

# The Current State of the OBI DICT Project: A Bilingual e-Dictionary of Oracle-Bone Inscriptions with AI Image Recognition

Yang Jin, *Faculty of Humanities, Georg-August-Universität Göttingen, Germany* ([yangjin306@hotmail.com](mailto:yangjin306@hotmail.com)) (<https://orcid.org/0009-0000-7353-9849>)  
and

Shuo Wen, *MLBIO lab, Ecole Polytechnique Federale de Lausanne, Switzerland* ([shuo.wen@epfl.ch](mailto:shuo.wen@epfl.ch)) (<https://orcid.org/0009-0006-7923-7925>)

---

**Abstract:** This article reports on the current state of the OBI DICT project, a bilingual e-dictionary of oracle-bone inscriptions (OBI), incorporating artificial intelligence (AI) image recognition technology. It first provides a brief overview of the development of the lexicographical works on oracle bones. Subsequently, it identifies deficiencies in existing oracle-bone dictionaries and underscores the pressing demand for the compilation of a new dictionary. In the subsequent two sections, the article delineates the project's initiation and objectives and then outlines its design. The four principal phases of the project, that is, material collection, literature review, content and user interface design, and search engine and AI image recognition design, are described in detail in the third section. In the concluding section, it expounds on how the OBI DICT addresses users' search requirements and maximizes usability, thereby offering substantial support to contemporary oracle-bone research, streamlining the learning process for novices, and expanding the readership interested in oracle bones.

**Keywords:** ORACLE-BONE INSCRIPTIONS, ORACLE-BONE LEXICOGRAPHICAL WORKS, ORACLE-BONE DATABASES, BILINGUAL DICTIONARY, AI IMAGE RECOGNITION, MACHINE LEARNING, DICTIONARY COMPILATION

**Opsomming:** Die huidige stand van die OBI DICT-projek: 'n Tweetalige e-woordeboek van orakelbeeninskripsies wat gebruik maak van KI-beeldherkenning. In hierdie artikel word verslag gelewer oor die huidige stand van die OBI DICT-projek, 'n tweetalige e-woordeboek van orakelbeeninskripsies (OBI), wat kunsmatige intelligensie- (KI-) beeldherkenningstechnologie gebruik. Daar word eers 'n oorsig gegee van die ontwikkeling van die leksikografiese werke oor orakelbeendere. Daarna word leemtes in bestaande orakelbeenwoordeboeke geïdentifiseer en die dringende behoefte aan die samestelling van 'n nuwe woordeboek word beklemtoon. In die daaropvolgende twee afdelings word die ontstaan en doelwitte van die projek uiteengesit en daarna word die ontwerp daarvan beskryf. Die vier hoof fases van die projek, nl. materiaalversameling, literatuurbeskouing, inhouds- en koppelvlakontwerp, en soekenjin- en

KI-beeldherkenningsontwerp, word in die derde afdeling in besonderhede beskryf. In die slotafdeling word breedvoerig uiteengesit hoe die die soekvereistes van gebruikers in die OBI DICT aangespreek en die bruikbaarheid gemaksimaliseer word, om sodoende aansienlike steun aan kontemporêre orakelbeennavorsing te verleen, wat die leerproses vir leke vergemaklik, en die leserstal wat belangstel in orakelbeendere, uitbrei.

**Sleutelwoorde:** ORAKELBEENINSKRIPSIES, ORAKELBEEN- LEKSIKOGRAFIESE WERKE, ORAKELBEENDATABASISSE, TWEETALIGE WOORDEBOEK, KI-BEELDHERKENNING, MASJIENLEER, WOORDEBOEKSAMESTELLING

## 1. Introduction

The earliest unambiguously attested Chinese writing is the oracle-bone inscriptions (OBI), which refer to the texts inscribed on bones and shell in the Late Shāng (ca. 1300–1046 BC) and Western Zhōu (ca. 1046–770 BC) periods,<sup>1</sup> generally known as *jiǎgǔwén* 甲骨文 in Chinese. These texts were first excavated in 1899 in Anyang, China, and primarily constituted royal divinations. About twenty years after the discovery, the first lexicographical work on oracle bones appeared.<sup>2</sup> In the following one hundred years, oracle-bone materials were continuously unearthed, and more and more scholars joined the research, helping to achieve many significant accomplishments.<sup>3</sup> In this process, lexicographical works have also been continuously adapted to keep up with the development of the oracle-bone studies, and at the same time, in turn, have promoted the development of the discipline.

Since the discovery of oracle bones, there have been four kinds of oracle-bone (lexicographical) works, namely (1) collections of graphic forms, (2) concordances of inscriptions, (3) collections of lexical research, and (4) dictionaries. Collections of graphic forms focus on gathering various graphic forms of oracle-bone signs. Concordances compile inscriptions that feature the same oracle-bone sign. Collections of lexical research gather the perspectives of various scholars on the interpretation of oracle-bone signs. Dictionaries provide comprehensive analyses of the graphic forms and explain the usages of oracle-bone signs in the inscriptions. From the late 1980s, digital technology began to be integrated into oracle-bone research. The first online OBI database, the CHANT database (漢達文庫), was constructed in 1988 by the D.C. Lau Research Centre for Chinese Ancient Texts at the Institute of Chinese Studies, Chinese University of Hong Kong. Currently, the largest online OBI database is the *Yīnqì wényuān* 殷契文淵 (2016–), developed by Anyang Normal University, China. This database contains 154 collections of oracle-bone rubbings, including 239,736 images and 34,591 works on oracle-bone studies.<sup>4</sup>

