A Dynamic Theory of Ontology

John F. Sowa

VivoMind Intelligence, Inc.

Abstract. Natural languages are easy to learn by infants, they can express any thought that any adult might ever conceive, and they accommodate the limitations of human breathing rates and short-term memory. The first property implies a finite vocabulary, the second implies infinite extensibility, and the third implies a small upper bound on the length of phrases. Together, they imply that most words in a natural language will have an open-ended number of senses — ambiguity is inevitable. Peirce and Wittgenstein are two philosophers who understood that vagueness and ambiguity are not defects in language, but essential properties that enable it to accommodate anything and everything that people need to say. In analyzing the ambiguities, Wittgenstein developed his theory of *language games*, which allow words to have different senses in different contexts, applications, or modes of use. Recent developments in lexical semantics, which are remarkably compatible with the views of Peirce and Wittgenstein, are based on the recognition that words have an open-ended number of dynamically changing and context-dependent *microsenses*. The resulting flexibility enables natural languages to adapt to any possible subject from any perspective for any humanly conceivable purpose. To achieve a comparable level of flexibility with formal ontologies, this paper proposes an organization with a dynamically evolving collection of formal theories, systematic mappings to and from formal concept types and informal lexicons of natural language terms, and a modularity that allows independent distributed development and extension of all resources, formal and informal.

Corresponding paper: http://www.jfsowa.com/pubs/dynonto.htm

A more recent paper that adds further detail: http://www.jfsowa.com/pubs/lgsema.pdf

How can we deal with incompatible systems?

Problem:

- Every agent (human, animal, or robot) is idiosyncratic.
- Their ontologies (implicit or explicit) are rarely, if ever, consistent with one another.
- How can they interoperate and cooperate?

Solution:

- When you talk to a dentist about a toothache, a plumber about a leak, or a waiter about the menu,
- You may revise some details of your ontology,
- But you don´t do a global realignment of your world knowledge.
- You merely revise that tiny task-oriented subset.
- Inconsistencies outside that subset are irrelevant.

Fundamental principle:

- Interoperability is always limited to a specific task.
- Global interoperability is impossible and irrelevant.

Task-Oriented Microtheories

Observation:

- **Every interaction is context dependent.**
- **What is relevant to an interaction (and to determining context) depends on the goals of the agents involved.**
- **Global alignment of the full ontologies of all the agents is unnecessary.**
- **Only tiny subsets of each agent´s ontologies are relevant.**
- **Everything outside those subsets can be ignored for the purpose of the interaction.**

Design principles:

- **Focus on the task and the information transferred in each interaction.**
- **Design a subontology (microtheory) to represent that information.**

Precision and Ambiguity

Charles Sanders Peirce:

> **"It is easy to speak with precision upon a general theme. Only, one must commonly surrender all ambition to be certain.**
>
> **"It is equally easy to be certain.**
> **One has only to be sufficiently vague.**
>
> **"It is not so difficult to be pretty precise and fairly certain at once about a very narrow subject."**

Ludwig Wittgenstein:

> **A word can have a fixed meaning only within a particular language game.**

Many linguists:

> **Lexical ambiguity can be avoided within a semantically restricted sublanguage, such as weather reports or airplane reservations.**

Question: How should we organize our knowledge about those "very narrow subjects," "language games," or "semantically restricted sublanguages"?

Wittgenstein´s Language Games

A language game is a sublanguage of a natural language

- **With a semantically restricted vocabulary and ontology,**
- **Possibly with some specialized syntactic features,**
- **Used to support or accompany a specific task or activity.**

New language games or new variations of old games can be created at any time.

A natural language is the totality of all language games that may be played with a given syntax and vocabulary.

Multiple Word Senses

The word "car" —

- **1900: A horseless carriage, complete with hooks to attach a real horse when it breaks down.**
- **Today: A metal cabin with more computing power than the supercomputers of the 1980s.**
- **Tomorrow: Acronym for Computerized Autonomous Rover.**

The word "file" in computer systems has different microsenses for every OS —

- **IBM mainframe: A bit string separated into records by the operating system.**
- **Unix: A character string separated by new-line characters.**
- **Macintosh: A character string separated by carriage-return characters.**
- **Windows: A character string separated by new-line plus carriage-return.**

Microsenses

A term coined by Alan Cruse for the infinite variability of word senses.

