

# Introduction to Ontology

## Application on Linguistic Studies

Author: *Huang (Patrick) HUANG*

Supervisors: *Prof. Heinz Preisig*

*Dr. John Morud*

# What I'm Trying to do...

- *To find out enough relevant information and sorted out logically.*
- *Put forward some “Patrick’s Hypotheses” for further research.*
- *Suggestions for future work.*
- ~~*Avoid professor sending my report to faculty of philosophy.*~~

# Structure

- *Background knowledge of ontology*
- *Review of relevant disciplines*
- *Analysis of ontology databases*
- *Assumptions & Hypotheses*
- *Discussions & Suggestions*

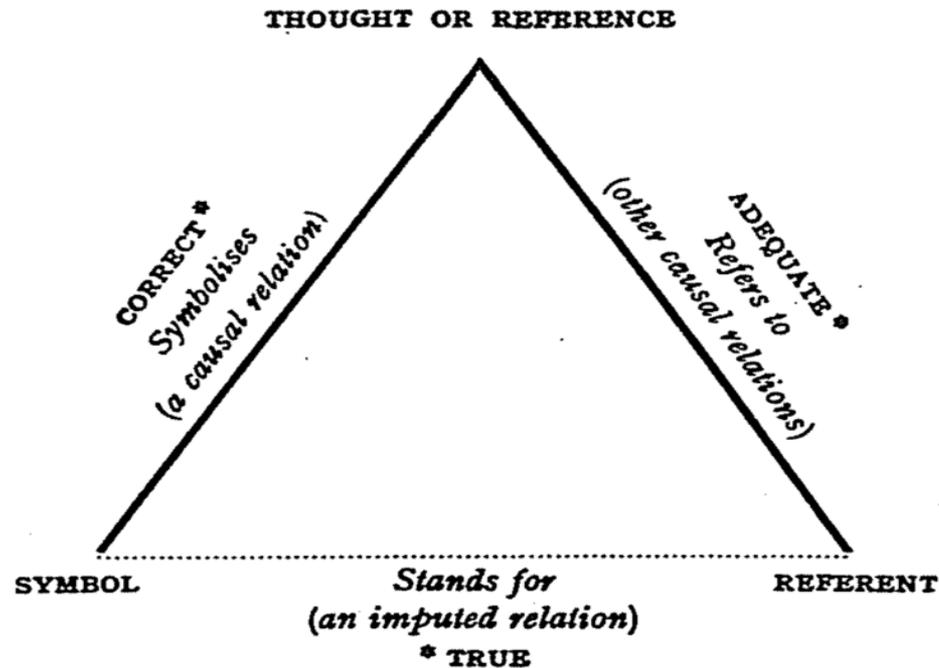
# Definition

- *Definition from Philosophical view:*
- *"Nature of being, becoming, existence, or reality".*
- *In this report, the key point is "whether a term is referring to something really exists, or just representing a concept".*

# Definition

- *Definition from Information Science:*
- *"Specification of a conceptualization".*
- *Specification – a detailed description*
- *Conceptualization – a very specific object / concept*
- *(Should be unique or with less ambiguity)*
- *"Ogden's Triangle of Reference"*

