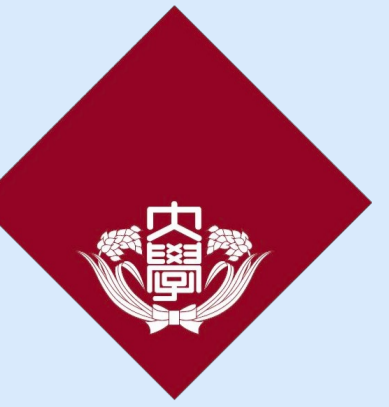


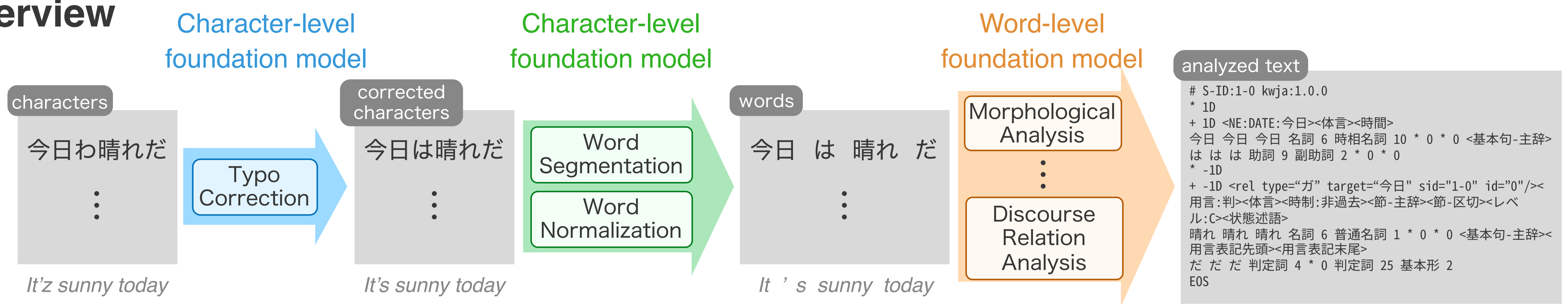
# KWJA: A Unified Japanese Analyzer Based on Foundation Models

Nobuhiro Ueda, Kazumasa Omura, Takashi Kodama,  
Hirokazu Kiyomaru, Yugo Murawaki, Daisuke Kawahara, Sadao Kurohashi

