Bronze Inscription Restoration

Anonymous EMNLP submission

Abstract

Bronze inscriptions from early China are often fragmentary, with missing or undeciphered characters limiting linguistic and historical analysis. Addressing this challenge requires models that can generalize across orthographic variation and diachronic script change. This paper introduces three contributions to support computational processing of bronze inscriptions: (i) a fully digitized and Unicode-encoded corpus of over 40,000 inscriptional characters; (ii) a glyph network linking diachronic variants to shared semantic anchors; and (iii) a masked language modeling (MLM) framework with variant-aware augmentation, alongside a periodization classification task. Experiments show that domain-adaptive pretraining and glyph-aware modeling substantially improve restoration accuracy. Our code is publicly available [1].

1 Introduction

Bronze inscriptions from the Chinese Bronze Age (c. 21st–3rd century BCE) are a primary source for the study of early Chinese writing, language, and state rituals. Found on ritual vessels, weapons, and musical instruments, these inscriptions record military achievements, feudal enfeoffments, oaths, and ancestral rites. Their forms vary in length and structure, and they offer direct evidence of pre-Qin language and institutional systems. However, many inscriptions are damaged or fragmentary due to age, and missing characters are common.

Traditionally, scholars restore missing characters by comparing graphic forms and using contextual inference. This process is time-consuming, expert-dependent, and difficult to scale. Recent work shows that pre-trained language models (PLMs) can support ancient script processing (Li, 2024), but most models are trained on modern Chinese and do not capture variant forms, phonetic substitutions, or syntactic evolution in early Chinese. This limits their effectiveness on bronze inscriptions, often leading to incorrect or implausible predictions.

In other domains, deep learning has demonstrated considerable success in restoring historical texts, including Akkadian cuneiform (Lazar et al., 2021) and ancient Greek inscriptions via the *Ithaca* model (Assael et al., 2022). However, no specialized system currently exists for the completion of Chinese bronze inscriptions, and general-purpose models cannot be directly applied due to the script's unique orthographic and diachronic characteristics. To bridge this gap, we introduce a masked language modeling framework for predicting missing characters in bronze inscriptions.

The contributions of this paper are threefold:

- It introduces the first fully digitized and Unicode-encoded corpus of Chinese bronze inscriptions to date, offering a structured foundation for computational modeling and downstream evaluation.
- It constructs a glyph network for Chinese bronze script that systematically aligns diachronic character variants with shared semantic anchors, facilitating variant-aware representation across historical periods.
- It develops a BERT-based masked language modeling framework for character reconstruction and period classification, explicitly tailored to the linguistic, orthographic, and diachronic properties of inscriptional Chinese.

2 Related Work

Corpus Compilation Bronze inscriptions have been extensively cataloged and interpreted by scholars throughout the past century. Foundational corpora were established by Rong (1985), Wu (2012), and researchers at CASS (2007), providing large-scale databases for further study. On the interpretive side, early contributions by Guo (1999)

[1] https://anonymous.4open.science/r/bir-0E4B/

Figure 1: An example of a damaged bronze inscription fragment (CCYZBI.02838) with annotation from us.

and Ma (1986), along with many subsequent philological studies, have clarified the content and structure of inscriptional texts. These efforts laid the groundwork for both traditional research and modern computational approaches.

Periodization The script style and grammar of bronze inscriptions change over time, making periodization a central task in the field. Chen (2004), Wang et al. (2017), and Du (2003) proposed widely adopted early-middle-late subdivisions of Western Zhou inscriptions based on inscription content and archaeological context. More recent approaches use finer features: Yan (2017) applies archaeological typology, while Deng (2015) explores dating by graphical components such as radicals and stroke patterns.

Digitization Multiple platforms now support computational access to bronze inscriptions. ECNU's system combines structured inscription catalogs with AI-based glyph recognition and transcription tools[2]. Academia Sinica maintains two complementary resources: a GIS-linked corpus of over 14,000 inscriptions[3] and a lexical database with concordance-style search across oracle, bronze, and bamboo texts[4]. However, many characters remain undeciphered or image-based, limiting downstream analysis.

