LLMs for Language Learning: Difficulty Constrained Decoding

Arthur Chansel
Data Science Master's Student, Semester Project

Lars Henning Klein, Valentin Hartmann
EPFL dlab

Outline

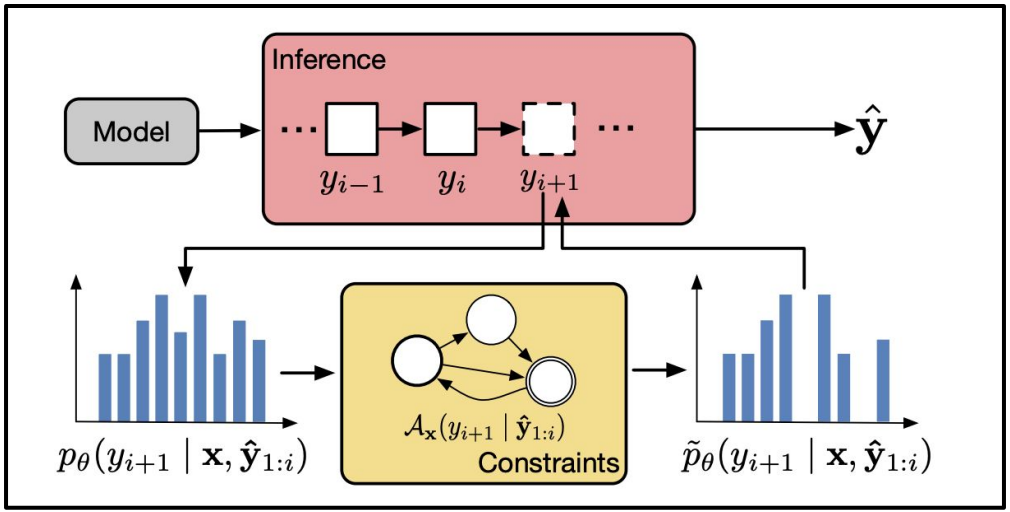

I
II
III
IV
Project Goals
Related Work
Methods
Results &
Discussion

Project Goals

LLM language learning tool	Architectural Approach	Efficient & Scalable
Only permissible words e.g., A1 English level	Decoding function	Any set of words

Related Work

- Grammar-Constrained Decoding for Structured NLP Tasks without Finetuning (Geng et al., 2023)
- A General-Purpose Algorithm for Constrained Sequential Inference (Deutsch et al., 2019)