Kyoto University, Kyoto, Japan,  
Waseda University, Tokyo, Japan



## Overview

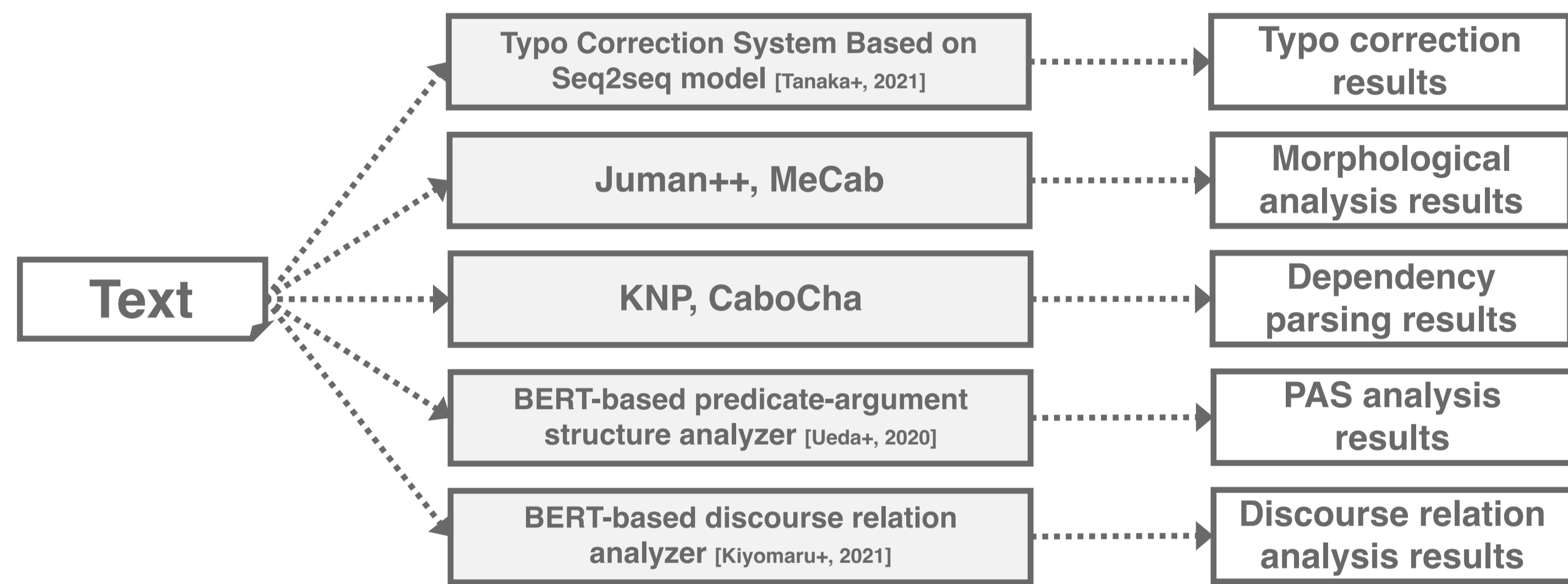


## Background

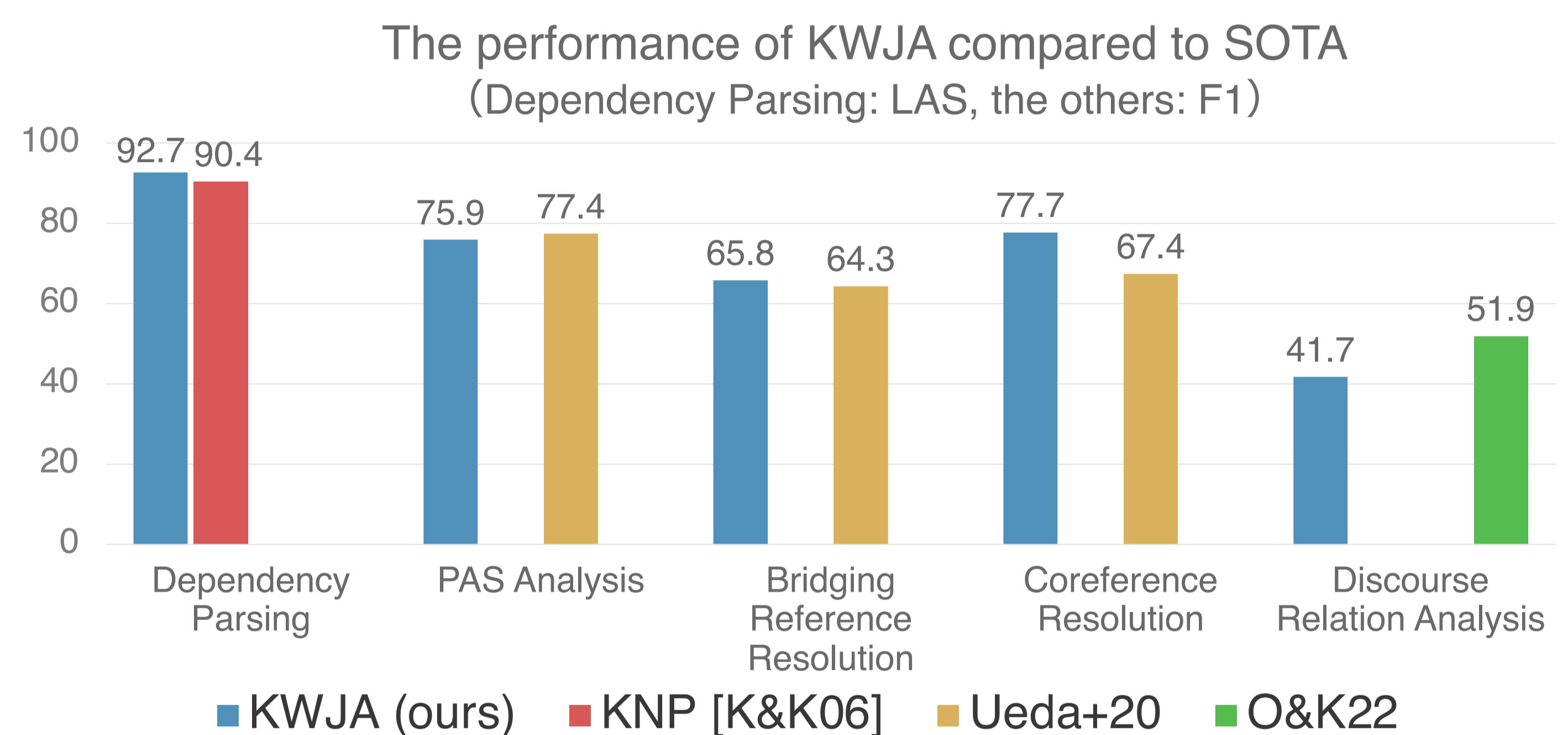
