difficult…"

...

"49,000 people died in the quake..."

"Switching schools makes it difficult"

"It is a company that records music..."

Added content!

Content exists in original text

Expert Pilot 1:
Elaboration Verification
(**301** elaborations)

Let's look at some elaborations!

Extract candidate elaborations from Newsela corpus

> 4K candidate elaborations sent to annotators

Pilot Annotation to understand elaborations

Elaborations

All content in these elaborations are added in to help readers further understand the original text.

Definitions or other background information about key entities in the original text

Explicit explanations or clarifications of implicit connections readers are expected to make in the original text

Reasoning or analysis to highlight a central idea or point made in the original text

And his wife, Maria, was inspired to get her GED. The general educational development (GED) is equal to a high school diploma. **It is for adults who were unable to finish high school.**

The National Mall is the stretch of parks and museums in Washington D.C. Its nickname is the "nation's front lawn."… **Every year millions of tourists visit it.** This October, there is an extra reason to go.

Claudia goes to two different schools each year. She gets straight A's at one. Her grades are lower at the other. **Switching between schools makes it more difficult to learn.** It's easy for kids like Claudia to fall behind.

Contextual Specificity

Low

And his wife, Maria, was inspired to get her GED. The general educational development (GED) is equal to a high school diploma. **It is for adults who were unable to finish high school.**
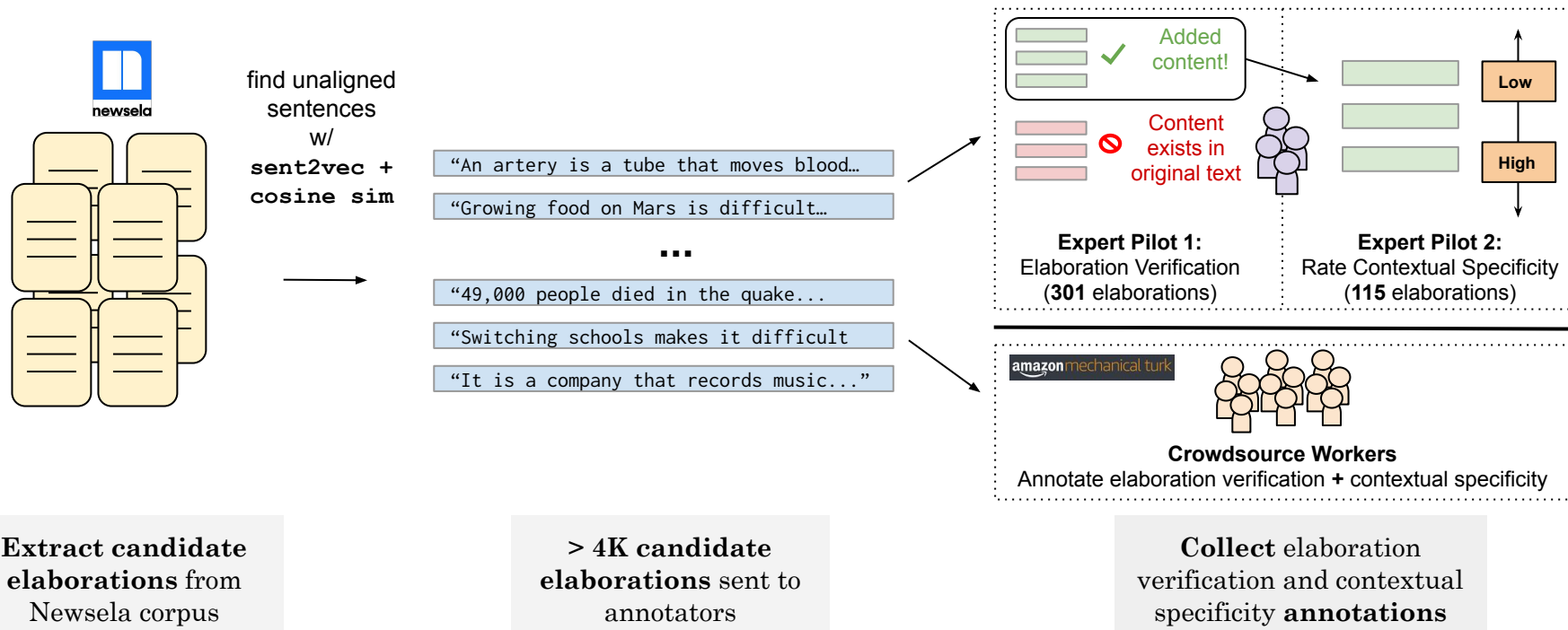
Medium

The National Mall is the stretch of parks and museums in Washington D.C. Its nickname is the "nation's front lawn."… **Every year millions of tourists visit it.** This October, there is an extra reason to go.