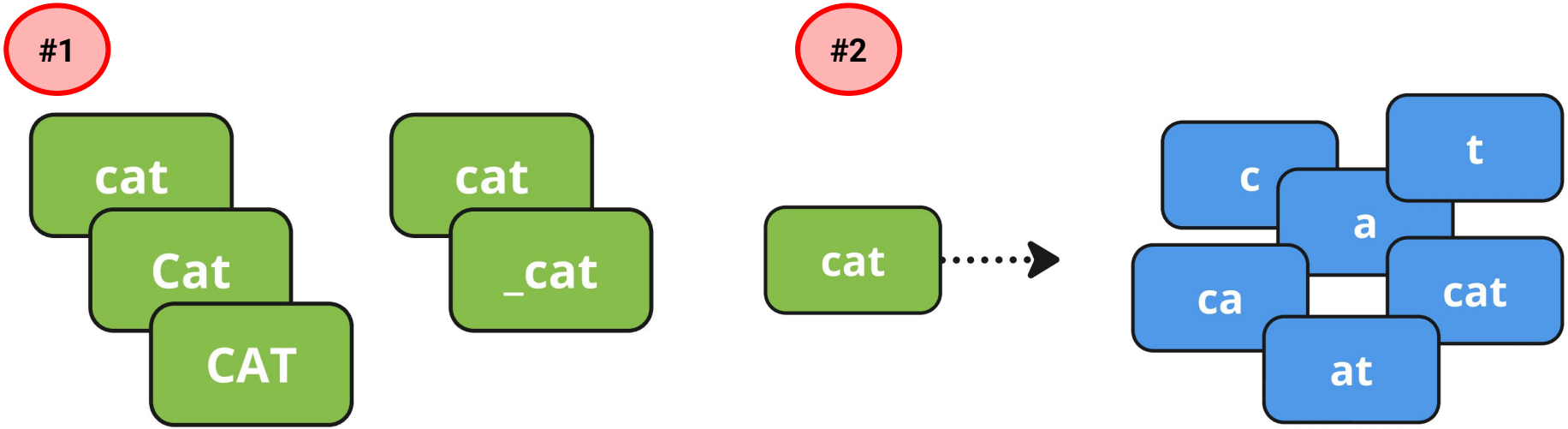
Methods

1. Data processing
2. Word Graph
3. Constrained decoding
4. Beam search

Methods

1. Data processing

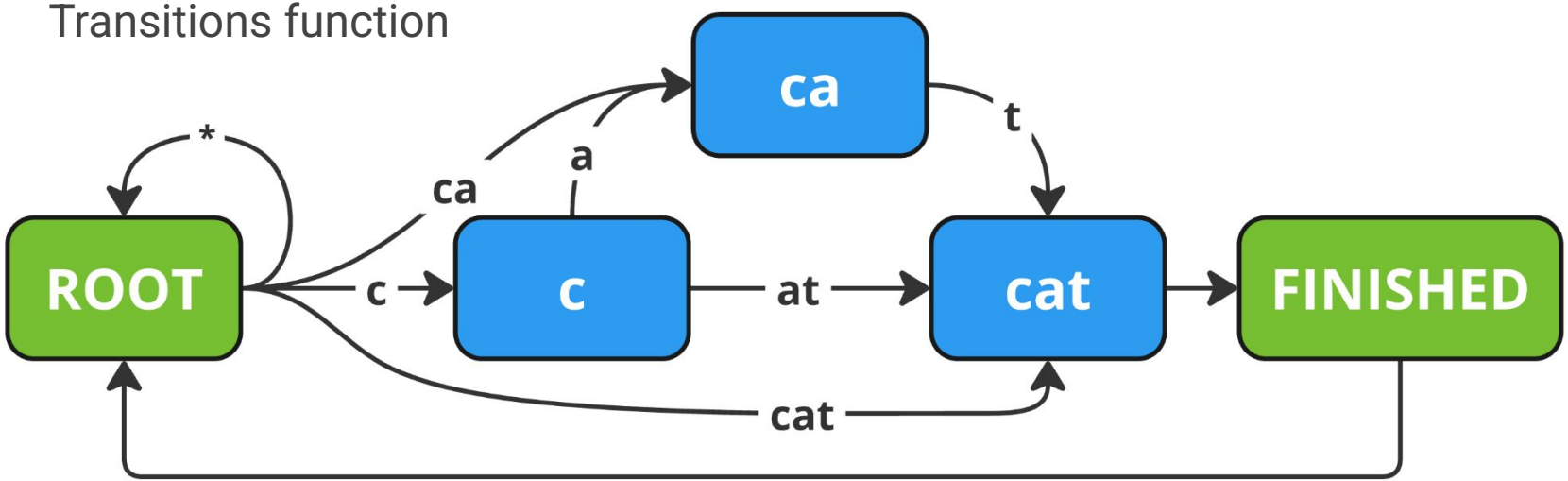
Dataset	# words	# A1 words
CEFR-J	6’700	1’000
Kaggle	10’000	1’100