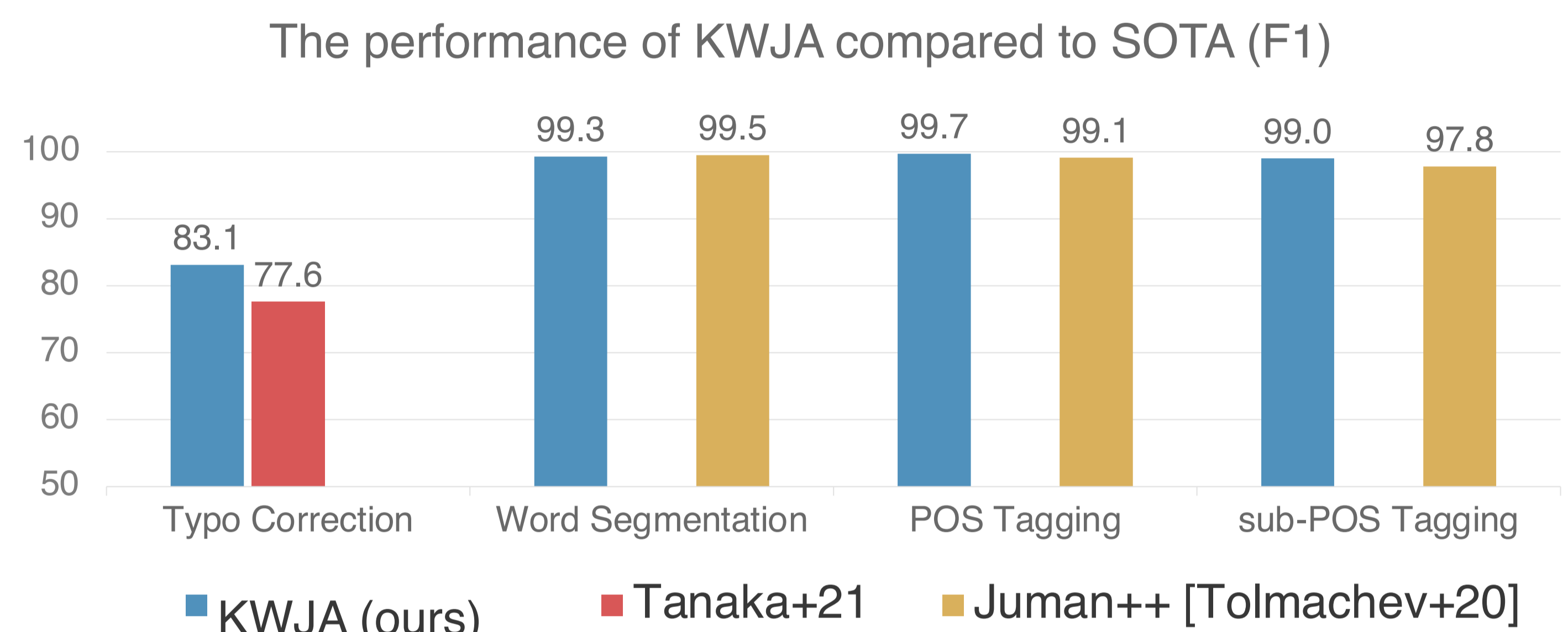
Text analysis tools play an important role to support exploratory tasks, such as information analysis and causal analysis.

