

Chinese Characters and Top Ontology in EuroWordNet

Shun Ha Sylvia Wong

Computer Science
Aston University
Birmingham, U.K.

Karel Pala

Faculty of Informatics
Masaryk University
Brno, Czech Republic

s.h.s.wong@aston.ac.uk pala@fi.muni.cz

presented by: Pavel Smrž

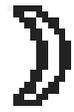
22nd January, 2002.

Motivation

- Various formalized lexical databases have been developed, e.g. WordNet 1.5 (Miller et al. 1990), CyC (Lenat & Guha 1990), HowNet (Dong & Dong 1999) and EuroWordNet 1, 2 (EWN) (Vossen et al. 1999).
- These databases all contain some hierarchy of language-independent concepts which reflects the important semantic distinctions of each concept.
- They differ in how they organize such concepts within the hierarchy.
- Two questions:
 1. How *faithfully* and *accurately* do such artificial constructs model the complicated groupings of real-world concepts?
 2. Would the *arbitrariness* which exists in such artificial constructs hinder their effectiveness in knowledge representation?
- Motivated by these questions, we studied how concepts are organized in Chinese and compared the result with EuroWordNet (EWN) Top Ontology (TO).
- Our ultimate goal is to derive new improved ways to knowledge representation.

Chinese Characters

- Chinese script has originated from picture-writing.



= 月 (*moon*)



= 魚 (*fish*)



= 寶 (*treasure*)

- The shapes of a Chinese character displays a fair amount of the meaning that it represents. E.g.:
 - the ancient form of 日 (*sun*) resembles ☉
 - 羊 (*sheep*) is a pictograph of a sheep with horns
 - putting 日 (*sun*) and 月 (*moon*) together forms 明 (*bright*)
 - putting two 木 (*tree*) together forms 林 (*forest*)

Chinese Characters (Cont'd)

- *An interesting thing ...*

A Chinese can look at a previously unknown Chinese character and be put in an appropriate context easily. Try:

魴 鮓 鮓 鮓 鮓 魴

Question: What do you think they mean¹?

- There exists a grouping of Chinese characters according to their meanings.
- In most cases, this grouping of Chinese characters is semantically motivated.

¹*They all mean some kind of fish.*

Entities in the 1st level of EWN TO

There are three types of entities distinguished in the first level of EWN TO (Vossen et al. 1999):

- 1st Order – any concrete entity publicly perceivable by the senses and located at any point in time, in a three-dimensional space, e.g. individual persons, animals and more or less discrete physical objects and physical substances. They are always denoted by (concrete) nouns.
- 2nd Order – any Static Situation (property, relation) or Dynamic Situation, which cannot be grasped, heard, seen, felt as an independent physical thing. They occur or take place rather than exist, e.g. continue, occur, apply, and also events, processes, states-of-affairs or situations that can be located in time belong here. They can be expressed by nouns, verbs and adjectives.
- 3rd Order – unobservable propositions which exist independently of time and space. They can be true or false rather than real. They can be asserted or denied, remembered or forgotten, e.g. ideas, thoughts, theories, plans, hypotheses, reasons, and they are always expressed by (abstract) nouns.

Chinese data and EWN 3rdOrderEntities

- Concepts in the 3rdOrderEntities display a propositional nature which make them fairly difficult to grasp.
- Unlike those in the 1st and 2nd order entities, the concepts in this entity list are not organized hierarchically.
- (Wong & Pala 2001) have observed that a direct correspondence between the concepts in this entity list and the Chinese radical does not seem to exist.
- We therefore turned to Chinese words for establishing a comparison.
- What we did:
 - We look up Chinese words which represent each basic concept in the 3rdOrder-Entities from various English-Chinese dictionaries, and
 - analyzed the meaning of individual characters which forms part of each word.
- The result shows that Chinese mainly uses a subset of characters to represent the concepts in the EWN 3rdOrderEntities.