High

Claudia goes to two different schools each year. She gets straight A's at one. Her grades are lower at the other. **Switching between schools makes it more difficult to learn.** It's easy for kids like Claudia to fall behind.

Elaborative Simplification Corpus Collection

We collected a corpus of elaborations using a combination of automatic extraction and human verification.

Extract candidate elaborations from Newsela corpus

> 4K candidate elaborations sent to annotators

Collect elaboration verification and contextual specificity **annotations**

Elaborative Simplification Corpus
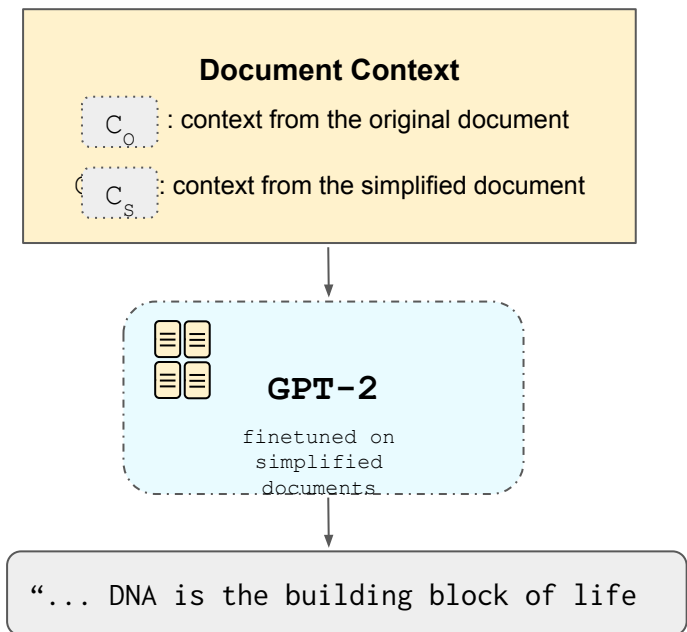
Low	Medium	High	Total
406	423	470	1299

Using this collection workflow, we collected around **1.3K instances** of elaborative simplification in the Newsela corpus.

Across the corpus we see a fairly balanced distribution of contextual specificity.
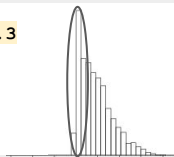
Elaboration Generation

We frame elaborative simplification as an NLG task, meant to mimic editors elaborating while simplifying documents.

Document Context

C_O : context from the original document

C_S : context from the simplified document

↓

GPT-2

finetuned on simplified documents

↓

"... DNA is the building block of life

Baseline:
Greedy Decoding

```
building: 0.3
...
most: 0.2
...
part: 0.1
```

Top-k:
Sample a sequence from LM using top-k sampling

```
building: 0.3
...
most: 0.2
...
part: 0.1
```

Contextual:
Sample 3 sequences using top-k sampling of low, medium, high contextual specificity & select sequence with gold specificity

DNA is the building block of life

We have thousands of DNA strands.

People look different because they have different DNA.