Generalized Characters Bronze texts exhibit widespread use of phonetic loans, allographs, and graphic variants, which complicates token-level modeling. Scholars have proposed generalized character mappings to unify semantically equivalent forms. Existing studies span Shang and Zhou corpora and document systematic correspondences across hundreds of variant graphs (Luo, 2013; Du, 2020; Qi, 2023). These mappings are crucial for reducing surface-level noise while preserving linguistic structure.

Restoration of Historical Texts Neural models have been successfully applied to restoring fragmentary ancient texts across diverse languages. Assael et al. (2022) introduced *Ithaca*, a transformer trained to complete Greek inscriptions, achieving 62% top-1 accuracy, improved to 72% with human collaboration. Lazar et al. (2021) framed Akkadian cuneiform reconstruction as a masked language modeling task, reaching 89% top-5 accuracy with limited training data. For ancient Arabic manuscripts, Miloud et al. (2024) proposed a BLSTM with modified attention, achieving 99.5% accuracy on 3,745 curated samples. These results show that neural restoration methods are effective even under low-resource conditions, and can meaningfully support human experts in epigraphic reconstruction.

3 Dataset

Domain-Adaptive Pretraining (DAPT) Following the principle of pretraining (Gururangan et al., 2020), we adapt a general-purpose encoder to the domain of early Chinese by continuing pretraining on 40 curated pre-Qin texts, including received classics (e.g., *Shangshu*, *Bamboo Annals*) and excavated manuscripts (e.g., Guodian slips, Mawangdui silk books). These sources provide syntactic and lexical coverage of the language continuum most proximate to bronze inscriptions. Full source details are listed in the Appendix 4.

Task-Adaptive Pretraining (TAPT) Our task-specific corpus builds on the *Complete Collection of Yin and Zhou Bronze Inscriptions* (CCYZBI), which compiles all known inscriptions published before 2007. We extend this base with philological updates from recent scholarship (e.g., Su (2016), Zhu (2007), Li (2023)), and annotate each inscription with temporal metadata. All transcriptions and

[2] https://jwdcdbz.ancientbooks.cn/index
[3] https://bronze.asdc.sinica.edu.tw
[4] https://inscription.asdc.sinica.edu.tw/c_index.php

Type	Count	Proportion
Identifiable	39,565	99.24%
Ambiguous (□)	236	0.59%
Unknown ([UNK])	56	0.14%

Table 1: Character types in the TAPT corpus.

datings were reviewed by trained paleographers to ensure accuracy.

Corpus Encoding All inscription data are normalized into machine-readable Unicode. We distinguish three character types: (1) parsed characters with known readings; (2) ambiguous characters marked as □ where glyphs are indistinct or damaged; and (3) unparsed yet visually legible characters, encoded as [UNK-xxxxx-x] for future recognition models. See Table 1 for corpus composition.

4 Model

4.1 Glyph Modeling

Inscriptions from different periods often render the same semantic unit in distinct glyphs (Qiu, 2013). We group such variants into realization clusters, where RealizationCluster(α) = $\alpha_1, \alpha_2, \alpha_3$ denotes the observed forms of semantic anchor α across time. Rather than collapsing this diversity through hard substitution, we hope to align these forms to a shared latent identity during encoding, and allow the model to generalize across periods while preserving surface sensitivity.

4.2 Glyph Net

To operationalize glyph clustering, we compile a cross-period inventory of character variants from the Shang to Eastern Zhou periods. Drawing on prior studies (Luo, 2013; Du, 2020; Qi, 2023), we assign Unicode-compatible IDs to normalized forms. This glyph net supports variant-aware encoding, augments training via injection strategies, and preserves diachronic structure. An example is shown in Appendix 3.

4.3 Training Architecture

We propose a BERT-based masked language modeling (MLM) framework for missing character reconstruction in bronze inscriptions. BERT has shown strong performance across NLP tasks due to its bidirectional context modeling, but it typically requires large-scale training data. This poses a challenge for our low-resource setting.