	theory	reason	hypothesis	idea/thought	structure	evidence	doctrine	policy	content	procedure	concept	plan/plan of action	communication	knowledge base	cognitive content	know-how	category	information/data point	abstract (info)
說 (theory/to explain/to say)	✓	✓				✓													
設 (to establish/to set up)			✓								✓								
識 (to know/knowledge)													✓	✓				✓	
學 (study/theory/discipline)	✓					✓													
理 (logic/reason/theory)	✓	✓		✓						✓									
原 (origin)	✓	✓																	
法 (law/method)			✓	✓	✓			✓											
主 (master/main)				✓		✓													
觀 (to observe/view)				✓						✓									
念 (to remember/to study)				✓						✓									
息 (breath/news)													✓						✓
義 (principle/morality/basic truth/meaning)						✓		✓											
方 (square/a place/direction/side/method)								✓			✓				✓				
案 (legal case/record/plan)								✓			✓								
要 (to need/important)								✓											✓
分 (to distribute/to distinguish)								✓										✓	
聞 (to hear/to smell)												✓	✓						
知 (to know/knowledge)														✓					✓

Figure 1: A distribution of Chinese characters used in different 3rdOrderEntity basic concepts

Chinese data and EWN 3rdOrderEntities (Cont'd)

- Many characters in this subset interact with each other to form a related or very different concept, e.g.:
 - 假設 (*fake/pseudo/to borrow + to establish/to set up = assumption*) versus 設計 (*to establish/to set up + to calculate/plan/scheme = design*)
 - 理論 (*logic/reason/theory + opinion/theory/discussion = theory*) versus 理由 (*logic/reason/theory + from = reason*) versus 理念 (*logic/reason/theory + to remember/to study = idea/concept*)
- *Back to EWN TO ...*

It seems to lack the dynamics which allows one to combine related primary concepts to form secondary concepts.
- It appears that the unique way of the evolution of Chinese script facilitates the study of meaning transformation because such phenomenon is more traceable in Chinese.

The Chinese way to represent concepts

- Chinese seems to organize concepts in a *contextual* manner:
Each Chinese radical serves as the characterizing basic concept in the respective context.
- To investigate how this works, we studied (the more well-known subset of) characters grouped under seven Chinese radicals.
- The result shows that each group of Chinese characters can be classified along *five* main lines:
 1. as an object
 2. as a property
 3. an typical event (situation, process)
 4. its component
 5. as an consequence

This classification captures the line in which the concept represented by a Chinese radical is projecting itself along.

The Chinese way to represent concepts (Cont'd)

- Such a classification suggests that relevant concepts in the same 'context' can be arranged in the form of a small semantic network whose structure may look like Figure 2.

