

# A SELF-LEARNING SYSTEM FOR THE CHINESE CHARACTERS

Georges FAFIOTTE  
GETA, IMAG (UJF & CNRS)  
BP 53X, 38041 Grenoble Cedex, France

## Abstract

We are prototyping a system for the self-learning of Chinese characters, presently on a Macintosh computer. The interactive information base provides the learner with basic universal properties of the characters (morphology, intrinsic meaning), extended with a quite comprehensive set of language-dependent aspects (phonetics, extended semantics, contextual or pragmatic attributes). The user is intended to have a professional or cultural non-academic motivation. The system allows to experiment on Heisig's proposal involving the separation of Chinese characters learning (or Japanese kanji) from that of the language. A prototype under HyperCard may be demonstrated on a subset of about 200 characters.

## Keywords

Chinese characters, Kanji, interactive information base, computer aided learning, personalized autonomous acquisition of Chinese characters.

## 1. Project motivation

The aim of the project is twofold: the first part is to model an interactive information base on characteristic properties of Chinese characters (or Japanese kanji), which would lend itself to personalized self-instruction in a free exploration or encyclopedic mode; and the second is to build the base on which we may subsequently explore an Intelligent Tutoring System architecture, which would take into account the expertise already gained by the learner over prior sessions. The work presented in the paper concerns the first aim.

The system is intended to provide a public of educated users with an adaptive environment for the autonomous learning and review of the properties of ideograms, such as their etymology, structure, graphic form, phonetics, semantics, composition (semantic and phonetic) within other characters... [CIYUAN 1984, NELSON A.N. 1974].

Our system is not for the study of the Chinese language or even of Chinese words (which are usually composed of more than one character), but solely for that of the characters as basic morphological units, even though instantiated with phonetic and semantic values in Chinese. We had initially intended to follow Heisig's

view [HEISIG 1977], which separates the acquisition of the universal qualities of characters (basically morphology and intrinsic meaning), from the learning of language-dependent aspects (such as the pronunciation, semantics, pragmatics in their use, etc). But for developing other applications, it seemed advisable to also include in our base a comprehensive knowledge of language-specific attributes, as it is first on Chinese language.

The normal target user is not a scholar. He (or she) is an active adult, who is not assumed to have the available time nor the focused motivation to undertake an intensive academic program (he is a scientist or a technical or industrial executive). Rather, he wishes to study the characters in a self-paced, extracurricular fashion, or to develop a multifaceted view of them according to his interests. The user wants to grasp and to memorize characters' important properties, to initiate his own "learning information base" on them, so, for example, not to feel illiterate during working visits in China or Japan. Later he might gradually enlarge and interrelate his personal knowledge items.

The learner will be given access, via an encyclopedic mode which provides a highly interactive user interface, to either basic properties or to extra specific information on the characters. He will be able to do so starting either from a key property he knows about a character, or from a partial (and even partially erroneous) description of it.

He will be also encouraged to record his personal discoveries or conclusions as additional materials: his personal mnemonics, his own mental images or self-built references on semantic links between characters, between their graphic form and meaning, etc. It has been shown that such an active and creative approach is very suitable to Western learners for autonomous acquisition and development of cognitive skills during the study of Japanese kanji [HEISIG 1977].

## 2. Overall view of the system

Central objects in the information base of the present project are: the current standard information base on characters (which merges basic and special material on their properties, textual, ideographic, and sound data), the learner's current personal information base (with his additional notes), and the learner's current profile (data he already accessed, drawn from a session journal or a global curriculum report).

The main functional handlers in the system are as follows: a learner and a developer interface, a query analyzer and character selector, a session monitor, a session observer, and a session and profile editor.

The identification and the selection of the working character (not a pattern recognition of a character drawn by hand) will interpret the learner's query, with respect to a subset of classical descriptive properties such as: the meanings, pinyin codes or stroke numbers of the character and/or of its semantic or phonetic radical if it has one. An expert assistant module will be added to enhance the interactive character selection, while managing missing or erroneous items in the query.

