Title: Chinese Characters and Top Ontology in EuroWordNet

Paper by: Shun Sylvia Wong &
Karel Pala

Presentation By: Patrick Baker

Introduction

- WordNet, Cyc, HowNet, and EuroWordNet each use a hierarchical structure of language independent concepts to reflect the important semantic differences between concepts
- EuroWordNet uses a hierarchy called Top Ontology (TO)
- This paper compares EuroWordNet's TO with the natural organization found in the pictographic based Chinese language

Top Ontology?

- Ontologies are artificial constructs built with the primary purpose to serve as the lexical databases for knowledge representation systems
- Top Ontology distinguishes between three types of entities
- This paper focuses on the third type

The Three Entity Types of TO:

- There are three types of entities distinguished at the first level of TO:
  1. 1st Order – any concrete entity publicly perceivable by the senses and located at any point in time, in a three-dimensional space (persons, animals, discrete objects)
  2. 2nd Order – any Static Situation (property, relation) or Dynamic Situation, which cannot be grasped, heard, seen, felt as an independent thing (events, processes, states-of-affair)
  3. 3rd Order – unobservable propositions which exist independently of time and space. They can be true or false rather than real (ideas, thoughts, theories, plans, reasons)

The Chinese Language

- Chinese script originated from picture-writing
- Only a couple hundred characters in the language are actual pictograms
- According to the etymological dictionary written by Xu Shen around 100 A.D., Chinese characters can be divided into six groups

Six Groups of Chinese Characters

1. Pictographs (≈4%): represent real-life objects by drawings
2. Ideographs (≈1%): represent positional and numeral concepts by indication
3. Logical Aggregates (≈13%): form a new meaning by combining the meanings of two or more characters
4. Phonetic Complexes (≈82%): form a character by combining the meaning of one character and another character which links through a shared sound
5. Associative Transformations (a small portion): extend the meaning of a character by adding more parts to the existing one
6. Borrowings (a small portion): to borrow the written form of a character with the same sound

The Chinese Language

- The average educated Chinese person knows only about 6000 of the 50,000 characters in the Chinese language
- Since many of the characters are combinations of simpler characters, knowing the meaning of one or more of the constituent characters allows deduction of the overall meaning

The Chinese Language

- Because Chinese characters can not be ordered alphabetically in a dictionary, they are ordered by Section Heads or Chinese Radicals
- There are 213 Chinese Radicals
- In most cases, a character is grouped under a certain Chinese Radical if its concept relates to the concept represented by the radical in some way

The Chinese Language and 3rd Order Entities

- The concepts in the 3rd Order Entity list are abstract and difficult to grasp; most are represented by use in the form of a sentence (e.g. “John thought the movie was good”)
- Wong & Pala (2001) have shown that no direct correspondence can be found between Chinese Radicals and the concepts in the 3rd Order list
- In most cases, the Chinese counterparts of these concepts are represented by more complicated lists of characters

The Chinese Language and 3rd Order Entities

- For each of the basic concepts in the 3rd Order list, the authors located their Chinese counterparts
- Each concept created a list of Chinese characters representing synonyms, hyperonyms, and/or meanings that collectively defined the scope of the concept
- The meanings of the component radicals of each character in the list were then examined

The Chinese Language and 3rd Order Entities

- The authors found that certain radicals (with specific meanings) were found associated with one or two 3rd Order concepts
- This association is called *Sense Transfer*
- e.g. the characters 理 (logic/reason/theory), 論 (opinion/theory/discussion), and 說 (theory/to explain/to say) appear more often under theory
- e.g. the characters 想 (to think/to consider) and 思 (to think/to contemplate) appear more often under idea/thought

Sense Transfer and Other Languages

- Sense transfer exists in most languages, though not necessarily to the extent as pictograph based languages
- English examples: care-free, side-light, un-think-able
- Czech example: u_-i-t-el (a root denoting the concept ‘teach’ + a verb-making affix + an infinitive affix + an agentives suffix = teacher)
- The inadequacy of existing ontologies to show this sense transfer property means there exists no way to derive the meaning for a new word even if its components already exist in the ontology

The Chinese Way to Represent Concepts

- Wong & Pala (2001) have observed that Chinese seems to organize concepts in a contextual manner, with each Chinese radical serving as the characterizing basic concept in the respective concept
- Through observation, the authors determined that many of the characters subsumed in the radicals can be classified along five main lines

The Chinese Way to Represent Concepts

- The five conceptual lines are:
  1. As an object
  2. As a property
  3. As a typical event (situation, process)
  4. It's component
  5. As a consequence
- e.g. the character  (fire) 'as an object' is part of  (stove) and  (charcoal), and 'as a typical event' is part of  (to burn) and  (to cremate)

Lexical/Conceptual Organization

- The Chinese way of organizing concepts (even abstract ones) from simpler, more concrete concepts/entities provides an alternative to the organization provided by existing ontologies
- Such an organization would form a semantic network as opposed to the tree structure found in such ontologies
- Such a semantic network is richer, more complete, and more transparent, as each concept is derived not from verbalized concepts, but a semantic context of discrete entities

Conclusion

- By comparing EuroWordNet's TO to the intrinsic structure provided by the natural language Chinese, it can be seen that:
  - Humans more naturally think of concepts as being composed of more concrete entities, as opposed to derived from abstract concepts
  - The more natural way to represent such concepts is a semantic graph, not the tree structure found in most existing ontologies