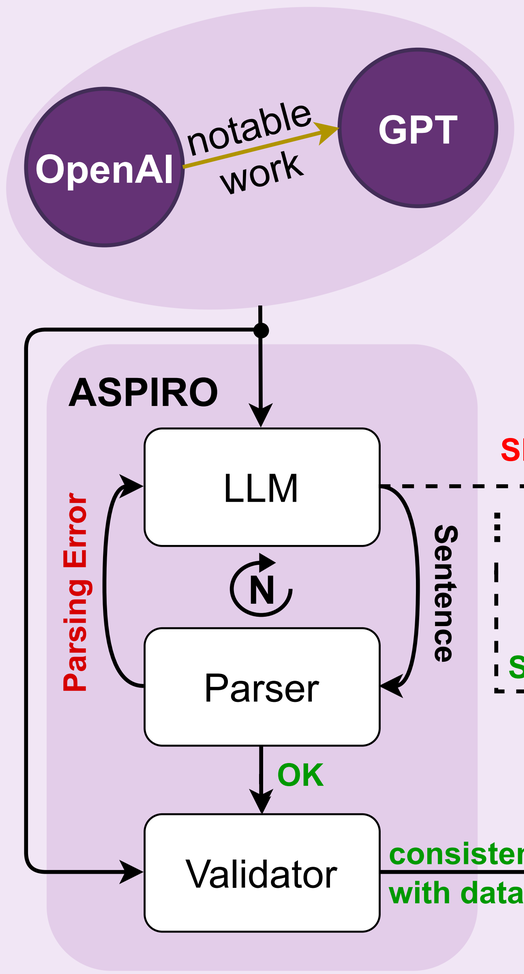
ASPIRO

Any-shot Structured Parsing-error-Induced ReprOmpting for Consistent Data-to-Text Generation

Martin Vejvar
Yasutaka Fujimoto

Graduate School of Engineering Science
Faculty of Engineering
YNU 横浜国立大学 YOKOHAMA National University



Background

- **Data to Text** using **language models** has two main branches:
  a) Train domain specific model: accurate, not universal
  b) Aim for **general domain**: no fine-tuning needed
- Recently, Large Language Models were applied successfully in the general domain.
- But simply prompting a LLM (**Shot 0**) is **prone to errors**.

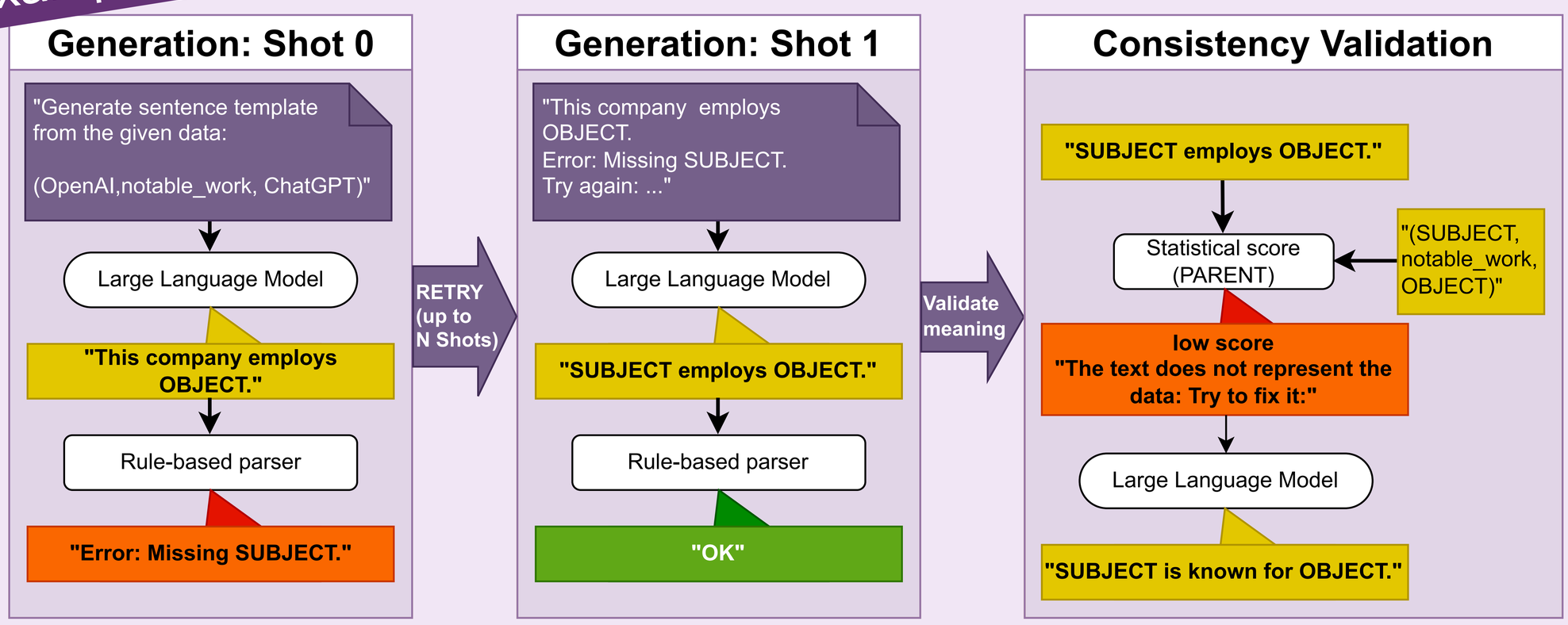
Problem

1) **Entity representation mismatch**
   - Leads to **parsing errors**
2) **Hallucinations / Omissions** of information
   - Text is **unfaithful to the facts** in the source data

Solution

LLMs + fixed rules + statistics

Example



Results

- ASPIRO with up to 5 retry Shots offers ca. **50% reduction in parsing errors** (Table 1) against simply prompting GPT3.5-Turbo (0 Shot) on DART dataset.
- Statistical metrics of ASPIRO on REL2TEXT dataset show it is **on par w/ fine-tuned PLMs** (Table 2).

Table 1: **Parsing error** counts /w GPT3.5-Turbo model on **DART** dataset (4299 samples in total).

	0 Shot	1 Shot	5 Shot
error count	996	727	497
error %	23.17	16.91	11.56

Table 2: Metrics of ASPIRO on **REL2TEXT test set** against REL2TEXT fine-tuned language model.

	BLEU	METEOR	BLEURT	NUBIA
fine-tuned BART (Kasner et al., EACL 2023)	**52.54**	44.86	0.54	**0.88**
5 Shot ASPIRO (w/ text-davinci-003)	51.40	**44.94**	**0.82**	0.87

Conclusion

- ASPIRO is **general domain data to text** pipeline.
- Tested on DART, REL2TEXT and WEBNLG, it consistently **produces simple factual sentences** from data triples.
- ASPIRO **reduces parsing errors** in text generated from data.
- Metrics show slight improvement over 0 Shot.
- **Customizable:**
  - choose your prompts
  - choose your LLMs
- **Limitations:**
  - single triples only
  - run time increase
  - cost increase

Questions?

vejvar-martin-km@ynu.jp