With the assistance of digital technologies and online databases, the vast majority of research endeavours and projects on OBI have made significant progress. For instance, in the latest version of the *Xīn jiǎgǔwén biān* 新甲骨文編 (*New Collection of Oracle-bone Graphic Forms*) (Liú et al. 2014), digital technology

is employed to realistically present oracle-bone glyphic forms, which can facilitate researchers' understanding of the oracle-bone graphic forms and their development in different periods. Online databases can also assist researchers in various ways. For example, database searches can identify inscriptions containing a specific oracle-bone sign within a few seconds, which enables them to serve as an electronic concordance. By combining searches in various online databases, researchers can efficiently compile a comprehensive concordance for a specific oracle-bone sign.

However, the development of oracle-bone dictionaries is apparently lagging behind. Although lexicographical works appeared not long after the discovery of oracle bones, the first oracle-bone dictionary, the *Jiǎgǔwén zìdiǎn* 甲骨文字典 (*Dictionary of Oracle-Bone Inscriptions*) (Xú 1988), was compiled in the late 1970s and was not published until 1988. Due to the influence of Chinese epigraphy, like most lexicographical works of ancient Chinese, this dictionary is handwritten and in vertical format. A revised version of this dictionary was published in 2022, with no significant alterations to its content but marked improvements in formatting. That is, the revised version no longer has a vertical layout in handwriting, but a horizontal layout in typeface. The oracle-bone dictionaries published after Xú (1988) primarily include: the *Xīnbīān jiǎgǔwén zìdiǎn* 新編甲骨文字典 (*Newly Compiled Dictionary of Oracle-Bone Inscriptions*) (Liú 1993, 2005), the *Jiǎnmíng jiǎgǔwén cídiǎn* 簡明甲骨文詞典 (*Concise Dictionary of Oracle-Bone Inscriptions*) (Cūi 2001), the *Yīnxū jiǎgǔwén shíyòng zìdiǎn* 殷墟甲骨文實用字典 (*Practical Dictionary of Oracle-Bone Inscriptions of Yīn Ruins*) (Mǎ 2008), and the *Shíyòng jiǎgǔwén zìdiǎn* 實用甲骨文字典 (*Practical Dictionary of Oracle-Bone Inscriptions*) (Chén 2019).

Among these dictionaries, some have better typesetting than Xú (1988), such as Mǎ (2008) and Chén (2019), which employ horizontal printing typesetting instead of vertical handwritten typesetting. This format presents no reading challenge for lay users as horizontal formatting aligns more with contemporary reading habits, which greatly reduce user difficulties and broaden readership. On the other hand, some of the layouts are not as good as Xú's (1988). For example, in Liú (1993, 2005), all kinds of information are listed together without obvious distinctions in each entry, which increases the difficulty of reading. Moreover, most of these latter dictionaries are not at the same level as Xú (1988) in explaining the usage of oracle-bone signs in the inscriptions. Let's take *mù* 目 as an example. The graphic form of this sign is the depiction of an eye. It has four usages in oracle-bone inscriptions, which are all listed in Xú (1988: 361-362, 2022: 248-249): (1) eye (n.), (2) spy, monitor (v.), (3) the name of a person, and (4) the name of a state,<sup>5</sup> while the same entry in Mǎ (2008: 88) and Chén (2019: 212) include only two usages. Besides, one of the usages in Mǎ (2008: 88), explaining the *mù* 目 as a name of sacrificial ceremony in oracle-bone inscriptions, is incorrect. Therefore, even though it has been published for more than thirty years, Xú (1988) is still the most widely used and authoritative dictionary in oracle-bone research.

However, the usages of oracle-bone signs in Xú (1988) reflect research from the late 1980s, failing to incorporate the research advancements made afterwards in the latest edition (Xú 2022). This inadequacy cannot satisfy the requirements of contemporary research. Therefore, the compilation of a better oracle-bone dictionary becomes an imperative necessity. Nowadays, the emergence of online oracle-bone databases and the application of interdisciplinary research approaches, such as the artificial intelligence (AI) image recognition technology combined with oracle-bone research method to recognize the oracle-bone signs (e.g. Huang et al. 2019, Liú 2020, Mén and Zhāng 2021, Jin 2023) or rejoin oracle-bone fragment images (e.g. Zhang et al. 2022), have brought new opportunities for dictionary compilation and oracle-bone research. If these new technologies and interdisciplinary research methods can be effectively utilized in the compilation of oracle-bone dictionaries, the newly compiled dictionary will undoubtedly promote the development of oracle-bone research and expand the readership of oracle-bone inscriptions.

## 2. Initiation and Aim of OBI DICT

The OBI DICT originated as part of a doctoral project at the Georg-August-Universität Göttingen, Germany (cf. Jin 2024). During the initial research phase of the project, it became evident that the development of oracle-bone dictionaries seriously lags behind oracle-bone research. Furthermore, it was noted that no English dictionary on oracle bones exists and other English reference works do not necessarily provide enough information on oracle-bone inscriptions. For example, the only comprehensive OBI introduction in English is Keightley's monograph published in 1978, which is now seen as outdated. English OBI readings published in recent years, such as Chen et al. (2017) and Takashima (2019), require high level of Chinese proficiency, which is not suitable or accessible for lay users. Based on the lack of user-friendly OBI dictionaries, we felt that there was a need for an accessible, modern bilingual oracle-bone dictionary. After investigating all previously mentioned oracle-bone lexicographical works and evaluating the feasibility of employing AI image recognition to the dictionary, the project started at Georg-August-Universität Göttingen at the end of 2022 (see also Jin and Wen 2024). Dr Yang Jin is the designer and coordinator of this project and Prof. Gordon Whittaker serves as the language consultant and academic advisor. Shuo Wen from the Ecole Polytechnique Federale de Lausanne provides AI technology support for the project. The project is expected to be completed within five years.

