Linguistically Conventionalized Ontology of Four Artifact Domains
A Study Base on Chinese Radicals

Chu-Ren Huang¹, Sheng-Yi Chen¹, Shu-Kai Hsieh², Ya-Min Chou³, Tzu-Yi Kuo¹

¹ Institute of Linguistics, Academia Sinica, Taiwan
² Department of English, National Taiwan Normal University
³ Department of International Business, Ming Chuan University

CIL18, Linguistic Studies of Ontology, Seoul, July 22, 2008
Research trend in linguistic studies of ontolog(y/ies).
- Formal vs. linguistic ontology.
- Chinese radical system offers a unique opportunity for contrast and comparison.
Historically,

- they have been widely used for over 2000 years.
- they have been used by languages that belong to different language families, (in which they are named as Hanzi/Kanji/Hanja/Chunom..., respectively).
Structurally,

- A Chinese character is an *ideogram* composed of mostly straight lines or “poly-line” strokes.
- Most of characters contain relatively independent substructures, called components (or glyphs), and some common meaning-bearing components (traditionally called radicals) are shared by different characters.
- Thus, the structure of Chinese characters can be seen to consist of a 3-layer affiliation network: character, component (glyph) and stroke.
- Traditional classification of Radicals: 540 Radicals (Shuo-Wen-Jie-Zi, Xyu Shen(121)), such as 艸、木、火，etc
- Examples: 金 (metal) → 銀 (silver), 鐵 (copper), 鐵 (iron), 鉛 (lead), ...
Linguistically, (with controversies)

- a Hanzi is regarded as an ideographic symbol representing *syllable and meaning* of a “morpheme" in spoken Chinese, or, in the case of polysyllabic word, one syllable of its sound.
- Namely, shape, morpheme and syllable are *triplicity* of a character.

Overall, the long-term historical development and broad geographical variation of Hanzi has made it a valuable resource for multi-linguistic and cross-cultural mediation in Asia, and thus as a linguistically conventionalized ontology, it is suitable for linguistic modeling and testing.
Basically, there are two types of components: Semantic components and Phonetic components.

- Semantic components are essential components of Chinese characters.
- *ShuoWenJieZi* is organized by regarding the Radical forms as semantic components.
- In *ShuoWenJieZi*, all Chinese characters are classified as derived from 540 radicals.

In this study, we assume that:

- These 540 radicals each represent a basic concept and that all derivative characters are conceptually dependent on that basic concept.
Previous Studies on Character/Radical Ontology

  - systematic attempt to couple character with ontology via WordNet-like structure
- SUMO-based conceptual mapping (cf. Hantology, Chou (2006))
  - systematic attempt to link character/radical to formal ontology
- Radicals and Generative Lexicon Theory (Pustejovsky (1995) ) (Chou and Huang (2007))
  - propose to account for radicals as linguistically conventionalized ontology by qualia structure
Assumptions in this Study

Following Chou and Huang (2007), we assume

- Radicals are relatively stable and attested ontology over thousands of years.
- Each radical group clusters as a domain ontology headed by one base concept.
- Shuo-Wen-Jie-Zi (Xu, (121))’s 540 radicals can reflect the conventionalized conceptualization.

In this study, we further examine in details four radicals of artifacts domains.
Goals of this Research: A Vision of Hanzi ontological semantics

We propose to:

**Short-term**
- construct and maintain an ontological lexical resource of based on Radical/Hanzi, which is cognitively sound and machine traceable, and based on that,
- elaborate on how shared experience and cognitive salience affects the formation of linguistic ontology.

**Long-term**
- Formulate (statistical) models that capture the evolution of Hanzi
- Facilitate the performance of relevant NLP tasks
By exploring the four radicals of artifacts domains, we would like to answer:

- if and how the conceptual extensions encoded by these radicals of artifacts differ from those by natural objects (Chou and Huang 2006)?
- do the design features of these artifacts play a role in their possible conceptual extensions?
- how human intension affects the formation of linguistic ontology?
Based on our previous studies, it shows that the conceptual clustering encoded in Radicals is not merely a simple taxonomy.

To capture how the base concept of one single radical forms a complete ontology through concept derivation, we take Aristotle’s *mode of explanation* (aitia, Physics II,3) and Pustejovsky’s Generative Lexicon Theory (Pustejovsky, 1995) as theoretical foundation, in which one of the goals is to explain the systematic relatedness between word senses in formal and predictable ways.
The Ontology of a Semantic Radical: Generative Lexicon
Approach

In particular, the network of **qualia structure**, which is viewed as expressing the componential aspect of a word’s meaning (Calzolari, 1992).

- **Formal**: (what distinguish it from others)
- **Constitutive**: (what constitute it)
- **Telic**: (what purpose it has)
- **Agentive**: (how it comes about)
Qualia Structure: system of relations that characterizes the semantics of nominals

- **Constitutive Role**: the relation between an object and its constituent parts;
  - Material
  - Weight
  - Parts and component elements

- **Formal Role**: the basic category of which distinguishes the meaning of a word within a larger domain;
  - Orientation
  - Magnitude
  - Shape
  - Dimensionality
  - Color

- **Telic Role**: purpose and function of the object
  - Purpose that an agent has in performing an act
  - Built-in function or aim that specifies certain activities

- **Agentive Role**: factors involved in its origin or “binging it about” an object
Some advantages:

- Compositional treatment of primitives (radicals/components): looking more at the generative or compositional aspects of lexical semantics rather than the decomposition into a specified number of primitives.

