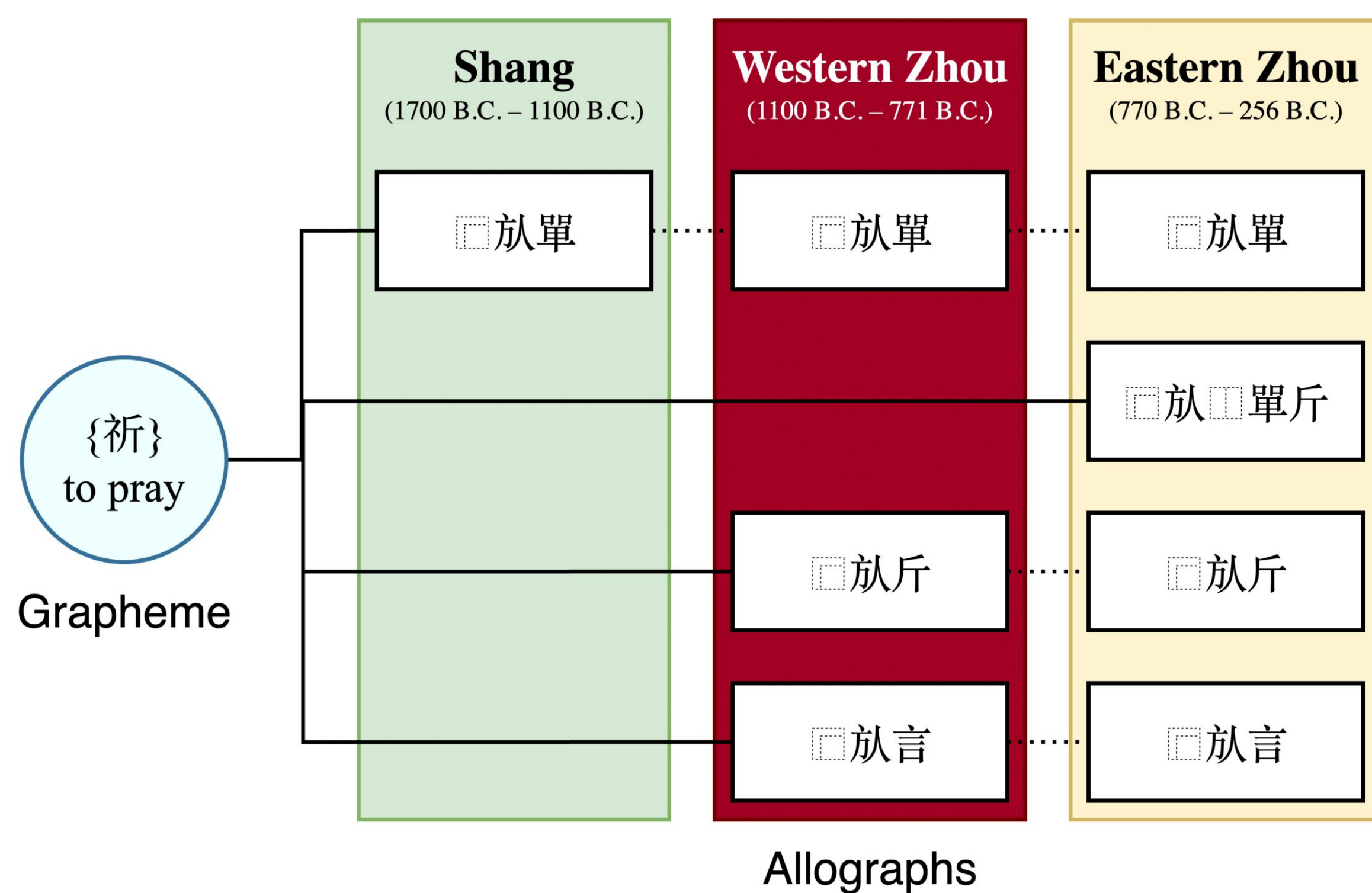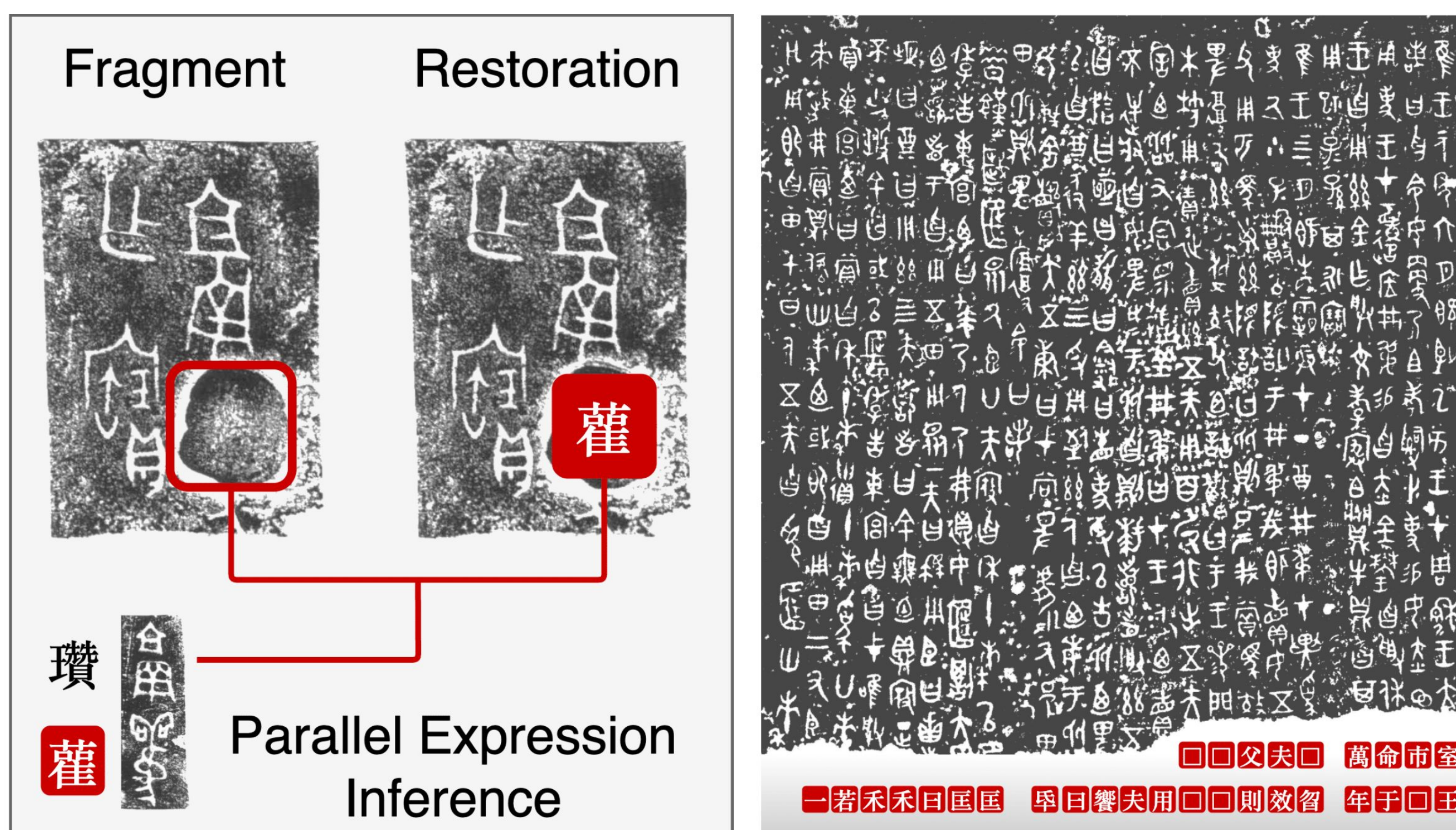# BIRD: Bronze Inscription Restoration and Dating