The digitalization of oracle bones, such as the construction of online databases, has rendered it more accessible to researchers and readers globally. Nonetheless, the effective utilization of these digital resources necessitates a high level of proficiency in the Chinese language, thus presenting a great challenge for both the general public with an interest in oracle bones and students in the early stages of their study worldwide. In addition to making up for the

dictionary's shortcomings by adding the latest research results, the OBI DICT also intends to improve the dictionary's usability and expand its readership. In light of these challenges and opportunities, this project aims to develop a bilingual oracle-bone dictionary in the form of an electronic application using an interdisciplinary approach, including AI image recognition.

### 3. Design of the OBI DICT

The OBI DICT project consists of four main phases: material collection, literature review, content and user interface design, and search engine and AI image recognition design. The project has gone through the initial stage, that is, material collection and content design, and is now focusing on literature review and AI image recognition design.

#### 3.1 Collection

The project has started with a comprehensive collection of academic works on oracle-bone inscriptions since late 1980s, including Chinese and Western sources. With regard to Chinese works, the collection is primarily built on the collections of oracle-bone lexical research as well as online databases and platforms. Collections of lexical research primary include the (1) *Jiǎgǔ wénzì gǔlín* 甲骨文字詁林 (*Collection of Explanations on Oracle-Bone Inscriptions*) (Yú 1996), (2) *Bǎinián jiǎgǔxué lùnzhùnmù* 百年甲骨學論著目 (*Bibliography of Oracle-Bones over the Past Century*) (Sòng and Cháng 1999), (3) *Jiǎgǔ wénxiàn jíchéng* 甲骨文獻集成 (*Collection of Oracle-Bone Literature*) (Sòng and Duàn 2001), and (4) *Jiǎgǔ wénzì gǔlín bǔbiān* 甲骨文字詁林補編 (*Supplement of the Collection of Explanations on Oracle-Bone Inscriptions*) (Hé 2017). Yú (1996) is a collection of lexical research containing diverse scholarly perspectives pertaining to the interpretation of oracle-bone signs, extending from the discovery of oracle bones down to 1989. Sòng and Cháng (1999) is a catalog, which includes research works in China and abroad from the discovery of oracle bones in 1899 to June 1999. Sòng and Duàn (2001) is a collection of research works in China and overseas, covering the period from 1899 to 1999. Hé (2017) serves as a supplement to Yú (1996), which gathers the perspectives of various scholars from 1990 to 2013 regarding the interpretation of oracle-bone signs. Moreover, the latest research works primarily come from the continuously updated online oracle-bone databases, such as the *Yīnqì wényuān* 殷契文淵, and online platforms, such as the CNKI (中國知網). For example, the database *Yīnqì wényuān* 殷契文淵 contains 34,591 oracle-bone research works, which can be searched by the works' titles, authors, keywords, abstracts, sources, and full texts.

As for overseas research works, in addition to Sòng and Cháng (1999) as well as Sòng and Duàn (2001) mentioned above, overseas research after the late 1980s primarily come from: (1) the *Xīguānhànjì: Xīfāng hànxié chūtǔ wénxiàn yánjiū*

gàiyào 西觀漢記: 西方漢學出土文獻研究概要 (*Chinese Annals in the Western Observatory: An Outline of Western Sinology's Contributions to the Study of Chinese Unearthed Texts*) (Shaughnessy 2018) and (2) the annual bibliography of the journal *Early China*. Shaughnessy (2018) is an historical introduction of Western research on Chinese unearthed materials, such as oracle-bone, bronze inscriptions, and bamboo and silk manuscripts. In the section of oracle bones, it has a list of Western research works from 1911 to 2015 (Shaughnessy 2018: 135-198). *Early China* is an annual journal, dedicated to original research covering every facet of China's culture and civilization, from ancient times to the Han dynasty (ca. AD 220). At the end of each volume, there is an annual bibliography, which is a list of English research works on early China for the whole year. Other research works on oracle-bones have also been gathered from online platforms such as Google scholar, ResearchGate and Academia.

### 3.2 Review

The collected works involve the interpretation of approx. 4,500 oracle-bone signs. Firstly, based on the (1) oracle-bone concordances, such as the *Yīnxū jiǎgǔ kècí lèi zuǎn* 殷墟甲骨刻辭類纂 (*Classified Compilation of Oracle-Bone Inscription from Yin Ruins*) (Yáo 1989) and the *Yīnxū jiǎgǔ wéncí lèibīān* 殷墟甲骨文辭類編 (*Classification and Compilation of Oracle-Bone Inscriptions in Yin Ruins*) (Chén 2021), and (2) online databases, such as the CHANT and *Yīnqì wényuān* 殷契文淵, the inscriptions for each individual oracle-bone sign are summarized. The collected research works have been studied, and the usage of each sign discussed in these works is summarized. In the next step, in order to determine the readable oracle-bone signs (approx. 2200) and their usage, the inscriptions are carefully analyzed, and the analysis is combined with the collected research works and the explanations provided in the dictionaries, such as Xú (1988, 2022), Cuī (2001), Liú (2005), Mǎ (2008), and Chén (2019). Then, appropriate inscriptional examples are selected for each usage. Finally, representative graphic forms for each sign are chosen based on previous collections of graphic forms, such as Liú et al. (2014) and Chén (2021), and the online database *Yīnqì wényuān* 殷契文淵.