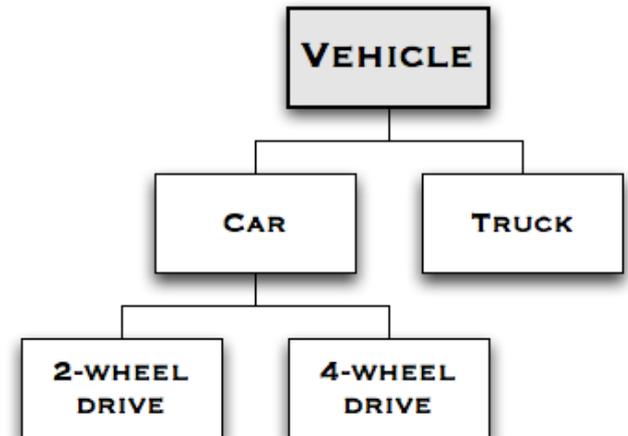
# Triangle of Reference



- *Some similarities with modern linguistics.. We will discuss it later.*

# Ontology Databases

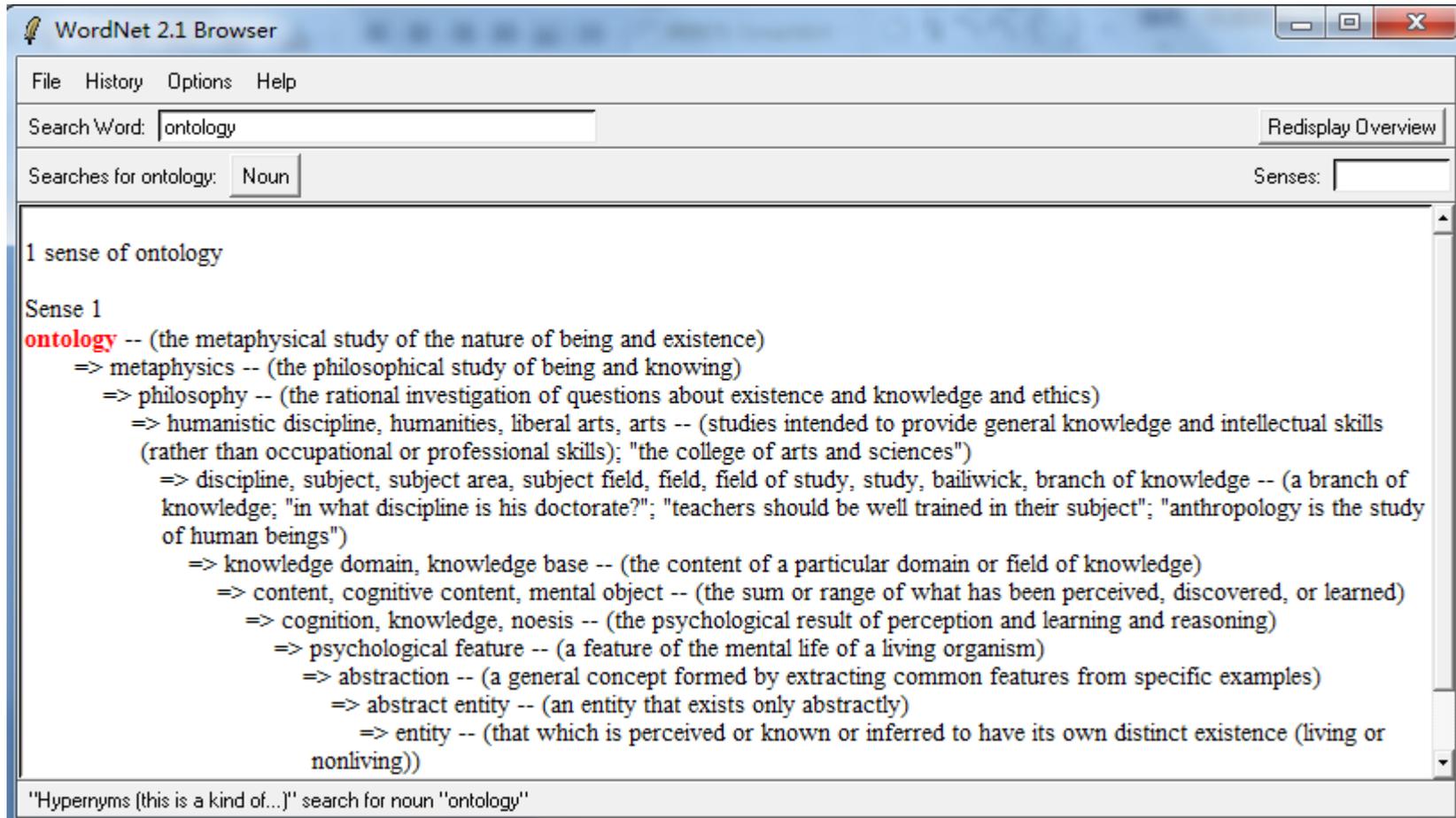
- *Advantages: Can clearly show the logical relationship between elements.*
- *Disadvantages: Weak for presenting the meaning of each term, especially in linguistic analysis.*



# WordNet Databases

- *Lexical referencing system*
- *Similar to dictionary, but more details*
- *Goal: readable for computers / AIs*
- *Helpful for automatic translation etc.*

# WordNet Example



WordNet 2.1 Browser

File History Options Help

Search Word:  Redisplay Overview

Searches for ontology:  Senses:

