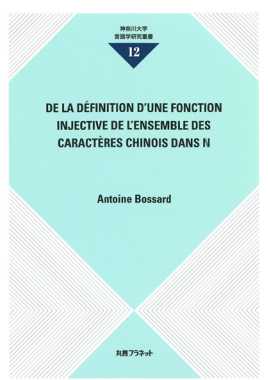


On the publication of a book related to the definition of an injective function from the set of Chinese characters to N.

Antoine Bossard



Those we name Chinese characters designate the logograms on which rely, even partially, for example, the Chinese, Japanese, Korean and Annamese writing systems. As such, it is no overstatement to say that these characters cement

Asian cultures. More precisely, and in addition, obviously, to the Chinese writing system, the Chinese characters make up the vast majority of the symbols involved in the Japanese writing system. That is, they are not alone: the Chinese characters in Japanese are very frequently accompanied by *hiragana* and *katakana* syllables. They are called *kanji* in this language. The Korean writing system is in a similar situation, although the Chinese characters therein are a minority; they are called *hanja* in this other language, where they accompany *hangul* symbols. In addition, the writing system based on Chinese characters used in Annamese is called the *chữ nôm* script.

From a technical point of view, the processing of Chinese characters on computer systems is a long-standing issue. One reason is the large number of symbols involved: tens of thousands. Another one is the absence of an exhaustive definition of these characters. Conventional approaches have thus had to make compromises: local encodings such as JIS X 0208 have considered a restricted set of characters, while

others such as Unicode have tried to include a larger set of characters. Nonetheless, some characters, albeit infrequently used ones, remain excluded from these standards and cannot thus be easily processed, let alone represented, by computer systems. Besides, the mapping of characters to code points is another issue: to this end, those conventional approaches have mostly relied on the character radical property and the stroke number property. Although convenient, ambiguity remains and makes using such codes, for instance for character lookup, difficult.

Considering the aforementioned two issues, namely a lacking character coverage and character lookup ambiguity, our research aims at unambiguously defining a unique “signature” for each and every Chinese character, that is, theoretically supporting any character, even those not currently supported and possibly those to be henceforth discovered. Defining a signature here simply means assigning a unique identifier to every character, which is somewhat related to the Four-corner character lookup method. This is a brief, informal presentation of our objective.

We believe that there is no need to further emphasise the important applications of this research: universal coverage and processing of all the Chinese characters, including variants (e.g. demotic forms) and local characters (e.g. Japanese *kokuji*) are essential in preserving the cultural heritage of Asian cultures. Aiming at describing a practical solution to the Chinese character mapping problem, this book provides a review of essential properties of the characters and may thus also be of interest to readers wishing to deepen their knowledge of the Chinese characters.

神奈川大学言語学研究叢書 第12巻 Antoine Bossard (丸善プラネット, 2023年)

『DE LA DÉFINITION D'UNE FONCTION INJECTIVE DE L'ENSEMBLE DES CARACTÈRES CHINOIS DANS N』