### 3.3 Design

#### 3.3.1 Content

The OBI DICT contains approx. 2200 entries. Each entry includes, as illustrated by the entry *mù* 目 (see figure 1), the following elements:

- (1) A head sign in Modern Chinese character with Pinyin (MCP):<sup>6</sup> e.g. 目 *mù*
- (2) Old Chinese reconstructions (OC), where possible and available: e.g. \*C.m(r)[u]k















<b>MCP</b>		<i>mù</i>	<b>OC</b>	*C.m(r)[u]k	
<b>OBI</b>	 師 H 20173	 子 H 21740	 賓 H 6946f	 賓 H 13620f	
	 賓 H 14787f	 歷 H 33237	 歷 H 34272	 無名 H 28374	
<b>AG</b>	Depiction of an eye.				
<b>DICTIONARY</b>	<b>NOUN</b>				
	1 <i>mù</i> 目, <i>yǎnjīng</i> 眼睛 'eye'				
	 貞: 王其疾(疒)目。(H 456 f, 賓, Period I) [The diviner] divined: Will the king have ailing eye(s)?				
	2 <i>rénmíng</i> 人名, <i>fāngguó míng</i> 方國名 'the name of a person or state'				
	 庚午卜, 賓貞: 子目媿, 嘉。(H 14034 f, 賓, Period I) Crack-making on the day <i>Gēngwǔ</i> 庚午 (the 7 <sup>th</sup> in the <i>gānzhī</i> cycle), Bīn 賓 (the diviner) divined: Would it be good that Zǐ Mù 子目 gives birth to a child?				
	<b>VERB</b>				
	<i>zhēnchá</i> 偵查, <i>xúnshì</i> 巡視 'spy, monitor'				
	 貞: 呼目舌方。(H 6194, 賓, Period I) [The diviner] divined: [Should the king] summon [someone] to monitor Gōngfāng 舌方 (one of the hostile countries of late Period I, to the west of the Shāng state)?				
<b>CW</b>	<a href="#">目方 <i>mùfāng</i></a>				
<b>FR</b>	Yú 1996: 0601; Cui 2001: 153; Xú 2014: 361-362, 2022: 248-249; Guō and Qiū 2021; Jīn 2024: 248-251.				

Figure 1: The entry *mù* 目 in the OBI DICTIONARY

Pinyin is the modern pronunciation of this Chinese character, and the Old Chinese reconstruction is the pronunciation of the Shāng and Western Zhōu periods, which primarily follows the reconstructions of the Gassmann and Behr (2011), Baxter and Sagart (2014, 2020) and Zhèng (2018).

- (3) Oracle-bone signs displayed in variant graphic forms with citations for their provenance from oracle-bone collections and diviner groups (OBI):<sup>7</sup>  
e.g.  (H 20173, 師 group),  (H 14787f, 賓 group)

One of the distinctive features of oracle-bone inscriptions, owing to its early developmental stage, lies in the variability of its graphic forms. In other words, a single oracle-bone sign may be represented in divergent graphic forms within the same historical period or in different periods. Therefore, it is imperative to compile a comprehensive list of these varied graphic forms, which aids readers in acquiring a thorough comprehension of the sign's graphic forms.

- (4) An analysis of graphic forms (AG), as needed:  
e.g. The graphic form of *mù* 目 is a depiction of an eye.

Oracle-bone inscriptions are comprised of numerous logograms, which serve as representations of lexical morphemes without explicit indication of word pronunciation. These logograms fall into two main categories: (1) those depicting objects or object parts, and (2) those depicting attributes, states, or actions.<sup>8</sup> Given this logographic nature, the analysis of graphic forms should be an essential part of oracle-bone dictionaries.

- (5) The usage of the oracle-bone sign in inscriptions, along with English translations (DICT):

e.g. *mù* 目 is used as both a noun and a verb in inscriptions:  
1. (n.) (1) *mù* 目, *yǎnjīng* 眼睛 'eye'; (2) *rénmíng* 人名; *fāngguó míng* 方国名 'the name of a person or state'.  
2. (v.) *zhēnchá* 侦查, *xúnrì* 巡视 'spy, monitor'.

- (6) Illustrative examples of usage with provenance from oracle-bone collections and diviner groups. These examples consist of oracle-bone transcriptions, transliterations in traditional characters, and English translations:

e.g. One of the inscriptional examples for the usage (n.) (1) is:  
真王其疒目  
真: 王其疾(疒)目。(H 456 f, 賓, Period I)  
[The diviner] divined: Will the king have ailing eye(s)?

The inclusion of inscriptional examples is imperative for an oracle-bone dictionary, and providing citations that indicate the provenance of the examples is



also of great importance. Such indications enable readers to revisit these inscriptions and form their independent assessments regarding the sign usage.

- (7) References to related compound words (CW), including links to their usage:

e.g. The compound word related to *mù* 目 is *mùfāng* 目方.