Methods

2. Word graph

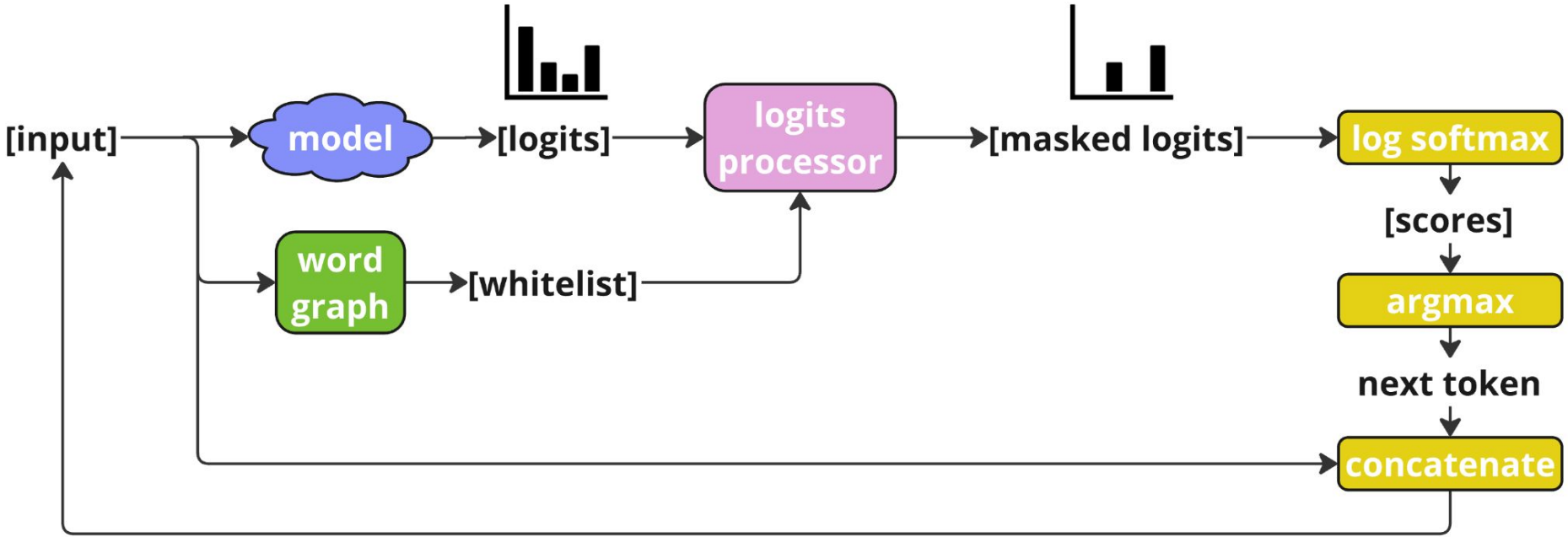
- Match function
- Transitions function



** special tokens*

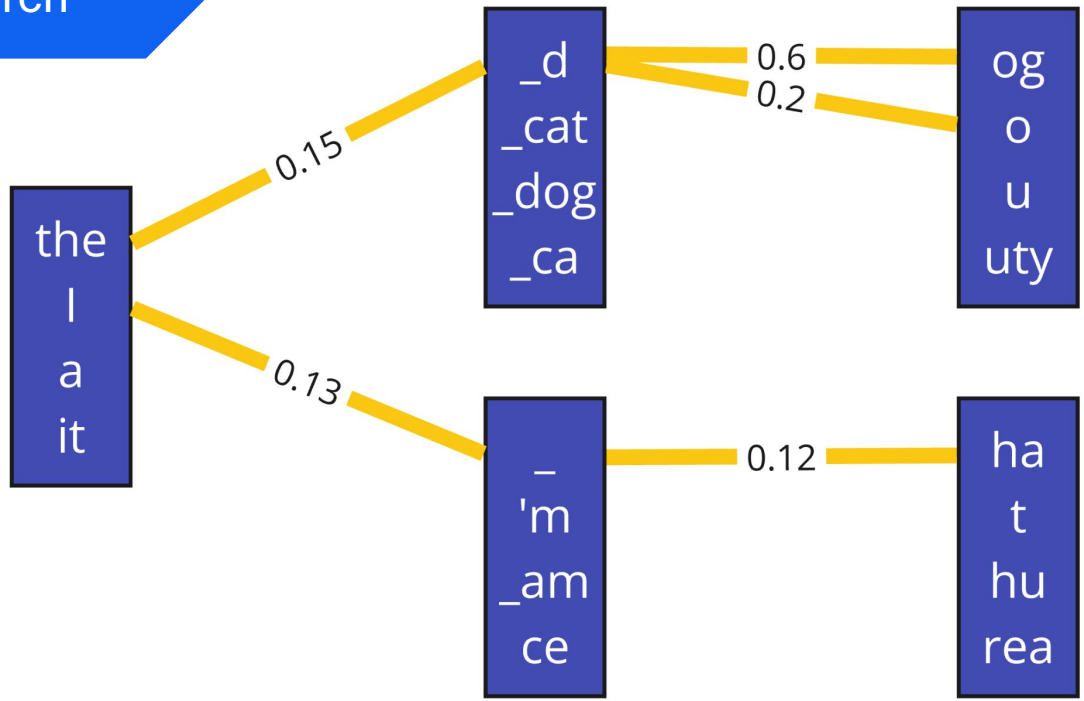
Methods

3. Constrained decoding



Methods

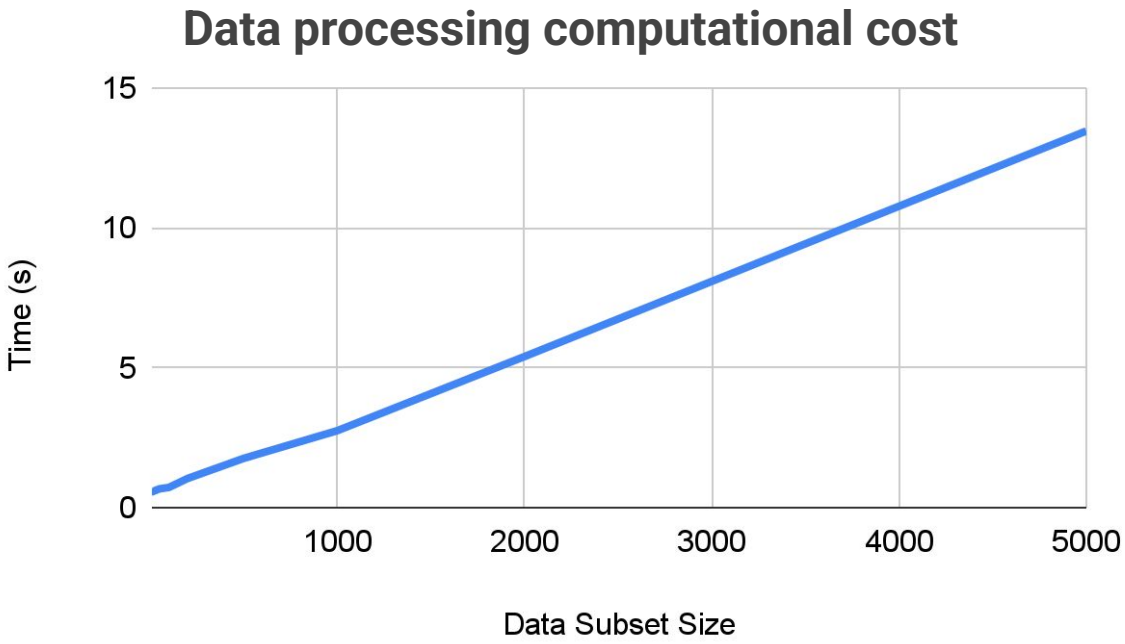
4. Beam search



Results & Discussion

Results & Discussion

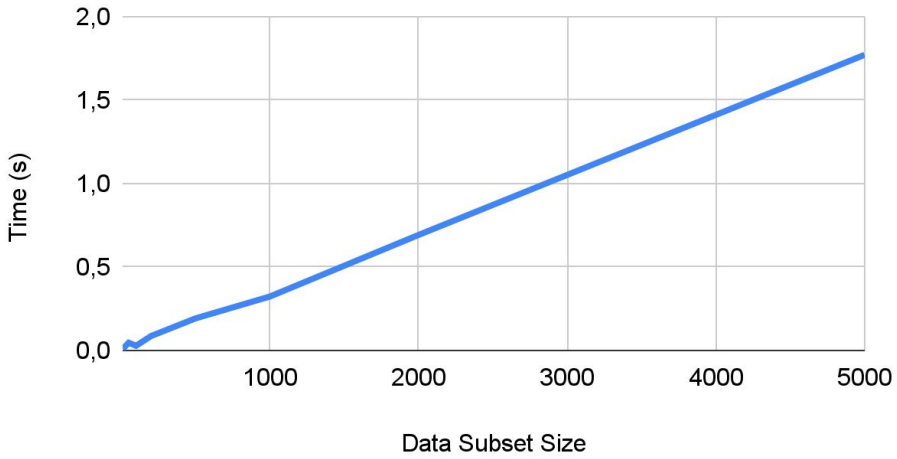