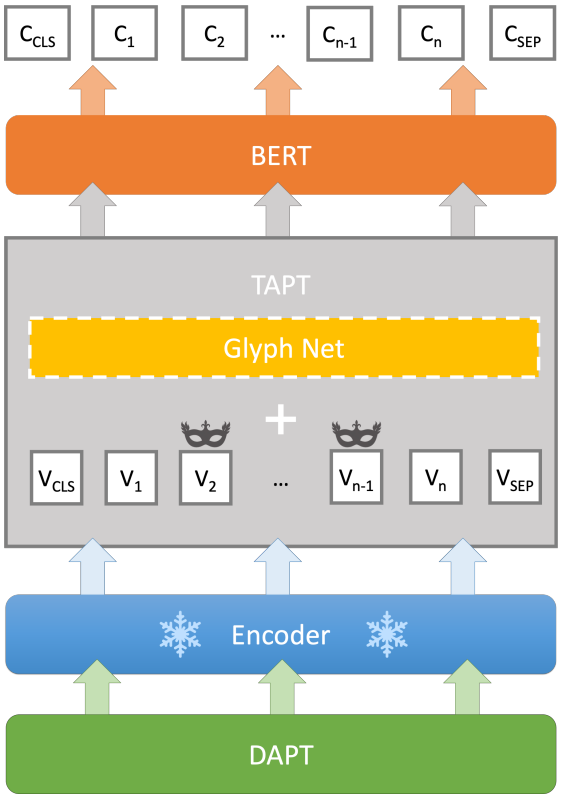

Figure 2: Overall framework of our masked language modeling for bronze inscription restoration.

As shown in Figure 2, our architecture integrates domain-adaptive pretraining with variant-aware augmentation. We first perform domain pretraining on a corpus of pre-Qin literature to expose the model to relevant syntactic and lexical patterns.

To handle glyph variation, we explore two augmentation strategies: **replace**, which maps all variants to the normalized form, and **inject**, which randomly replaces tokens with members of their realization cluster. The latter encourages robustness to surface variation and supports generalization across unattested forms.

5 Experiments

5.1 Setup

We benchmark two pretrained encoders: **multilingual BERT (mBERT)** [5], which is trained across 104 languages including Chinese, and **SIKU-BERT** [6], a domain-specific variant trained on Classical Chinese corpora such as the *Siku Quanshu*.

For MLM tasks, we evaluate Top-1 and Top-10 accuracy on character reconstruction and the

[5] https://huggingface.co/google-bert/bert-base-multilingual-cased

[6] https://huggingface.co/SIKU-BERT/sikubert

Model	DAPT	GN	Top-1 / -10
mBERT	–	replace	.4580 / .6270
	–	inject	.4797 / .6531
	+	replace	.4570 / .6378
	+	inject	**.4833 / .6580**
SIKU-BERT	–	replace	.5049 / .6871
	–	inject	.5353 / .7290
	+	replace	.5111 / .7012
	+	inject	**.5420 / .7263**

Table 2: Masked character prediction accuracy (Top-1 / Top-10). All settings include TAPT. DAPT = domain-adaptive pretraining; GN = glyph net.

Classifier	Dynasty	Period (avg)
Logistic Regression	.7407	.5003
Naive Bayes	.7222	.5371
Linear SVM	**.7840**	**.5932**
Random Forest	.7747	.5661

Table 3: Dynasty and period-level classification accuracy (mean).

impact of DAPT and TAPT, both with and without GN. Following standard practice in low-resource modeling, we freeze the encoder during DAPT and fine-tune during TAPT. Details of the corpora used are provided in Appendix A.1.

For periodization, we formulate diachronic classification as a hierarchical prediction task over four major periods (Shang, Western Zhou, Spring and Autumn, Warring States) nested within two dynasties (Yin, Zhou). We evaluate four classifiers—Logistic Regression, Naive Bayes, Linear SVM, and Random Forest.

5.2 Masked Language Modeling

Table 2 presents masked character prediction accuracy for mBERT and SIKU-BERT with DAPT and GN variants. TAPT consistently improves baseline performance. GN (inject) yields higher accuracy than GN (replace), reaching .4833 (Top-1) for mBERT and .5420 (Top-1) for SIKU-BERT. SIKU-BERT uniformly outperforms mBERT, confirming domain-specific advantages.

