

# Finding structure in logographic writing with library learning II: Grapheme, sound, and meaning systematicity

**Guangyuan Jiang**

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

**Matthias Hofer**

MIT, Cambridge, Massachusetts, United States

**Jiayuan Mao**

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

**Lionel Wong**

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

**Joshua Tenenbaum**

MIT, Cambridge, Massachusetts, United States

**Roger Levy**

Massachusetts Institute of Technology, Cambridge, Massachusetts, United States

## Abstract

Writing systems are structured to depict the various facets of human language, from sounds to meanings. Chinese writing, as a logographic system, offers a distinctive opportunity to study the structural relationships between written forms and their sounds and meanings all at once. In this companion paper to Jiang et al. (2024), we explore a computational model based on library learning that can capture the compositional structure of Chinese characters and their relationship to sound and meaning. We extend the written-only library learning framework from Jiang et al. (2024) by incorporating written-sound joint compression and distributional semantic representations. The joint compression component allows the model to uncover structural relationships between a character's graphical components and its pronunciation, mirroring the function of phonetic and semantic radicals in Chinese orthography. With distributional semantics, the model also learns systematic links between the graphical structure and the meaning of characters, enabling it to predict the meanings of unseen characters based on their constituent parts. Moreover, our model allows us to explore historical shifts in how written Chinese has represented spoken language. We anticipate that our library learning model to be a unified computational account of writing's interaction with multi-level structures of human language. Full paper available at <https://jiang.gy/assets/pdf/jiang2025grapheme.pdf>