- (8) Additional readings related to the specific oracle-bone sign (FR):

e.g. For more discussions on *mù* 目 see also Yú (1996: 0601), Cūi (2001: 153), Xú (1988: 361-362, 2022: 248-249), and Guō and Qiū (2021), Jin (2024: 248-251).

Similar to the indication of the examples' provenance, the additional readings related to the specific oracle-bone sign help readers establish a comprehensive understanding of the various discussions of the usage of the specific sign and form their independent views.

### 3.3.2 User interface

The oracle-bone signs are searchable by three modes in the OBI DICT: (1) Modern Chinese (both traditional and simplified), Pinyin or English through the search engine, (2) image (allowing users to upload pictures or take photos), and (3) handwriting input. These modes are designed to meet user search needs and improve usability to the largest extent (see figure 2).

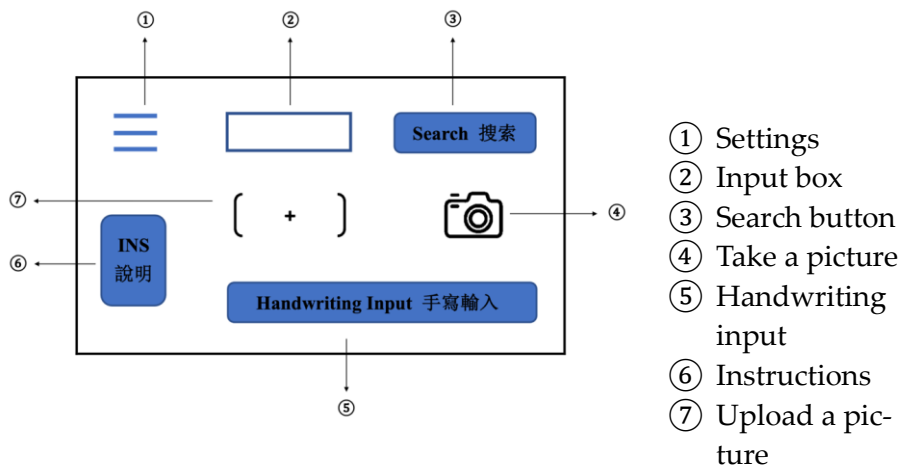


Figure 2: The user interface of the OBI DICT

Though professional oracle-bone researchers and users proficient in Chinese may comfortably employ Chinese characters or Pinyin for searching, it is also essential to meet needs of lay users and individuals less proficient in Chinese. The incorporation of AI image recognition (uploading pictures, taking photos and handwriting input) will significantly ease the retrieval and reduce the challenges associated with dictionary utilization for those not proficient in Chinese, greatly broadening the user base.

### 3.3.3 Maintenance and updates

One of the primary issues inherent in extant oracle-bone dictionaries, whether in print or digital form, is their failure to integrate the recent research findings. Therefore, the OBI DICT not only assimilates contemporary research results during its compilation but also commits to ongoing maintenance and updates post-completion, thereby aligning itself with the latest advancements in research. This practice guarantees the dictionary's status as a dynamic and up-to-date resource for users.

### 3.4 Methodology

There are mainly two parts in the method applied: (1) the image recognition model and (2) an OBI search engine. After the handwriting or photograph of an oracle-bone sign is input into the OBI DICT, it will be first recognized as a specific OBI sign by the image recognition model, and then the search engine will search and output the explanation of the sign (see figure 3).

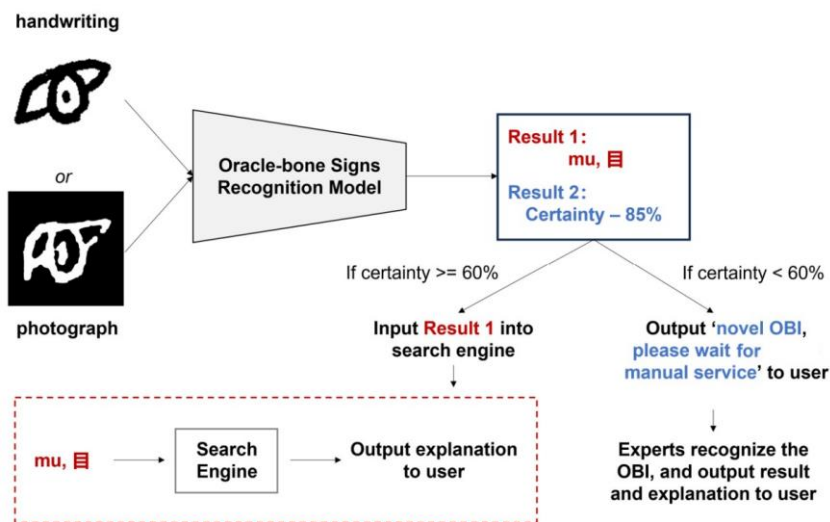


Figure 3: Recognition model and search engine of the OBI DICT

### 3.4.1 Image recognition

Machine learning methods are employed to identify oracle-bone signs in images. Specifically, identifying oracle-bone signs is considered as a classification task. In the first place, 2,200 predefined oracle-bone signs are used as 2,200 classes. Then, if input images contain oracle-bone signs, the model will output predefined classes representing the corresponding oracle-bone signs. Supervised learning is utilized to develop the classification model.

To train the model with supervised learning, a dataset must be created. Since the aim is to employ supervised training, both images and labels are required. To enable the model to recognize both photograph and handwriting of the OBI signs, images of both formats are collected. In short, the dataset consists of printed and handwritten OBI images accompanied by their respective labels. Let's again take the entry *mù* 目 as an example. We first collect 100 images of both photograph and handwritten format and resize them into 256 × 256 resolution. Then, 70, 20, and 10 images are put into training, validation, and test sets. The training sets are used for model training; the validation sets aid in hyperparameter selection, and the test sets assess the model's performance.