5.3 Periodization Classification

Classification results appear in Table 3. Linear SVM obtains highest accuracy (.7840 dynasty; .5932 period-level average). Random Forest closely follows; Naive Bayes and Logistic Regression lag behind. Lower period-level scores highlight the difficulty of fine-grained temporal classification.

6 Discussion

Results from both tasks indicate three key factors influencing performance: domain adaptation, GN normalization, and pretraining source. DAPT consistently improves accuracy across configurations, with mBERT benefiting most and substantially closing the gap with SIKU-BERT. This confirms that exposure to inscriptional data enables models to acquire lexical and syntactic patterns specific to bronze texts.

GN normalization further enhances prediction. Injecting variant forms (GN-inject) outperforms strict replacement (GN-replace), demonstrating that preserving surface variation improves generalization to both attested and unattested forms. In contrast, replacement reduces diversity in training signals and limits the model's ability to recover historically accurate glyphs.

Pretraining source exerts the strongest effect. SIKU-BERT, pretrained on classical Chinese, consistently outperforms mBERT under all conditions. Its alignment with the target domain's language and script structure leads to more stable and accurate predictions, both before and after adaptation.

7 Conclusion

This paper presents two new resources for the study and modeling of Chinese bronze inscriptions: a fully digitized, Unicode-encoded corpus of pre-Qin inscriptions, and a glyph network that links diachronic character variants to shared semantic anchors. These structured lexical tools provide a foundation for variant-aware modeling and enable robust generalization across script forms.

Built on these resources, a masked language modeling framework is proposed for character reconstruction and period classification. By incorporating domain- and task-adaptive pretraining, along with glyph-level augmentation, the model effectively handles orthographic variation and data sparsity. Experimental results confirm that glyph-aware strategies and domain alignment significantly enhance performance, underscoring the value of combining historical lexical structure with modern neural methods in low-resource epigraphic NLP.

8 Limitations

Despite promising gains in both reconstruction and periodization, several limitations remain. First, the current framework does not incorporate archaeological typology or vessel metadata, which play a central role in traditional period determination (Wang et al., 2017). Features such as vessel shape, decorative motifs, and casting techniques constitute an independent chronological signal and could significantly improve classification when integrated with textual modeling.

Second, the treatment of graphic variation relies on a manually curated glyph network and surface-level augmentation. While effective for known variants, this strategy may fail to capture more complex historical substitution patterns, particularly those involving phonetic loans. The current model lacks explicit phonological supervision, limiting its ability to resolve homophonic ambiguity—a common phenomenon in bronze script. Incorporating phonetic embeddings or diachronic sound correspondences may offer better disambiguation in future work.

Third, data scarcity remains a core constraint. Although the corpus released here represents the most complete Unicode-based resource for bronze inscriptions to date, it still covers only a portion of the known epigraphic record. Many inscriptions remain lost, unpublished, or undeciphered. This sparsity hinders generalization, especially for rare or stylistically irregular cases. Techniques such as synthetic augmentation, cross-modal training with image data, or semi-supervised learning on partial transcriptions could help mitigate this limitation.

References

Yannis Assael, Thea Sommerschield, Brendan Shillingford, et al. 2022. Restoring and attributing ancient texts using deep neural networks. *Nature*, 603(7900):280–283.

CASS. 2007. *Yīn Zhōu Jīnwén Jíchéng* 殷周金文集成 *[Complete Collection of Yin and Zhou Bronze Inscriptions (CCYZBI)]*. Zhōnghuá Shūjú 中華書局, Běijīng 北京.

Mengjia Chen. 2004. *Xī Zhōu Tóngqì Duàndài* 西周銅器斷代 *[Dating Western Zhou Bronzes]*. Zhōnghuá Shūjú 中華書局, Běijīng 北京.

Kai Deng. 2015. Jīnwén zìxíng gòujiàn duàndài fǎ chūtàn 金文字形構件斷代法初探 [a preliminary study on dating bronze characters by graphical components]. *Yīndū Xuékān* 殷都學刊 *(Journal of Yindu Studies)*.