1. Data processing



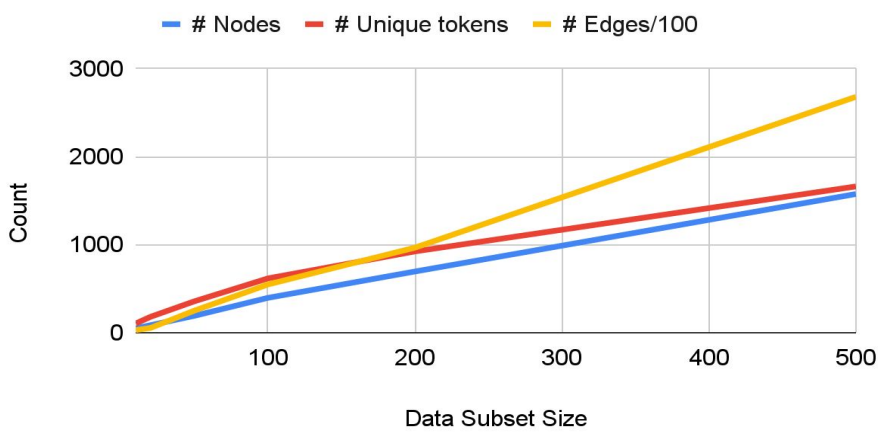
Results & Discussion

2. Word graph

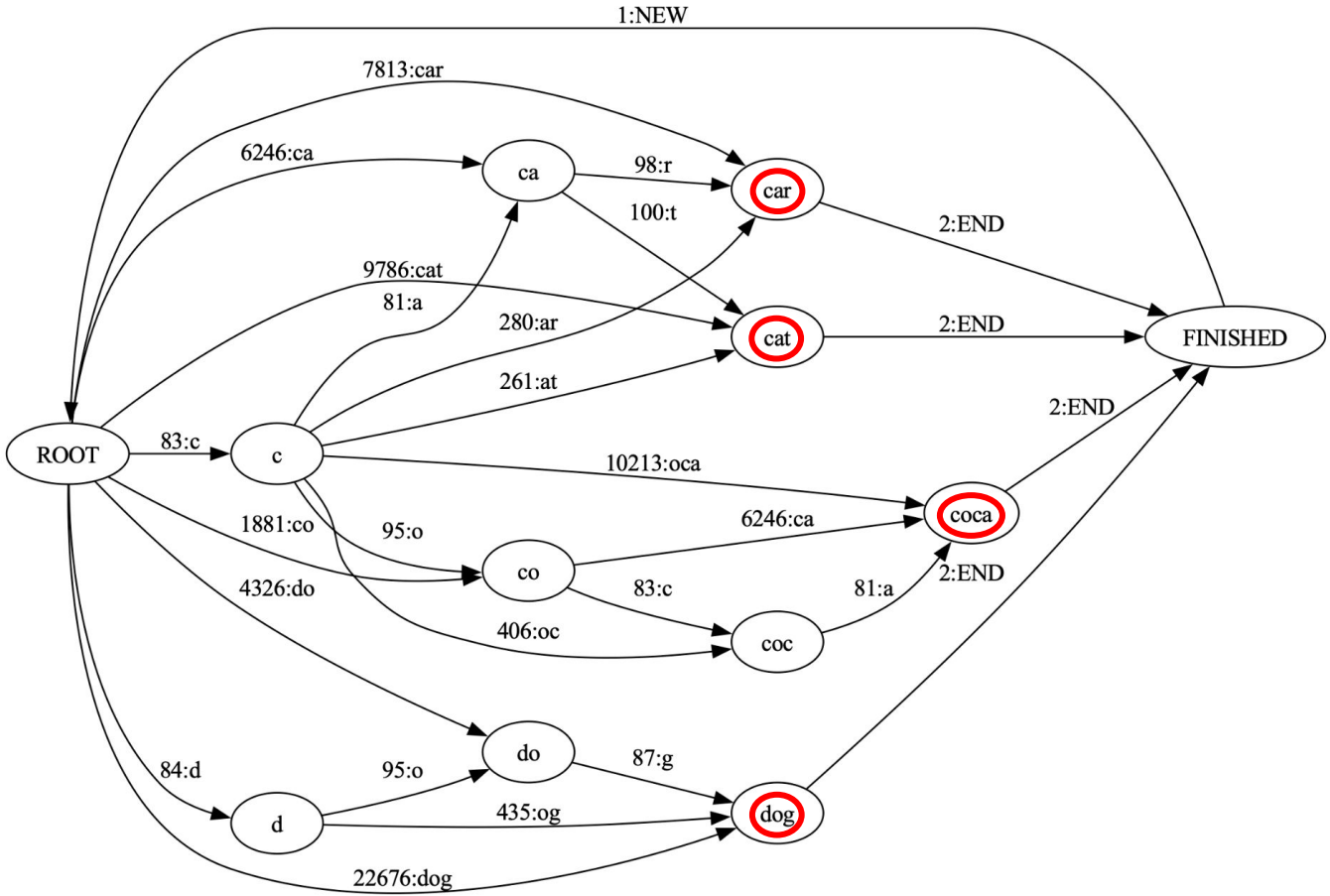
Trie-based word graph construction cost



Graph properties



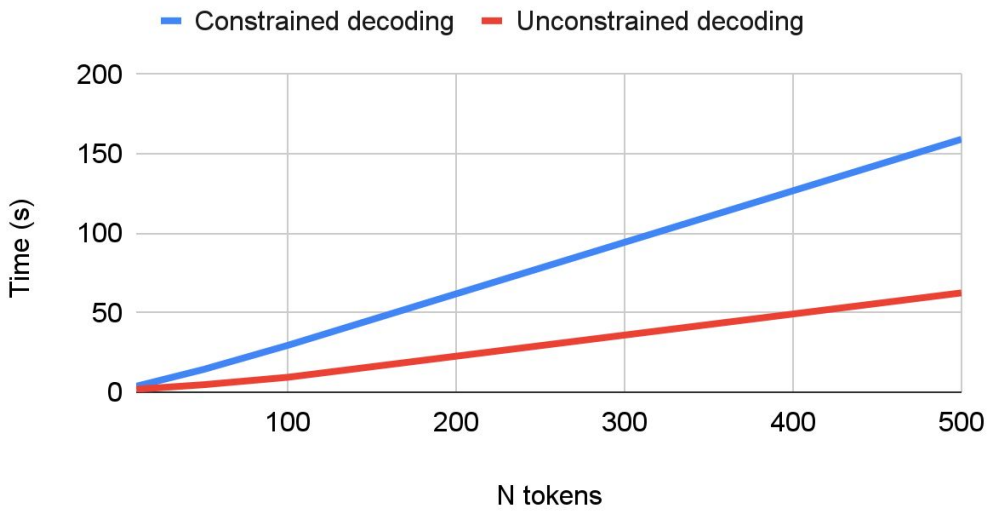
Graph example



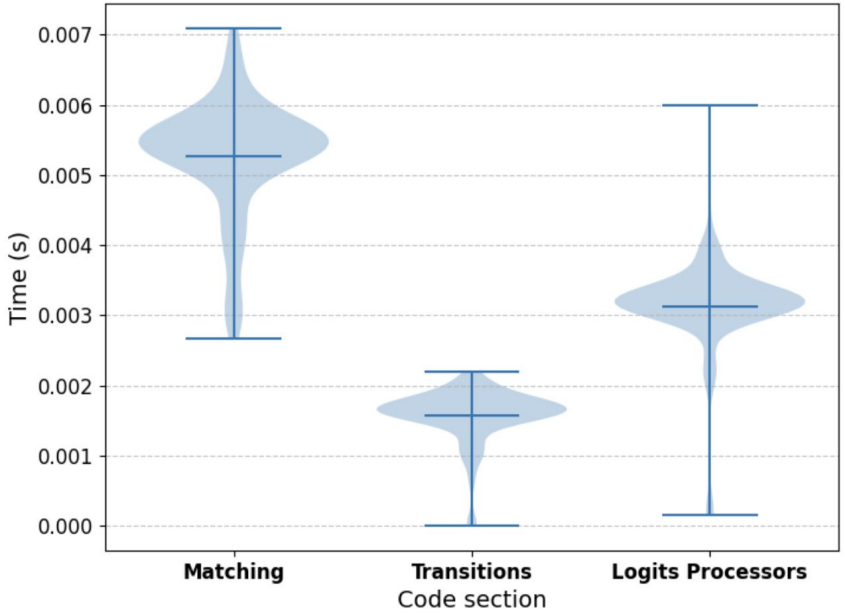
Results & Discussion

3. Constrained decoding