### 3. The information thesaurus of Chinese characters

Activities proposed in the system allow the study of a comprehensive set of properties of a character. In the information base there are actually two levels of accessibility: basic essential information, and additional more detailed (or erudite) material [WIEGER 1972, RYJIK 1980]. They are all listed here, in the context of the Chinese language.

#### Morphology:

- the etymology of the character, its iconographic origin and evolution,
- its generic category (among 6 classical ones),
- the calligraphy (the stroke order, the different writing styles, their evolution) [ZHONGGUO SHUFA DA ZIDIAN 1983],
- the structure (synthetic representation of the morphological tree of the character, semantic and/or phonetic radicals within it),
- the use in derivation or composition within other characters.

#### Phonetics:

- the pinyin encoding, the tone,
- the standard pronunciation (from a digitized sound base),
- ultimately, different provincial or colloquial pronunciations.

#### Semantics:

- the usual meaning,
- mnemonics proposed to the learner,
- learner's personal mnemonics,

- common "false friends" (misleading similarities),
- other characters in homomorphy, homonymy, usually confusing homophones,
- synonyms, antonyms.

## 4. Current prototyping

We first modelled the pertinent material on characters, and specified an interaction scheme for the user, then the learner interface.

We refer to an iterative cycle for the software development. We have prototyped first a simplified version of the system on a very small subset of the character base, in order to validate the data structure, the design of the main functionals and the user interface.

We have chosen object-oriented programming tools as well-suited to the incremental realization scheme. Thus far, this first version is being developed on Apple Macintosh under the HyperTalk-HyperCard environments, regarded as fair facilities for implementing hypertextual and voiced applications. A second level of prototyping is expected on a Xerox AI workstation using LOOPS and Common Lisp.

## 5. Further development

#### Short-term steps:

We are currently initiating the first model validation. Then are planned both a moderate quantitative, and a qualitative, extension of the system. We will first enlarge the character set to about 300 units, while monitoring both systematic assembly cost, and system efficiency measurements.

Qualitatively, the complementary properties in the character base will be completed, and the phonetics will be voiced. Next to be worked on, in the prototype, are a session journal manager, and the interactive character identification and selection function.

#### In the future:

A possible trend leads towards a system of formative use, with a real scale character base and ergonomic enrichment.

On another line for evolution, the system is a basis towards exploring knowledge based architectures, which then incorporate objects and functional handlers inherited from the design of Intelligent Information Retrieval Systems [BRUANDET 1989, CHIARAMELLA 1987] or Intelligent Tutoring Systems [WINKELS 1988, WENGER 1988].

## Conclusion

The project focuses on characters only, yet a vast field of investigation for foreigners (and one of practice and review, for native users). Moreover, extending the system capability to word formation, then to structural or pragmatic views on the language itself, would undoubtedly require much dedicated work on the language didactics.

It would be of interest however, in order to confirm or to infirm Heisig's hypothesis, to experience different practice strategies on an adequate version of the system, and to value whether one better learns characters while separating their study from that of the language, or while merging them.

A rewarding aspect lies in the scopes of future system developments: a realistic and versatile pedagogic use on widely accessible micro-stations, as well as a contribution to stepwise modelling of built-in intelligence, for Computer Based Learning Systems; and last, in the attempt to develop tools for giving larger access to Chinese characters - a vehicle for communication between over one billion people - in the frame of intercultural development.

## Acknowledgements

My deep thanks will go indeed to François Tcheou. Without his wide expertise on the Chinese language, and distinguished calligraphy, this work could certainly not have been carried out. Many thanks as well to Mohan Embar for patient reading of the first draft.