Naisong Du. 2003. Qīngtóngqì de fēnqī yǔ duàndài 青銅器的分期與斷代 [periodization and dating of bronze inscriptions]. *Jíjīn Wénzì yǔ Qīngtóng Wénhuà Lùnwénjí* 吉金文字與青銅文化論集.

Yinqin Du. 2020. Shāngdài jīnwén tōngyòngzì zhěnglǐ yǔ yánjiū 商代金文通用字整理與研究 [studies on generalized characters in shang bronze inscriptions]. Master's thesis, Southwest University 西南大學, Chongqing 重慶.

Moruo Guo. 1999. *Liǎng Zhōu Jīnwén Cí Dàxì Túlù Kǎoshì* 兩周金文辭大系圖錄考釋 *[Annotated Catalog of Zhou Inscriptions]*. Shànghǎi Shūdiàn上海書店, Shànghǎi 上海.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8342–8360, Online. Association for Computational Linguistics.

Koren Lazar, Benny Saret, Asaf Yehudai, Wayne Horowitz, and Gabriel Stanovsky. 2021. Filling the gaps in ancient akkadian texts: A masked language modelling approach. *arXiv preprint arXiv:2109.04513*.

Chuntao Li. 2024. Rén'gōng zhìnéng yǔ jīnwén yánjiū zhǎnwàng 人工智能與金文研究展望

prospects for integrating artificial intelligence and bronze inscription research

. *Chinese Social Sciences Today*. https://www.cssn.cn/skgz/bwyc/202408/t20240809_5769948.shtml.

Xueqin Li. 2023. *Jīnwén yǔ Xī Zhōu Wénxiàn Hézhèng* 金文與西周文獻合證 *[Bronze Inscriptions and Western Zhou Textual Corroboration]*. Tsinghua University Press 清華大學出版社, Beijing 北京.

Tingting Luo. 2013. Dōngzhōu jīnwén tōngjiǎzì 東周金文通假字 [phonetic loan characters in eastern zhou bronze inscriptions]. Master's thesis, Yunnan University 雲南大學, Kunming 昆明.

Chengyuan Ma. 1986. *Shāng Zhōu Qīngtóngqì Míngwén Xuǎn* 商周青銅器銘文選 *[Selected Inscriptions of Shang and Zhou Bronzes]*. Wénwù Chūbǎnshè 文物出版社, Běijīng 北京.

Kamline Miloud, Moulay Lakhdar Abdelmounaim, Beladgham Mohammed, and Bendjillali Ridha Ilyas. 2024. Restoration of ancient arabic manuscripts: A deep learning approach. *Studies in Engineering and Exact Sciences*, 5(2):1–22.

Ruihua Qi. 2023. Xīzhōu jīnwén tōngjiǎ guānxì zhěnglǐ yǔ yánjiū 西周金文通假關係整理與研究 [collation and study of tongjia in western zhou bronze inscriptions]. Master's thesis, Jilin University 吉林大學, Changchun 長春.

Xigui Qiu. 2013. *Wénzìxué Gàiyào* 文字學概要 *(Introduction to Chinese Paleography)*. Shāngwù Yìnshūguǎn 商務印書館, Běijīng 北京.

Geng Rong. 1985. *Jīnwén Biān* 金文編 *[Compendium of Bronze Inscriptions]*. Zhōnghuá Shūjú 中華書局, Běijīng 北京.

Wenying Su. 2016. *Xī Zhōu Jīnwén Yìtǐzì Yánjiū* 西周金文異體字研究 *[A Study of Variant Characters in Western Zhou Bronze Inscriptions]*. Phd dissertation, Southwest University 西南大學, Chongqing 重慶.

Shimin Wang, Gongrou Chen, and Changshou Zhang. 2017. *Xī Zhōu Qīngtóngqì Fēnqī Duàndài Yánjiū* 西周青銅器分期斷代研究 *[Research on the Periodization of Western Zhou Bronzes]*. Wénwù Chūbǎnshè 文物出版社, Běijīng 北京.