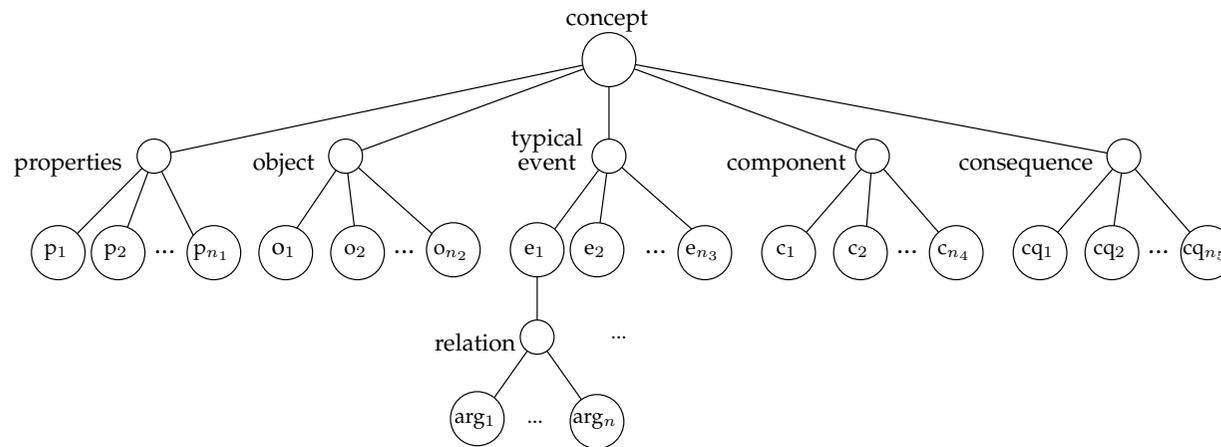


Figure 2: A new way for organizing concepts – a schema

- When compared with EWN TO, the realization of such an organization would be richer as it is centered around a *semantic* context, rather than *syntactic* categories.
- We believe that this organization of concepts reflects how humans organize and process conceptual knowledge better.

Conclusion

- The presented study is a continuation of Wong & Pala's (2001) work.
- Our main findings are:
 - Chinese data offer some new views of concept organization.
 - This organization could systematically enrich the existing EWN TO and make it more natural and better structured.
 - The origin atomic concepts in the present EWN TO could be viewed as graph structures for capturing concept relations.
 - Such structures could inspire us to derive a better formulation of inference rules for making a more realistic and intelligent reasoning possible.
- Please direct further discussions to:

Shun Ha Sylvia Wong & *Karel Pala*
s.h.s.wong@aston.ac.uk pala@fi.muni.cz

References

- Baker, M. A. (1998), 'Mandarin Chinese outpost – characters', [Online]. Available at: http://chinese_outpost.tripod.com/chars.html [2001, March 21].
- Dong, Z. & Dong, Q. (1999), 'HowNet', [Online] Available at: http://www.keenage.com/zhiwang/e_zhiwang.html [2001, June 7].
- Harbaugh, R. (1996), 'Zhongwen.com – Chinese Characters and Culture', [Online]. Available at: <http://www.zhongwen.com/> [2001, March 19].
- Harbaugh, R., ed. (1998), *Chinese Characters: A Genealogy and Dictionary*, Han Lu, Taipei. Also appeared in: [Online]. <http://www.zhongwen.com/m/search.htm> [2001, March 19].
- Hornby, A. S., ed. (1984), *Oxford advanced learner's English-Chinese dictionary*, Oxford University Press; Keys Publishing, Hong Kong.
- Lenat, D. & Guha, R. (1990), *Building Large Knowledge-based Systems – Representation and Inference in the CyC Project*, Addison Wesley.

- Lu, A. Y.-C. (1998), *Phonetic Motivation – A Study of the Relationship between Form and Meaning*, PhD thesis, Department of Philology, Ruhr University, Bochum.
- Miller, G. A. et al. (1990), *Five papers on WordNet*, Technical report, Princeton University. CSL Report 43, Cognitive Science Laboratory.
- Muller, A. C. (2000), 'Dictionary of East Asian literary terms', [Online]. Available at: <http://www.human.toyogakuen-u.ac.jp/acmuller/dicts/dealt/index.htm> [2001, March 19].
- Peterson, E. (2000), 'Chinese character dictionary', [Online]. Available at: http://www.mandarintools.com/chardict_rs.html [2001, March 19].
- Vossen, P. et al. (1998), *The EWN base concepts and top ontology*, Technical report, University of Amsterdam, Amsterdam. Deliverables D017, D034, D036, EuroWordNet, LE2-4003, Final Version.
- Vossen, P. et al. (1999), *Final report on EuroWordNet 2*, Technical report, University of Amsterdam, Amsterdam. [CD ROM].

Wah Tung Committee, ed. (1983), 詞淵 (*a Chinese word dictionary*), Wah Tung, Hong Kong. [In Chinese].

Wong, S. H. S. & Pala, K. (2001), Chinese Radicals and Top Ontology in WordNet, *in* 'Text, Speech and Dialogue—Proceedings of the Fourth International Workshop, TSD 2001, Pilsen, 10–13 September 2001', Lecture Notes in Artificial Intelligence, Subseries of Lecture Notes in Computer Sciences, Faculty of Applied Sciences, University of West Bohemia, Springer, Berlin.