1 sense of ontology

Sense 1

**ontology** -- (the metaphysical study of the nature of being and existence)

- => metaphysics -- (the philosophical study of being and knowing)
  - => philosophy -- (the rational investigation of questions about existence and knowledge and ethics)
    - => humanistic discipline, humanities, liberal arts, arts -- (studies intended to provide general knowledge and intellectual skills (rather than occupational or professional skills); "the college of arts and sciences")
      - => discipline, subject, subject area, subject field, field, field of study, study, bailiwick, branch of knowledge -- (a branch of knowledge; "in what discipline is his doctorate?"; "teachers should be well trained in their subject"; "anthropology is the study of human beings")
        - => knowledge domain, knowledge base -- (the content of a particular domain or field of knowledge)
          - => content, cognitive content, mental object -- (the sum or range of what has been perceived, discovered, or learned)
            - => cognition, knowledge, noesis -- (the psychological result of perception and learning and reasoning)
              - => psychological feature -- (a feature of the mental life of a living organism)
                - => abstraction -- (a general concept formed by extracting common features from specific examples)
                  - => abstract entity -- (an entity that exists only abstractly)
                    - => entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

"Hypernyms (this is a kind of...)" search for noun "ontology"

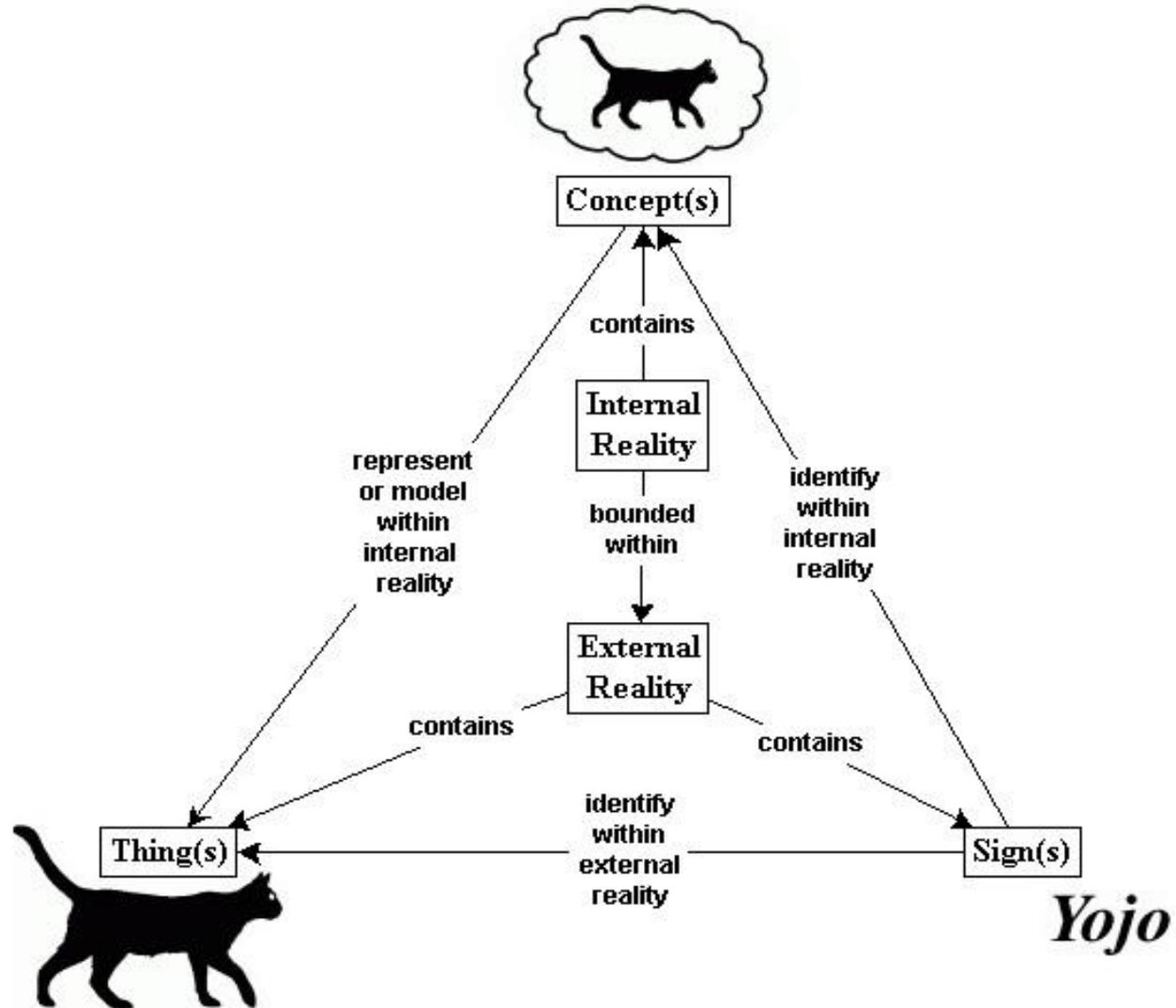
# Database Improvement

- *WordNet structure could be used.*
- *Improvement may apply.*
- *Will be introduced after “Theory of Description”.*