Zhenfeng Wu. 2012. *Shāng Zhōu Qīngtóngqì Míngwén Jí Túxiàng Jíchéng* 商周青銅器銘文暨圖像集成 *[Corpus of Inscriptions and Images on Shang and Zhou Bronzes]*. Shànghǎi Gǔjí Chūbǎnshè 上海古籍出版社, Shànghǎi 上海.

Zhibin Yan. 2017. *Shāngdài Qīngtóngqì Míngwén Yánjiū* 商代青銅器銘文研究 *[Study of Shang Bronze Inscriptions]*. Shànghǎi Gǔjí Chūbǎnshè 上海古籍出版社, Shànghǎi 上海.

Fenghan Zhu. 2007. *Xī Zhōu Qīngtóngqì Fēnqī Duàndài Yánjiū* 西周青銅器分期斷代研究 *[Periodization and Dating of Western Zhou Bronze Vessels]*. Science Press 科學出版社, Beijing 北京.

A Appendix

A.1 Pretraining Configuration

DAPT was conducted over the full corpus of pre-Qin texts listed in Table 4. Training ran for one epoch, with the bottom six layers of the encoder frozen to reduce overfitting. TAPT was performed on the curated corpus of bronze inscriptions for three epochs. All training phases employed the AdamW optimizer with a learning rate of 5×10^{-5}, weight decay of 0.01, batch size of 8, and a masking probability of 15%. Experiments were run on two NVIDIA A100 GPUs.

A.2 Glyph Net and Realization Clusters

Figure 3 illustrates a subset of our diachronic glyph network. Each node corresponds to a character form g_i attested in the corpus, and edges denote either unidirectional normalization relations

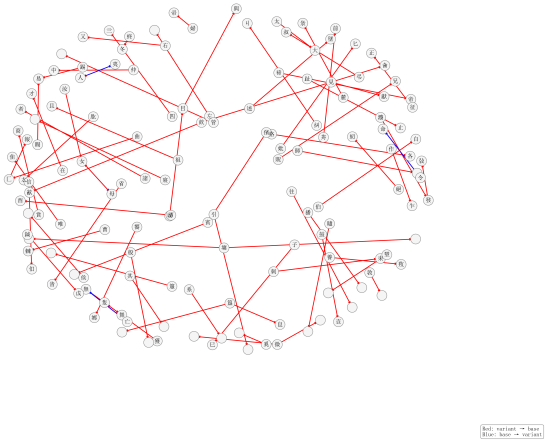

Figure 3: Excerpt from the glyph network showing variant relationships in the Shang period.

($g_i \to g_j$, red) or reversible substitution ($g_i \leftrightarrow g_j$, blue). These mappings form the structural basis for glyph-aware encoding and variant injection during training.

Bronze inscriptions frequently exhibit diachronic glyph variation. A semantic unit α may appear as distinct forms over time—for example, α_1 (Shang), α_2 (Western Zhou), and α_3 (Eastern Zhou)—linked by a chain:

$$\alpha_1 \to \alpha_2 \to \alpha_3$$

All such forms are grouped into a *realization cluster*:

$$\mathrm{RC}(\alpha) = \{\alpha_1, \alpha_2, \alpha_3\}$$

where $\mathrm{RC}(\alpha)$ denotes the set of all known surface realizations of α. During encoding, any $g \in \mathrm{RC}(\alpha)$ is mapped to a shared latent representation $[\alpha]$. At decoding, the model selects $g' \in \mathrm{RC}(\alpha)$ that maximizes contextual likelihood:

$$g' = \arg\max_{g \in \mathrm{RC}(\alpha)} P(g \mid \text{context})$$

This structure avoids the limitations of hard substitution (e.g., $g \mapsto g_{\text{modern}}$), which discards diachronic signal and precludes reverse mapping. Instead, GN maintains bijective paths between surface forms and semantic anchors, enabling: - Generalization: $g_{\text{unseen}} \in \mathrm{RC}(\alpha)$ can be semantically interpreted even if absent from training. - Alignment: tokens are grouped not by appearance, but by historical and semantic equivalence. - Reversibility: decoding remains faithful to inscriptional form, preserving chronological granularity.