Elaboration Generation

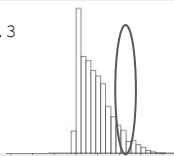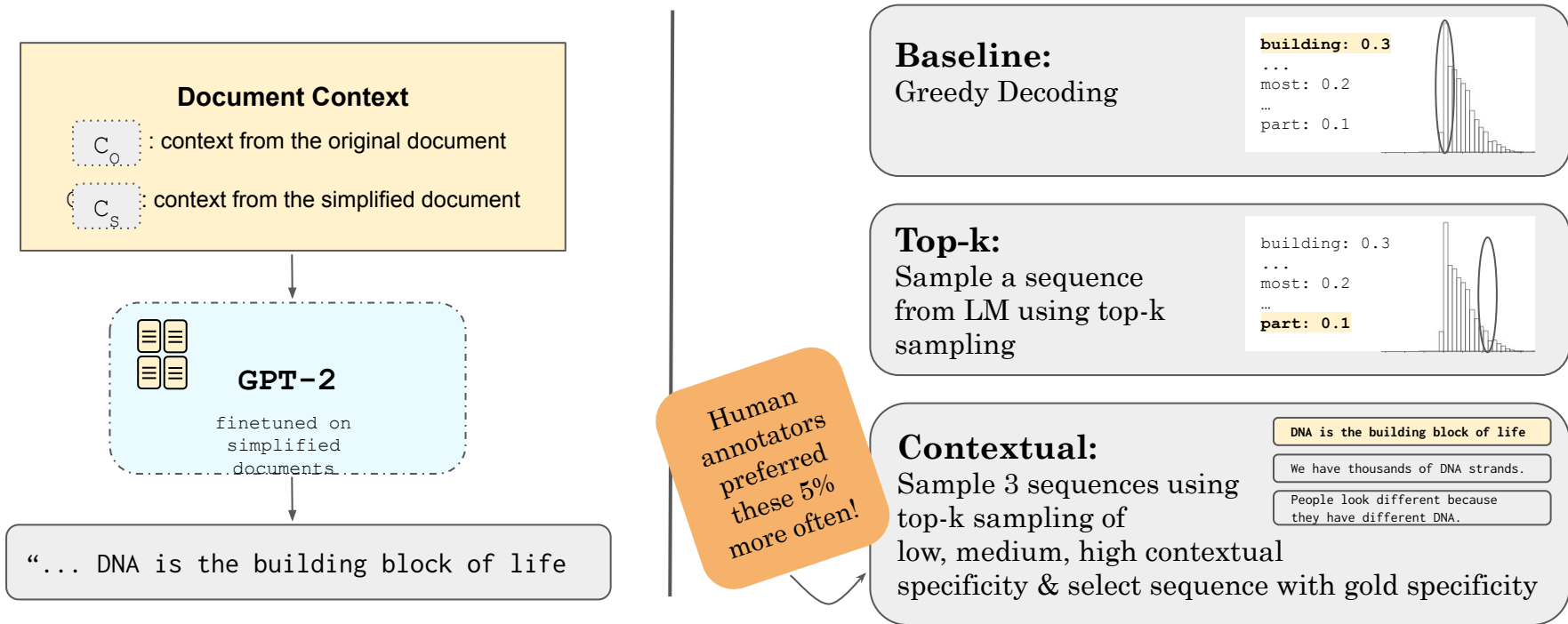We frame elaborative simplification as an NLG task, meant to mimic editors elaborating while simplifying documents.

Document Context

C_O : context from the original document

C_S : context from the simplified document

GPT-2

finetuned on simplified documents

"... DNA is the building block of life

Baseline:
Greedy Decoding

building: 0.3
...
most: 0.2
...
part: 0.1

Top-k:
Sample a sequence from LM using top-k sampling

building: 0.3
...
most: 0.2
...
part: 0.1

Human annotators preferred these 5% more often!

Contextual:
Sample 3 sequences using top-k sampling of low, medium, high contextual specificity & select sequence with gold specificity

DNA is the building block of life

We have thousands of DNA strands.

People look different because they have different DNA.

Example Result

Simplified Text

Because they look like a poisonous butterfly, they are left alone by predators. Looking like another species works just fine for them. While they avoid getting eaten, the male and the masculine-colored female are still at risk. For a long time, biologists have studied how animals try to look like other animals. It teaches them a lot about natural selection. Natural selection is sometimes called "survival of the fittest." **It means that animals that can change will survive.**

Generated Elaborations

Greedy: It is a way of thinking about how animals live.

Top-k: It means that the more an animal looks like another animal, the more likely it is to survive.

Contextual: It's a way of thinking about how organisms adapt to change.

Summary

Highlights

- Predicting the contextual specificity of elaborations (or the expected contextual specificity) is possible, but challenging!
- Informing a generation model with contextual specificity is helpful in both automatic and human evaluation.

Code & Annotation at: https://github.com/nehasrikn/elaborative-simplification

Request Newsela data at: https://newsela.com/data

Read our paper at: https://arxiv.org/pdf/2010.10035.pdf