# Linguistics

- *Saussure Linguistics*
- *Saussure turned traditional linguistics studies from Humanities to Science of Symbol.*
- *Well known as “Semiology” or “Semiotics”.*
- *“Signifier” -> “Signified”*
  
- *There is also a type of linguistic that use cryptology to research unknown language.. It's worth considered but not described in this report...*

# Examples



# Chomsky's Theory

- *“Universal Grammar Theory”*
- *Main idea: The studying process of different languages are common. i.e. children will follow the same rule to study different languages.*
- *Inference: Language only studied by the universal grammar, without touching with the specific cultural background is sufficient.*
- *Chomsky's “Generative Grammar”.. Introduced later..*

# Quantitative Linguistics

- *“Martin’s Law”*
- *“Several lexical chains obtained by looking up the definition of a word in dictionary, then looking up the definition of the definition obtained and so on. Those lexical chains are with different "levels", which forms a hierarchy of more and more general meanings.”*
- *More explanations in later sessions.*

# Philosophical Background

- *“Identity”*
- *Aristotle - start discussing*
- *Descartes - "Cogito ergo sum" (which means I think, therefore I exist), no further discussion.*
- *Kant – Identity Issue*
- *Russell – Theory of Description*
- ~~*Wittgenstein – Ignored, I don't want to make everyone sleepy.*~~

# Philosophical Background

- *“Kant’s Identity Issue”*
- *“Patrick is Patrick” - Analytic proposition, logically true but meaningless.*
- *“Patrick is an NTNU student” - Synthetic proposition. Not possible to judge by logic, should be determined by fact (therefore database is required).*

# Philosophical Background

- *Russell's Law of Description*
- *Patrick (Identity) is an NTNU student (Description).*
- *Problem solved.*
- *This also inspires Chomsky's "Generative Grammar", but this topic is too huge...*

# Analysis

- *Use of Russell's Law*
- *The "tags" for word explanation can be series of descriptions. – (May be applied in analytical philosophy studies also).*
- *Advantages:*
  - *Shorter descriptions.*
  - *Database simplified.*

# Suggestion of Databases

- *Adding “tags” for “judgment”.*
- *Examples: Ontology of “Patrick Huang”*
- *Male (true, leave it)*
- *Lives in Oslo (false, delete)*
- *23 years old (false, but becomes true next year – the database could be dynamic)*
- ~~*Failed the module “Advanced Process Simulation”  
(Unknown now, depends on our professor).*~~

# Suggestion of Databases

- *Examples discussed in report:*
- *Definite articles translation between English and German.*

# Patrick's Hypothesis

- *1. Simulation of Proto-language*
- *Proto-language: The beginning state of a language.*
- *May be simulated by developing process of Pidgin – Creole – “Stabled” Language.*

# Patrick's Hypothesis

- *Pidgin: simplified and undeveloped language by communication between people without common languages.*
- *Creole: After Pidgin was stabilized and become mother language of second generation...*

# Patrick's Hypothesis

- *Example of Pidgin Languages*
- *Pidgin was lack of vocabulary, so there are three ways: creating some new words, borrow words or explain new things by existing words.*
- *This is a common phenomenon... We'll talk about it later..*

# Patrick's Hypothesis

- *Pidgin in Papua New Guinea*
- *Accordion: liklik box you pull him he cry you push him he cry.*
- *Bank: Money House.*
- *Beard: Grass belong face.*
- *A very thin person: Bone nothing*
- *May be simplified, or become fixed noun or slang / saying / twister etc..*

# Patrick's Hypothesis

- *2. Simulation of language development*
- *e.g. English:*
- *Anglo-Saxon & Celtic -> mixing-> pidgin state -> development -> Normandy Conquest -> mixing -> pidgin (fully pidgin-  $a:b = 1:1$ , nothing affected –  $a:b = \text{infinity:zero}$ ) ->pidgin state -> development.. etc.*

# Patrick's Hypothesis

- *2. Simulation of language development*
- *e.g. English:*
- *Simulated as a mixed model with specific properties.*
- ~~*Potential argument –*~~
- ~~*Englishmen said: Our language is NOT a liquid!*~~

# Patrick's Hypothesis

- 3. *"Equilibrium" of Language Development*
- *Three methods for adding objects: Create new words, Borrow from somewhere else, Represent new objects by existing words.*
- *The Pidgin language shows that we may have a very big "error" (ambiguity) when using something to represent (e.g. "Grass belong face" may be eyebrow also..)*