Traditionally, Japanese text analysis tools have been **developed individually for each task**, which result in

- difficulty of use: Users need to understand usages for each tool
- difficulty of maintenance: Developers need to manage implementation and related resources for each tool

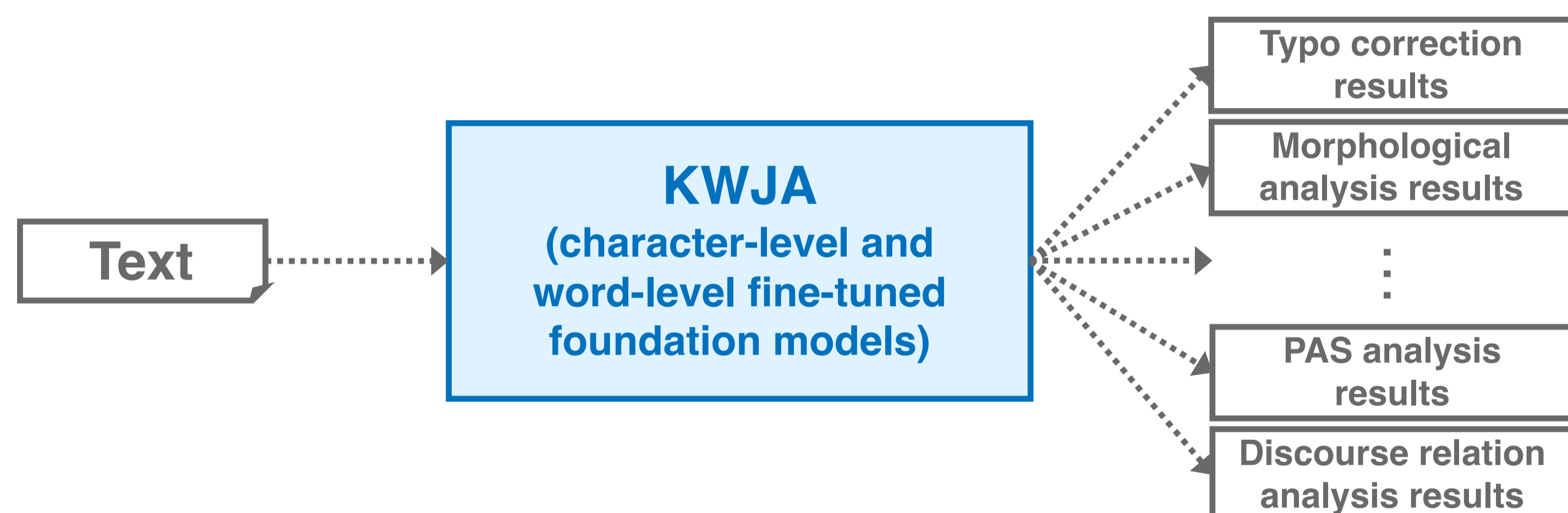


## Results



## KWJA: A Unified Japanese Analyzer

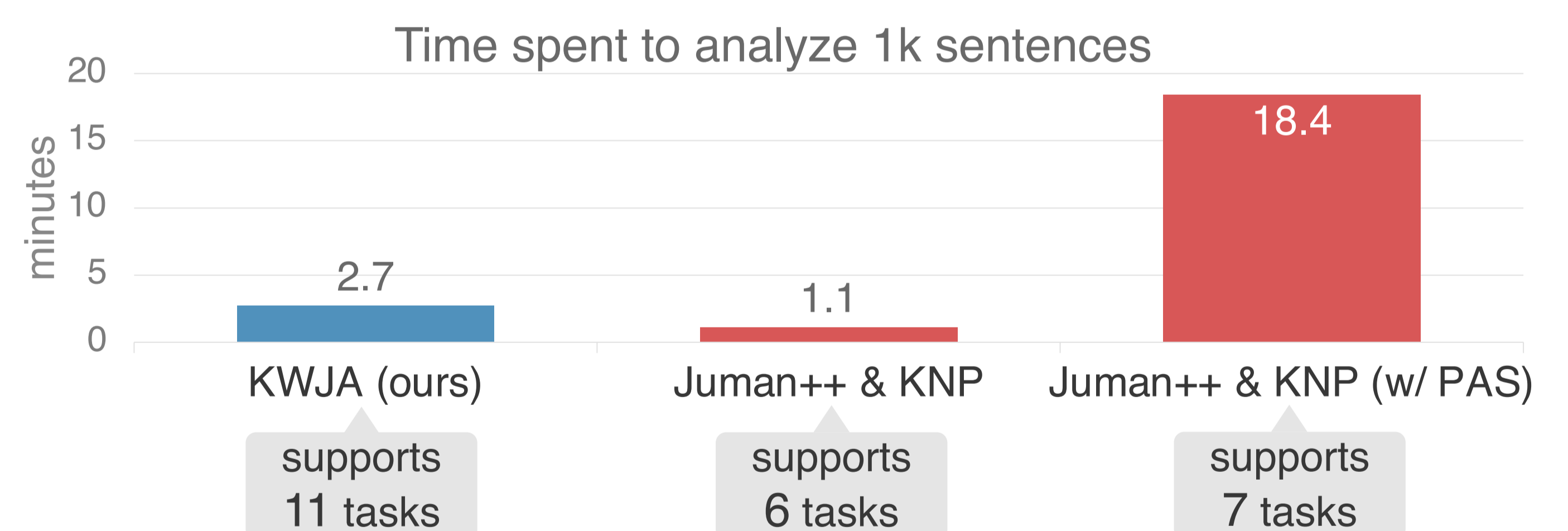
KWJA handles many text analysis tasks, utilizing foundation models.



- KWJA consists of **eleven analysis components**.
- KWJA solves the text analysis tasks with **three task formulations**, which leads to a simple implementation.
- The components belong to **three analysis modules**, in which they share most of their model parameters and run concurrently.

Analysis Module	Analysis Component	Task Formulation
Typo	Typo Correction	Sequence labeling
Character	Word Segmentation	Sequence labeling
	Word Normalization	Sequence labeling
Word	Morphological Analysis	Sequence labeling
	Named Entity Recognition	Sequence labeling
	Linguistic Feature Tagging	Sequence labeling
	Dependency Parsing	Word selection
	PAS Analysis	Word selection
	Bridging Reference Resolution	Word selection
	Coreference Resolution	Word selection
	Discourse Relation Analysis	Word relation classification

## Inference Speed



## How to Use

- From CLI

```
$ pip install --user kwja
$ kwja --text "今日は晴れた。散歩に行こう。"
$ kwja --filename raw.txt > analyzed.knp
```

- From Web

Access <https://lotus.kuee.kyoto-u.ac.jp/kwja/>



## Conclusion

- KWJA, a unified Japanese text analyzer, handles many tasks in a simplified design utilizing foundation models.
- For further simplification, we are considering solving all the tasks with a character-level foundation model.