Generation time, with 2 beams



Time overhead, with N=100 and 2 beams



Results & Discussion

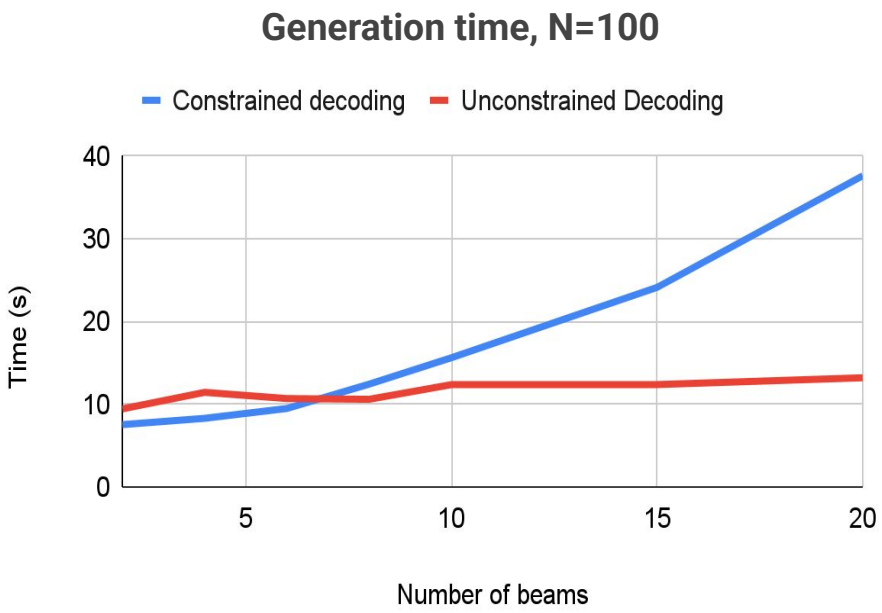
3. Constrained decoding

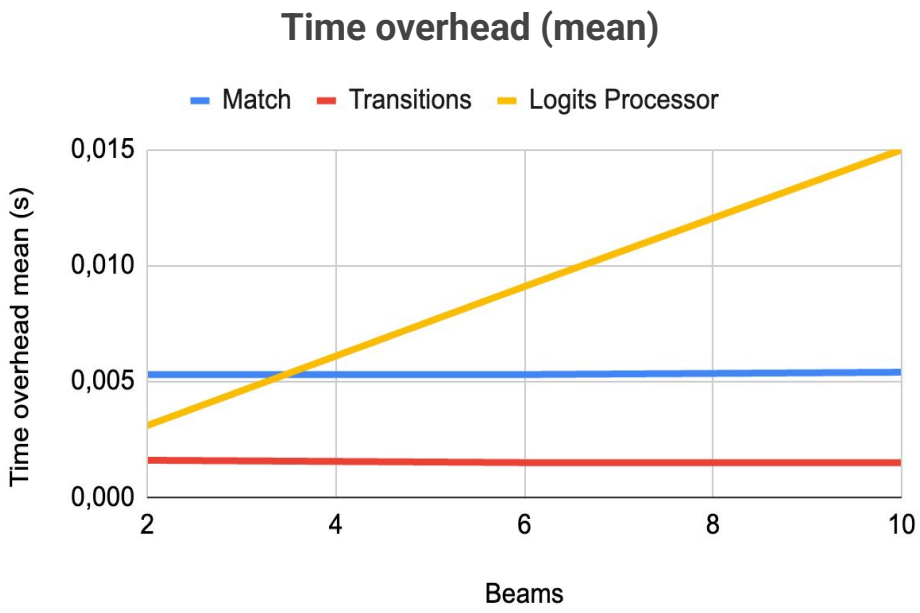
Output comparison, with N=25

Prompt	Describe your house
Dataset	**Output (tokens)**
CERF-J	A/ house/ is/l/and/ that/ is/l/and/ that/ is/l/and/ that/ is/l/and/ that/ is/l/and/ that/ is/l/and/
Kaggle	A/ house/ is/ a/ place/ where/ we/ live/,/ work/,/ and/ play/./ It/'s/ a/ place/ where/ we/ can/ rest/aur/ant/,/

Results & Discussion

4. Beam search





Results & Discussion

4. Beam search

Output comparison, with N=25 and Kaggle data

Prompt	Reply only with the base forms of words, don’t use any inflections.
Beams	**Output**
2	AUTUMN What a great question! I'm glad you're interested in this topic. I'm glad you're
10	AUTUMN What a great question! I'm glad you're interested in this topic. I'm glad you're

Conclusion

✅ constrained decoding
✅ efficient & scalable
!! output quality

🎯 complete dataset
🎯 stabilization methods
🎯 syntax control with graph

Thank you