## Annex

一 二 三 四 五 六 七 八 九 十

乙 口 日 月 田 目

Première leçon (16 caractères)

古 吾 冒 朋 明 唱 晶 品 呂 昌  
早 旭 世 胃 旦 胆 肉 亘 凹 凸

Deuxième leçon (20 caractères)

## References

- BRUANDET M.F. (1989)  
*Outline of a knowledge base model for an Intelligent Information Retrieval System.*  
Information Processing and Management, Vol 25, N°3.
- CHIARAMELLA Y. & al. (1987)  
*A prototype of an intelligent system for Information Retrieval.*  
Information Processing and Management, Vol 23.
- CIYUAN (1984)  
*Comprehensive Dictionary of Chinese Characters and Words.*  
3rd edition, Shang Wu, Beijing.
- HEISIG J.W. (1977)  
*Remembering the Kanji.*  
Japan Publications Trading, Tokyo.
- NELSON A.N. (1974)  
*The Modern Reader's Japanese-English Dictionary.*  
Tuttle, 2nd rev. edition.
- RYJIK K. (1980)  
*L'Idiot Chinois.*  
Ed. Payot.
- WENGER E. (1988)  
*Artificial Intelligence and Tutoring Systems.*  
Morgan Kaufmann Pub. Inc., Los Altos.
- WIEGER L. s.j. (1972)  
*Caractères chinois. Etymologie. Graphies. Lexique.*  
8ème édition, Kuangchi Press, Taichung.
- WINKELS R. & al. (1988)  
*Didactic discourse in intelligent help systems.*  
Int. Conf. on Intelligent Tutoring Systems, Montréal ITS88.
- ZHONGGUO SHUFA DA ZIDIAN (1983)  
*Comprehensive Dictionary for Chinese Calligraphy.*  
6th edition, Zhong, Wai Ed., Hong Kong.

Ecriture 3 (8) **Prononciation** san [san]  
**Signification** trois (le nombre 3) **Sens**  
 Mnemos  
 Triple le chiffre 'un'  
 (comme le chiffre romain III)  
 Notes  
 Décomp Compos 3 1  
 O C M P S ? ↺

Ecriture 3 (8) **Prononciation** san [san]  
**Signification** trois (le nombre 3) **Sens**  
**Etymologie: san**  
 Nombre qui représente 'les 3 puissances': le ciel, la terre et l'humanité.  
 C'est l'humanité qui se trouve entre le ciel et la terre.  
 Etymologie  
 Calligraphie  
 Clé S  Clé P  
 Homographes  
 Faux Amis  
 Décomp Compos 3 1  
 O C M P S ? ↺

Ecriture 4 **Prononciation** yue [yue]  
**Signification** - lune  
 - mois **Sens**  
**Calligraphie: yue** Traces (mas)  
 丨 冂 月 月 月  
**Calligraphie: yue** Styles (s)  
 月 月 月 月 月 月 月  
 Etymologie  
 Calligraphie  
 Clé S  Clé P  
 Homographes  
 Faux Amis  
 Décomp Compos 14 14  
 D C M P S ? ↺

Ecriture 4 **Prononciation** yue [yue]  
**Signification** - lune  
 - mois **Sens**  
**Homographes: yue**  
 月, la chair, est utilisé comme clé en composition.  
 Les 2 traits horizontaux touchent les traits verticaux de la clé.  
 Utilisé seul (comme caractère) il s'écrit:  
 肉  
 Etymologie  
 Calligraphie  
 Clé S  Clé P  
 Homographes  
 Faux Amis  
 Décomp Compos 14 14  
 D C M P S ? ↺

Ecriture 5 **Prononciation** mu [mou] /moux de tête/  
**Signification** œil **Sens**  
**Etymologie: mu**  
 L'œil humain D'abord l'orbite avec les 2 paupières et la pupille.  
 Puis la pupille disparaît.  
 Enfin la figure est redressée et allongée, pour prendre moins de place.  
 目 目 目 目  
 Etymologie  
 Calligraphie  
 Clé S  Clé P  
 Homographes  
 Faux Amis  
 Décomp Compos 16 16  
 D C M P S ? ↺