1. Automatic Text Simplification

- **Rewrite** complex text into **simpler language** while retaining the original meaning.
- Treated as a **text-to-text generation** task.
- Involves **3 edit operations** - delete, split, and paraphrase.

According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.

- Problems with SOTA simplification systems:
  - Lack **controllability** in terms of edit operations.
  - Perform only **deletion**.

2. Our Approach: Controllable Text Generation

**Candidate Generation**

- Split the input sentence using 35 linguistic rules (Niklaus et al., 2019) + a seq2seq model.
- The exhibition, which opened Oct. 8 and runs through Jan. 3, features 27 self-portraits.

**Candidate Ranking**

- Rank all intermediate candidates.
- Loss function:
  \[
  L_{MR} = \frac{1}{m} \sum_{k=1}^{m} \frac{1}{n_k} \sum_{i=1}^{n_k} \max(0, 1 - d_{ij}^{y^*})
  \]

**Paraphrase Generation**

- Paraphrase top-ranked candidate.
- Data **Augmentation** with additional training data that focuses on lexical paraphrasing.
- Copy-control token as a soft constraint to control paraphrasing.
- **Auxiliary task** (if a word should be copied).

**Combine splits as candidates.**

3. Experiments

- Trained on **Newsela-Auto** (Jiang et al. 2020).
- 259,778 train + 32,689 valid + 33,391 test pairs.
- Human evaluation on 100 random simplifications.

4. Edit-focused Evaluation

- Evaluation on sections of Newsela-Auto test set.
- We report SARI, the main automatic metric of simplification.

- Control over 3 types of edit operations.
- Inject linguistic knowledge into neural models.
- New dataset to evaluate lexical paraphrasing.
- Evaluation setup for edit operations.

Code/data at [https://github.com/mounicam/controllable_simplification](https://github.com/mounicam/controllable_simplification)