- QS and the Compositional Interpretation of Compounds: Instead of a taxonomy of the concepts wired in Hanzi/components, this approach could provide us the generative device to present the minimal semantic configuration of a given character, and a set of character association (字組) (collocation/compound).

- In practice, radical may be considered as ILI (Inter-Lingual-Index)-like among Sinosphere.
Through the analysis of Shuo-Wen-Jie-Zi, we suggest that conceptual extensions from the base concept encoded by a radical can be classified into seven main types:

- Formal
- Constitutive
- Telic
- Agentive
- Participant
- Participating
- Descriptive (state/manner)
Extended Qualia Structure

- 物質 Formal
  - 感官 senes
    - 視覺 vision
    - 聽覺 hearing
    - 嗅覺 smelling
    - 味覺 taste
  - 特性 characteristic
  - 専名 proper names
  - 非典型 atypical

- 組成 Constitutive
  - part
  - member
  - group

- 功用 Telic: concepts related to function or usage.
- 產生 Agentive: the relationship between the radical and its meaning cluster coming from production or giving birth are classified into agentive.
參與者 Participant relations are put in this type when the gloss in ShuoWenJieZi mentions the participant in particular.

事件 Participating: according to different events,

- action
- state
- purpose
- function
- tool
- others

描述狀態 Descriptive

- state
- manner
Some Examples of Seven types of Conceptual Extensions

- **FORMAL**: (sense, characteristic, proper names ... ) ex: 銀，白金也。
- **CONSTITUTIVE**: (part, member) ex: 睫，目旁毛也。磊，眾石貌。
- **TELIC**: ex: 鍾，酒器也。
- **PARTICIPATING**: ex: 呼，外息也。吸，內息也。
- **PARTICIPATANT**: ex: 驅，驅馬也。(人是參與者)
- **DESCRIPTIVE**: (state/manner) ex: 含，嘔也。嘔，口有所銜。/ 吐，寫也。
- **AGENTIVE**: ex: 狩，五月生羔也。鍊，冶金。
Working Interface: Search by SUMO Class

Linguistically Conventionalized Ontology of Four Artifact Domains
Linguistically Conventionalized Ontology of Four Artifact Domains
Analysis of Four Radicals of Artifacts

皿 (min3): basin / container. (說文：皿，飯食之用器也。象形。)
耒 (lei3): plow / a farm tool. (說文：耒，手耕曲木也。木推。)(即雜草)
刀 (dao1): knife / weapon. (說文：刀，兵也。象形。)
网 (wang3): weaving a net / catching/fishing. (說文：网，庖羲所結繩，以田以漁也。)
The Qualia Structure on Derivative Concepts of 皿

皿 (min3) Basic concept: container
耒 (lei3) Basic concept: a farm tool

Diagram:

- Basic concept: a farm tool
- Telic: 耒、耜
- Participating: 耒、耘、耤、耰
The Qualia Structure on Derivative Concepts of 刀

刀 (dao1) Basic concept: 1. knife 2. weapon

Linguistically Conventionalized Ontology of Four Artifact Domains
The Qualia Structure on Derivative Concepts of 网

网 (wang3) Basic concept: 1. catching/fishing, 2. net
Findings 1: Conceptual Dependency

The primary meanings of characters that shared the same radical symbol are indeed conceptually dependent on the basic concept of that radical.

- 网 (wang3) two key meanings:
  1. catching / fishing, ex: 罗 (luo2): a tool to catch bird
  2. weaving a net, ex: 网 舞 (wu3): a latticed window that looks like a reticulation.
Findings -2: Dimensions of conceptual extensions

- Natural objects v.s Artifacts
Findings -2: Dimensions of conceptual extensions

- Artifacts are designed with a specific functionality
  - so, most of the types of conceptual extensions belong to telic.
- The concept of an artifact can best be understood by how it is used
  - hence a character often denotes a typical event in which the artifact is a main participant
Findings-3: Semantic coverage and Generative power

- Different generative power
  - 脈 (container; 28 derived characters)
  - 粿 (a farm tool, 8 derived characters)

粿 is a kind of farming tool, so its event function is task-oriented and socially defined. Therefore, the generative power is more restricted; 脈 on the contrary, as a container, is a basic tool with generic purpose, so its capability of generating new characters is less restricted.
an artifact that is a human imitation of natural object or function is conceptually more versatile and can serve as the base of conceptual extensions similar to natural object.

a human invention with functional components, is directly restricted by its intended function and limited in conceptual extensions.

in both cases, however, eventive conceptual extension occurs frequently based on the event associated with the function of that artifact.
Further research

- Further analysis on other categories of Chinese radicals
  - Investigate ontological analogy and characteristic of different categories of Chinese radicals
- Establish the ontology of Chinese radicals systematically, e.g., formally represent the resultant ontology by mapping it to other formal ontology.

In conclusion, we believe that this work can provide a solid foundation that is flexible enough to capture the generative nature of Chinese lexicon.