Our primary challenge lies in the scarcity of data available for robust model training. To address this issue, transfer learning should be employed. Broadly, transfer learning entails initializing the network using a large dataset that lacks oracle-bone signs, before fine-tuning the network using our OBI dataset. Another challenge is that the user can input rare OBIs which do not belong to the 2,200 predefined oracle-bone signs. In this case, out-of-distribution detection will be needed. Specifically, the model outputs a certainty for each input image. If the certainty is lower than a threshold, the model will output 'unknown OBI'. In this situation, if the user is willing to get further information, he or she can send a request and wait for assistance from a designated expert (manual service). The manual service can inform the user of the correct result or the information that this sign has not been deciphered yet.

### 3.4.2 Search engine

After identifying oracle-bone signs in images, results and explanations should be provided to the users. Thus, a search engine is needed to search the results and explanations according to the oracle-bone signs. The method for implementing the search engine is as follows: The dataset for the OBI DICT entries is in the first place constructed, including approx. 2200 recognized oracle-bone signs. Following that, a structured database is established using MySQL or PostgreSQL to store these entries. The database is then populated with entries from the DICT dataset, ensuring that each entry is inclusive of Chinese (both simplified and traditional), Pinyin, and English as searchable fields. Subsequently, the search logic on the backend is implemented, so that when a Modern Chinese character, Pinyin, or English input is provided by a user, the database is que-

ried, and the corresponding entry is retrieved by the system. For example, to search for the oracle-bone sign corresponding to the word "eye", the user can enter the Modern Chinese character "目", Pinyin "mu" or English word "eye" in the search engine, and the OBI DICT will query the database and retrieve the corresponding entry *mù* 目 (see figure 1).

#### 4. Conclusion

As discussed in section 1, the primary issue inherent in extant oracle-bone dictionaries is their failure to incorporate the latest research findings. Apart from that, there are some other shortcomings. For example, traditional handwritten and vertical formats present a tough reading challenge for lay users or non-native Chinese users. Moreover, the usages included in each entry are not comprehensive, which fails to enable users to understand the meaning of the same sign in different inscriptions. Furthermore, some dictionaries, such as Xú (1988, 2022) and Liú (1993, 2005), only have the lookup tables of the radicals and stroke counts,<sup>9</sup> which are usable for experts and users proficient in Chinese but are not friendly to lay persons and individuals less proficient in Chinese. Likewise, in most dictionaries, the pronunciation is not indicated, making it inconvenient to use for non-native Chinese users. The ongoing project OBI DICT has the potential to address all these problems. In the first place, it integrates the recent research findings and provides comprehensive usages for each entry and will also conduct an ongoing maintenance and updates to align itself with the latest advancements in research and provide up-to-date resource for users. Moreover, the OBI DICT employs horizontal typesetting instead of the traditional vertical handwritten format, which, as noted, aligns more with contemporary reading habits. Furthermore, the oracle-bone signs are searchable by three modes in the OBI DICT: Modern Chinese (both traditional and simplified), Pinyin or English through the search engine, image (allowing users to upload pictures or take photos) and handwriting input. The incorporation of AI image recognition or handwriting input will significantly assist the retrieval of data and reduce the challenges associated with dictionary utilization for those not proficient in Chinese. In addition, pronunciation for each entry, both Pinyin and Old Chinese reconstruction, Pinyin and English for the explanations of usages and English translations for illustrative examples are provided in the OBI DICT, making it user-friendly to a wider range of users. Therefore, the ongoing project OBI DICT aims to fully meet users' search needs and improve user usability to the largest extent. It provides support to current oracle-bone research, facilitates the learning process for beginners, and broadens the readership for oracle bones, including those whose first language is not Chinese.

The next stage of the OBI DICT project will focus on issues related to AI image recognition, such as improving the accuracy of image recognition, the recognition of undeciphered oracle-bone signs, and data privacy. In the maintenance phase, as mentioned in Section 3, the latest research on the approx. 2200

recognized oracle-bone signs will be continuously updated, and important discussions on approx. 2300 unrecognized oracle-bone signs will also be gradually added to the OBI DICT to facilitate users' research and learning. It would also be ideal if the OBI DICT could be a formal part of existing oracle-bone databases or platforms in the future. If the OBI DICT can be combined with these databases, it will greatly reduce learning difficulty, improve research efficiency, and promote the development of the discipline.

## Endnotes

1. The chronology of the Late Shāng and Western Zhōu periods follows the latest research results of the Xia-Shang-Zhou Chronology Project (夏商周斷代工程) in China (Xià Shāng Zhōu duàndài gōngchéng zhuānjiǎzǔ 2022).
2. The earliest lexicographical work on oracle bones is the *Fǔshì Yīnqì lèizǔǎn* 籒室殷契類纂 (*Collections of Graphic Forms of Yin Inscriptions in Fǔshì*), edited by Xiāng Wáng 王襄 and published in 1920.
3. For the development of academic research of oracle-bone inscriptions see e.g. Wáng and Yáng (1999), Wáng and Koo (2019) and Jin (2024: 68-74).
4. The database is still being continuously updated, and these are the statistics as of May 19, 2024.
5. For recent discussions on the oracle-bone sign mù 目 see Jin (2024: 248-251).
6. "Signs" refer to graphic forms representing Chinese before the Han Dynasty, when their graphic forms were not fully standardized, while "characters" refer to graphic forms representing Chinese after the Han Dynasty, when their graphic forms were standardized.
7. "H" is the abbreviation for the *Jiǎgǔwén héjí* 甲骨文合集 (*Collection of Oracle-Bone Inscriptions*) (Guō and Hú 1978–1982), the largest collection of OBI rubbings.
8. For recent discussions on the logograms in early Chinese writing see Jin (2024: 233-255).
9. Radicals, known as *bùshǒu* 部首 in Chinese, refer a component or a character conveying the lexical meaning of a logogram in Chinese, and Chinese dictionaries arrange characters under radicals.