# Patrick's Hypothesis

- 3. *"Equilibrium" of Language Development*
- *The analysis was shown in the report.. (why grass belong face can be used and why there is error..)*
- *[Grass: green coloured, strip shape, will grow, organism, etc.]*
- *[Unknown substance (beard): black coloured, strip shape, will grow, etc.]*
- *[Face includes: "eye", "nose", "mouth", etc.]*

# Patrick's Hypothesis

- 3. *"Equilibrium" of Language Development*
- *If we assume an "ideal language" – every identities with a specific word, there is no error, but meaningless – it occupied infinity of human's memory.*
- *Nothing was perfectly identical, so using something to represent others makes error. If making maximum replace, the error is largest (affects communication).*
- *So there is an ideal "equilibrium" point...*

# Discussions

- *Simplification of English:*
- *Too much useless and repetitive words! – in our words, a lot of identities could be replaced by words of descriptions with a small error.*
- *Not good for propagation, since you need to remember a lot of words..*

# Discussions

- *Simplification of English:*
- *Compared with Chinese:*
- *English: June, Chinese: 六月 (six month).*
- *English: Pork, Chinese: 猪肉 (pig meat).*
- *English: Diabetes, Chinese: 糖尿病 (sugar urine disease).*

# Discussions

- *Technical Problem:*
- *Lack of database!*
- *Traditional linguistic research has sufficient data but not readable by computers, ontology research database is readable but very limited, some of the databases are even deadlink..*

# Discussions

- *Technical Problem:*
- *The topic is too large...*
- *Cannot make everything deep enough, could only do some introductory work.*

# Discussions

- *Technical Problem:*
- *Simulation may not representing the fact..*
- *For example, we use pidgin to simulate proto-language, which may not be true, since we don't have much relevant information..*
- *(As we know many great economics models have big errors...)*

# Discussions

- *Technical Problem:*
- *Not every relevant things were talked about..*
- *Cryptography – probably be used*
- *Chomsky's Generative Grammar – One semester is too short to me...*
- *Artificial language – Could be used for design a “reasonable” language.. Maybe useful for further simulation..*

# Conclusions

- *Many basic concepts could be used and understood by an ontology view.*
- *Indicators could be made to explain some concepts, e.g. level of "literature".*
- *"Linearization" of language.*
- *Final sentence: With the development of ontology, the view of linguistic studies could be completely changed.*

# References

- Altmann, G. (1993). *Science and linguistics*. In *Contributions to quantitative linguistics* (pp. 3-10). Springer Netherlands.
- Chandler, D. (2007). *Semiotics: the basics*. Routledge. Chicago
- Chazelle, B. (n.d.). *Introductory models & basic concepts: semiotics*. [ONLINE] Available at:  
<http://www.cs.princeton.edu/~chazelle/courses/BIB/semio1.html>.
- Chomsky, N., R, DiNozzi. (1972). *Language and mind*. Harcourt Brace Jovanovich. New York
- Cooke, H.P. (1983) *Aristotle Categories Vol. 1*. Harvard University Press. Cambridge.

# References

- George A. Miller (1995). *WordNet: A Lexical Database for English*. Communications of the ACM Vol. 38, No. 11: 39-41.
- Gruber, T. R. (1993). *A translation approach to portable ontology specifications*. Knowledge acquisition, 5(2), 199-220.
- Kaye, A. S., & Tosco, M. (2001). *Pidgin and creole languages: A basic introduction (Vol. 5)*. Lincom Europa. Chicago
- Kitcher, P. (1982). *Kant on self-identity*. The philosophical review, 41-72. Chicago
- Nöth, W. (1995). *Handbook of semiotics*. Indiana University Press. Bloomington

# References

- Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning*. New York.
- Pharo, N. 2014. *The semantic web*. [ONLINE] Available at: [http://www.jbi.hio.no/bibin/dig\\_korg/sem\\_web.htm](http://www.jbi.hio.no/bibin/dig_korg/sem_web.htm).
- Reicher, M. (2010). *Nonexistent objects*. [ONLINE] Available at: <http://stanford.library.usyd.edu.au/entries/nonexistent-objects/>
- Smith, B. (1998). *An introduction to ontology*.
- The Guardian. (2012). *Prince Charles in Papua New Guinea: how to speak pidgin English like a royal*. [ONLINE] Available at: <http://www.theguardian.com/uk/shortcuts/2012/nov/05/prince-charles-papua-new-guinea>

**Have a nice exam &  
Merry Christmas!**

*Huang (Patrick) HUANG*

# Thank you!

*Huang (Patrick) HUANG*