Wenjie Hua[1]   Hoang Huu Nguyen[2]   Gangyan Ge[3]

[1] Wuhan University
[2] University of Illinois, Chicago
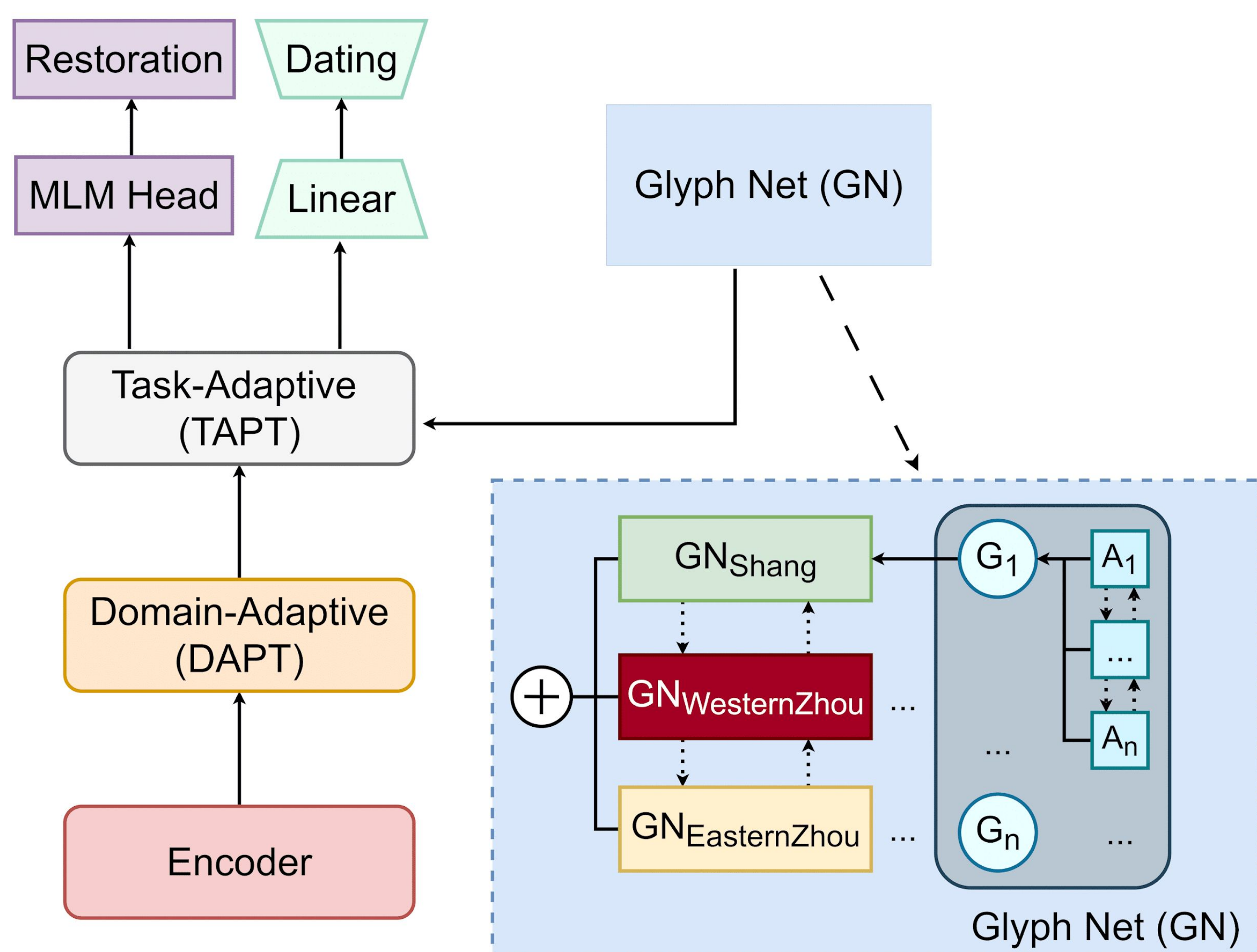[3] Xinjiang University

Motivation

Low-resource and allography make NLP on bronze inscriptions hard

Fragment   Restoration

瓚
蘱
Parallel Expression Inference

一若未末国国區
昂目繫夫用国国效智
年于国王

| Shang (1700 B.C. – 1100 B.C.) | Western Zhou (1100 B.C. – 771 B.C.) | Eastern Zhou (770 B.C. – 256 B.C.) |

{祈} to pray — Grapheme

Allographs

Methods

DAPT on Pre-Qin texts; **TAPT** on BIRD
+ Glyph Net (GN) (graphemes ↔ allograph)
+ Stride-based masking
+ Glyph-biased sampling

Evaluation

Restoration needs allographic equivalence
Dating needs diachronic differences

Model	Params	E@1	E@5	E@10	F@1	F@5	F@10
BiLSTM	20M	39.02	42.98	53.10	**57.41**	57.63	62.50
SikuRoBERTa	109M	**49.47**	**65.20**	**70.15**	54.32	**68.05**	**73.07**
mBERT	110M	43.55	58.57	63.71	46.93	61.28	65.92
XLM-Base	278M	43.51	58.35	62.94	44.28	59.49	64.03
XLM-Large	550M	45.64	60.92	64.91	47.16	61.17	65.36

Model	Params	Dynasty		Hier-Dyn		Hier-Per	
		Acc	F1	Acc	F1	Acc	F1
SVM	0.08M	75.31	49.44	76.32	42.67	58.55	49.43
SikuRoBERTa	109M	**86.42**	**77.83**	**84.21**	**54.32**	**67.11**	62.91
mBERT	110M	84.57	74.77	82.24	53.12	63.82	58.63
XLM-Base	278M	79.01	50.34	80.92	51.32	62.50	57.34
XLM-Large	550M	84.01	74.60	81.58	53.12	65.13	**62.96**

- **Restoration** (SikuRoBERTa + GN):
  - Exact@1 49.47%, Family@10 73.07%
- **Dating** (SikuRoBERTa + Bias):
  - Dynasty Acc 86.42%, Macro-F1 77.83
- **Hu Ding inscription** case study:
  - Exact@1 50% (11/22) + undeciphered slots
- **Ablation**
  - Restoration: GN ↑ + 2.0 p.p.
  - Dating: Glyph-biased sampling ↑ + 2.7 p.p.

Conclusion

- **BIRD**, dataset of encoded bronze inscriptions, cleaned, deduplicated, and chronologically labeled for NLP tasks
- **Glyph Net** (allographs → grapheme families)
- **Glyph-biased sampling** (diachronic variation)
- A foundation for future work

Dataset	Ava.	Dedup.	Filt.	Enc.	Chron.
Jihewang	✗	✗	✗	Partial	✓
AS DABII	✗	✗	✗	Partial	✓
AS Lexicon	✗	✗	✗	Partial	✓
BIRD	✓	✓	✓	Full	✓