## References

### Dictionaries

- Chén, N. [陳年福] (Ed.). 2019. *Shíyòng jiǎgǔwén zìdiǎn* [實用甲骨文字典] (*Practical Dictionary of Oracle-Bone Inscriptions*). Chengdu: Sichuan cishu chubanshe [四川辭書出版社].
- Cuī, H. [崔恒昇] (Ed.). 2001. *Jiǎnmíng jiǎgǔwén cídiǎn* [簡明甲骨文字典] (*Concise Dictionary of Oracle-Bone Inscriptions*). Hefei: Ānhuī jiàoyù chūbǎnshè [安徽教育出版社].
- Liú, X. [劉興隆]. 1993. *Xīnbiān jiǎgǔwén zìdiǎn* [新編甲骨文字典] (*Newly Compiled Dictionary of Oracle-Bone Inscriptions*). Beijing: Guójì wénhuà chūbǎn gōngsī [國際文化出版公司].
- Liú, X. [劉興隆]. 2005. *Xīnbiān jiǎgǔwén zìdiǎn* [新編甲骨文字典] (*Newly Compiled Dictionary of Oracle-Bone Inscriptions*). Revised Edition. Beijing: Guójì wénhuà chūbǎn gōngsī [國際文化出版公司].

- Mǎ, R.** [馬如森]. 2008. *Yīnxū jiǎgǔwén shíyòng zìdiǎn* [殷墟甲骨文實用字典] (*Practical Dictionary of Oracle-Bone Inscriptions of Yīn Ruǐns*). Shanghai: Shànghǎi dàxué chūbǎnshè [上海大學出版社].
- Xú, Z.** [徐中舒] (Ed.). 1988. *Jiǎgǔwén zìdiǎn* [甲骨文字典] (*Dictionary of Oracle-Bone Inscriptions*). Chengdu: Sìchuān císhū chūbǎnshè [四川辭書出版社].
- Xú, Z.** [徐中舒] (Ed.). 2022. *Jiǎgǔwén zìdiǎn* [甲骨文字典] (*Dictionary of Oracle-Bone Inscriptions*). Horizontal Layout Edition. Chengdu: Sìchuān císhū chūbǎnshè [四川辭書出版社].

### Other lexicographical works

- Chén, N.** [陳年福] (Ed.). 2021. *Yīnxū jiǎgǔ wéncí lèibiān* [殷墟甲骨文辭類編] (*Classification and Compilation of Oracle-Bone Inscriptions of Yīn Ruǐns*). Chengdu: Sìchuān císhū chūbǎnshè [四川辭書出版社].
- Hé, J.** [何景成] (Ed.). 2017. *Jiǎgǔ wénzì gǔlín bǔbiān* [甲骨文字詁林補編] (*Supplement of the Collection of Explanations on Oracle-Bone Inscriptions*). Beijing: Zhōnghuá shūjú [中華書局].
- Liú, Z.** [劉釗], **Y. Hóng** [洪飈], **X. Zhāng** [張新俊] and **Z. Zhōu** [周忠兵] (Eds.). 2014. *Xīn jiǎgǔwén biān* [新甲骨文編] (*New Collection of Oracle-bone Graphic Forms*). Revised Edition. Fuzhou: Fújiàn rénmín chūbǎnshè [福建人民出版社].
- Wáng, X.** [王襄]. 1920. *Fǔshì Yīnqì lèizuǎn* [簠室殷契類纂] (*Collections of Graphic Forms of Yīn Inscriptions in Fǔshì*). Tianjin: Tiānjīn bówùguǎn [天津博物館].
- Yáo, X.** [姚孝遂] (Ed.). 1989. *Yīnxū jiǎgǔ kècí lèi zuǎn* [殷墟甲骨刻辭類纂] (*Classified Compilation of Oracle-Bone Inscription from Yīn Ruǐns*). Beijing: Zhōnghuá shūjú [中華書局].
- Yú, X.** [于省吾] (Ed.). 1996. *Jiǎgǔ wénzì gǔlín* [甲骨文字詁林] (*Collection of Explanations on Oracle-Bone Inscriptions*). Beijing: Zhōnghuá shūjú [中華書局].