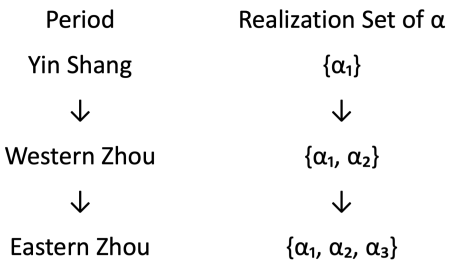

Figure 4: An illustrative example of diachronic glyph realization. The semantic unit α has a unique form α_1 in the Shang period, gradually expanding to $\{\alpha_1, \alpha_2\}$ in Western Zhou and $\{\alpha_1, \alpha_2, \alpha_3\}$ by the Eastern Zhou. This realization chain reflects both orthographic evolution and semantic continuity.

This formulation reflects both philological convention and neural modeling requirements, providing a scalable solution for lexical generalization across historical scripts.

A.3 Pre-Qin Corpus Composition

The DAPT corpus consists of 40 classical and excavated texts, covering all major schools of early Chinese thought. Table 4 provides a representative subset categorized by philosophical or historical affiliation. These texts provide broad coverage of the syntactic and lexical patterns most proximate to inscriptional Chinese.

A.4 Perplexity Evaluation

Table 5 and Table 6 report perplexity (PPL) scores for masked language modeling across mBERT and SIKU-BERT. Lower perplexity indicates better predictive confidence under the MLM objective.

For both encoders, DAPT and glyph-aware augmentation significantly reduce PPL relative to the baseline. Injecting variant forms (GN-inject) consistently outperforms replacement (GN-replace), confirming that preserving surface variation improves generalization. SIKU-BERT achieves lower PPL than mBERT across all configurations, consistent with its pretraining alignment with classical Chinese. The lowest overall PPL (.13.86) is obtained with SIKU-BERT under the DAPT + TAPT + GN-inject setting, demonstrating the combined benefit of domain adaptation and glyph-variant integration.

Category	Titles
Confucianism	*Analects*, *Mengzi*, *Liji*, *Xiao Jing*, *Xunzi*, *Yili*
Mohism	*Mozi*
Daoism	*Laozi*, *Zhuangzi*, *Liezi*, *He Guan Zi*, *Yu Liaozi*
Legalism	*Hanfeizi*, *Shang Jun Shu*, *Shenzi*, *Jian Zhu Ke Shu*, *Guanzi*
School of Names	*Gongsunlongzi*
School of the Military	*Sunzi Bingfa*, *Wu Zi*, *Liu Tao*, *Si Ma Fa*
Miscellaneous Schools	*Gui Gu Zi*, *Lü Shi Chun Qiu*
Histories	*Guo Yu*, *Yanzi Chun Qiu*, *Zhan Guo Ce*, *Mutianzi Zhuan*, *Zhushu Jinian*, *Zuo Zhuan*
Ancient Classics	*Book of Poetry*, *Shang Shu*, *Book of Changes*, *Rites of Zhou*, *Chu Ci*, *Shan Hai Jing*, *Yizhoushu*
Etymology/Medicine/Excavated	*Huangdi Neijing*, *Guodian*, *Mawangdui*

Table 4: Subset of pre-Qin texts included in the DAPT corpus.

Scenario	mBERT PPL	Std. Dev.
Baseline	169.39	± 28.72
TAPT + GN (inject)	16.14	± 2.00
TAPT + GN (replace)	18.27	± 1.84
DAPT + TAPT + GN (inject)	**15.98**	± 1.98
DAPT + TAPT + GN (replace)	17.48	± 1.69

Table 5: Perplexity results for mBERT under different training configurations.

Scenario	SIKU-BERT PPL	Std. Dev.
Baseline	1253.66	± 281.67
TAPT + GN (inject)	14.05	± 1.72
TAPT + GN (replace)	18.25	± 2.44
DAPT + TAPT + GN (inject)	**13.86**	± 1.78
DAPT + TAPT + GN (replace)	17.15	± 2.15

Table 6: Perplexity results for SIKU-BERT under different training configurations.