### Online databases

- The CHANT Database (漢達文庫): <http://www.chant.org/>  
The Yīnqì wényuān 殷契文淵 database: <http://jgw.aynu.edu.cn/>

### Other literature

- Baxter, W.H. and L. Sagart.** 2014. *Old Chinese: A New Reconstruction*. New York: Oxford University Press.  
Online supplementary materials updated on October 21, 2020:  
<http://ocbaxtersagart.lsa.umich.edu/>
- Chen, K., Z. Song, Y. Liu and M. Anderson (Eds.).** 2017. *Reading of Shāng Inscriptions* [商代甲骨中英讀本]. Shanghai: Shanghai People's Publishing House.
- Gassmann, R.H. and W. Behr.** 2011. *Antikchinesisch — Ein Lehrbuch in zwei Teilen*. Bern: Peter Lang AG.
- Guō, J.** [郭靜云] and **S. Qiū** [邱詩瑩]. 2021. *Jiǎgǔwén zhōng yǐ tāotiè yǎnjīng wéi "mù" "chén" de zìxíng kǎo* [甲骨文中以饜饕眼睛為"目""臣"的字形考] (A Study on the Graphic Forms of the Sign "Eyes" and "Officials" Based on Gluttons' Eyes in Oracle-Bone Inscriptions). *Jiǎgǔwén yǔ yīnshāngshǐ* [甲骨文與殷商史] 11: 280-291.
- Guō, M.** [郭沫若] and **H. Hú** [胡厚宣] (Eds.). 1978–1982. *Jiǎgǔwén héjí* [甲骨文集] (*Collection of Oracle-Bone Inscriptions*). Beijing: Zhōnghuá shūjú [中華書局].

- Huang, S., H. Wang, Y. Liu, X. Shi and L. Jin.** 2019. OBC306: A Large-scale Oracle Bone Character Recognition Dataset. *2019 International Conference on Document Analysis and Recognition, ICDAR 2019, Sydney, Australia, 20–25 September 2019*: 681-688.
- Jin, Y.** 2023. The Potential Benefits and Limitations of Artificial Intelligence Technology Used in Oracle-Bone Studies. *Irish Journal of Technology Enhanced Learning* 7(2): 8-20.
- Jin, Y.** 2024. *A Comparative Study of the Origins of Chinese and Mesoamerican Writing*. Ph.D. Dissertation. Göttingen: Georg-August-Universität Göttingen.  
<http://dx.doi.org/10.53846/goediss-10371>
- Jin, Y. and S. Wen.** 2024. A Critical Analysis of OBI Lexicography and Online Databases, and a Brief Introduction to an Ongoing OBI e-Dictionary Project. (to appear)
- Keightley, D.N.** 1978. *Sources of Shang History: The Oracle-Bone Inscriptions of Bronze Age China*. Berkeley: University of California Press.
- Liú, G.** [刘国英]. 2020. Jiyú shēndù xuéxí de jiǎgǔ wénzì jiǎncè yǔ shíbié [基于深度学习的甲骨文文字检测与识别] (Oracle-Bone Inscriptions Detection and Recognition Based on Deep Learning). *Yīndù xuékān* [殷都学刊] 3: 54-59.
- Mén, Y.** [门艺] and **C. Zhāng** [张重生]. 2021. Jiyú réngōng zhīnéng de jiǎgǔwén shíbié jìshù yǔ zìxíng shùjùkù gòujiàn [基于人工智能的甲骨文识别技术与字形数据库构建] (Artificial Intelligence Based Oracle-Bone Inscriptions Recognition Technology and Graphic Forms Database Construction). *Zhōngguó wénzì yánjiū* [中国文字研究] 1: 9-16.
- Shaughnessy, E.L.** 2018. *Xīfāng hànxié chūtǔ wénxiàn yánjiū gàiyào* [西觀漢記: 西方漢學出土文獻研究概要] (*Chinese Annals in the Western Observatory: An Outline of Western Sinology's Contributions to the Study of Chinese Unearthed Texts*). Shanghai: Shànghǎi gǔjí chūbǎnshè [上海古籍出版社].
- Sòng, Z.** [宋鎮豪] and **Y. Cháng** [常耀華] (Eds.). 1999. *Bǎinián jiǎgǔxué lùnzhù mù* [百年甲骨學論著目] (*Bibliography of Oracle-Bones over the Past Century*). Beijing: Yǔwén chūbǎnshè [語文出版社].
- Sòng, Z.** [宋鎮豪] and **Z. Duàn** [段志洪] (Eds.). 2001. *Jiǎgǔ wénxiàn jíchéng* [甲骨文獻集成] (*Collection of Oracle-Bone Literature*). Chengdu: Sīchuān dàxué chūbǎnshè [四川大學出版社].
- Takashima, K.** 2019. *A Little Primer of Chinese Oracle-Bone Inscriptions with Some Exercises*. Second revised edition. Wiesbaden: Harrassowitz.
- Wáng, Y.** [王宇信] and **S. Yáng** [楊昇南]. 1999. *Jiǎgǔxué yībǎinián* [甲骨學一百年] (*One Hundred Years of Oracle-Bone Studies*). Beijing: Shèhuì kēxué wénxiàn chūbǎnshè [社會科學文獻出版社].
- Wáng, Y.** [王宇信] and **Y.H. Koo** [具隆會]. 2019. *Jiǎgǔxuéfāzhǎn 120 nián* [甲骨學發展120年]. (*120-Years' Development of Oracle-Bone Studies*). Beijing: Zhōngguó shèhuì kēxué chūbǎnshè [中國社會科學出版社].
- Zhang, Z., A. Guo and B. Li.** 2022. Internal Similarity Network for Rejoining Oracle Bone Fragment Images. *Symmetry* 14(7): 1464.
- Zhèng, Z.** [鄭張尚芳]. 2018. *Shàngǔ yīnxì* [上古音係] (*Ancient Phonological System*). Shanghai: Shànghǎi jiàoyù chūbǎnshè [上海教育出版社].