In any ontology, each word sense corresponds to a distinct type, and each microsense of that word corresponds to a distinct subtype.

In computer systems, every release and every patch to the software creates new microsenses whose formal definitions may be inconsistent with the old definitions.

Ambiguity is not a property of natural languages, but the result of using and reusing a finite vocabulary to talk about an arbitrarily complex and ever-changing world.

A Neo-Wittgenstenian Model of Language

Developed by Margaret Masterman —

One of six students in Wittgenstein´s course of 1933-34 whose notes were compiled as *The Blue Book*.

Founded the Cambridge Language Research Unit (CLRU) in the late 1950s.

Emphasized semantics, not syntax:

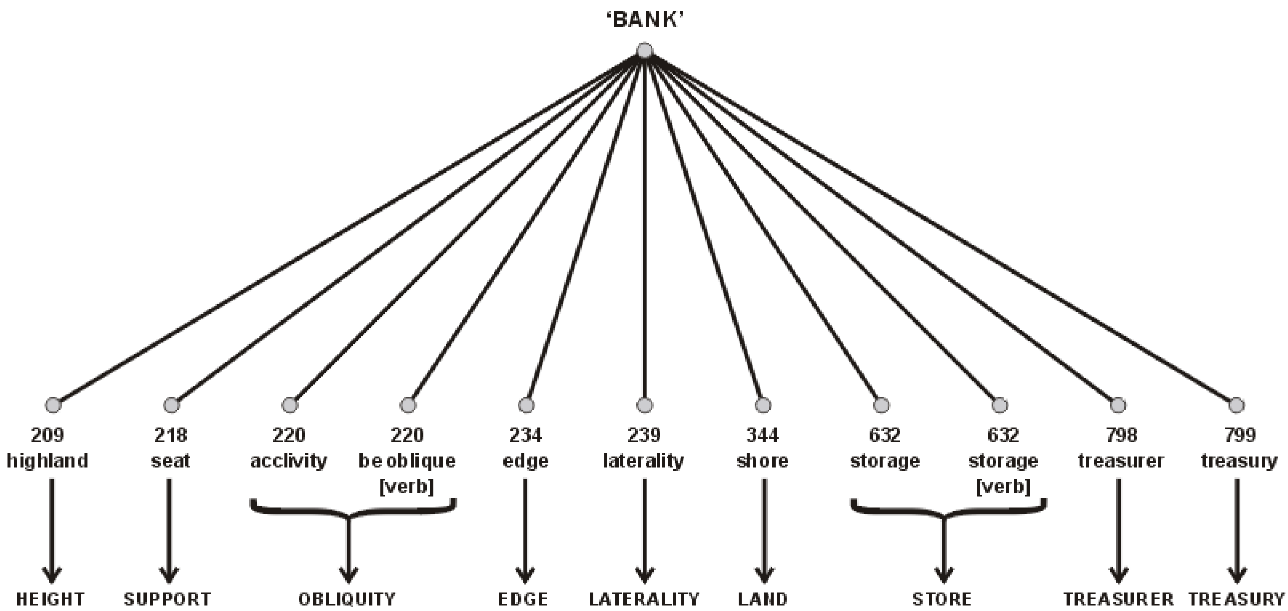
> "I want to pick up the relevant basic situation referring habits of a language in preference to its grammar."

Developed a context-dependent method of analysis:

1. Thesaurus with words grouped by areas of use.
2. Word "fans" radiating from each word type to each area of the thesaurus in which it occurs.
3. Dynamically generated combinations of fans for word tokens.

Recognized that analogy and metaphor are fundamental to the creation of novel uses of language, especially in science.

A Word Fan for "BANK"



Numbers and labels represent areas in Roget´s *Thesaurus*.

Method of Disambiguation

Example: ***up the steep bank*** **and** ***in the savings bank*****.**

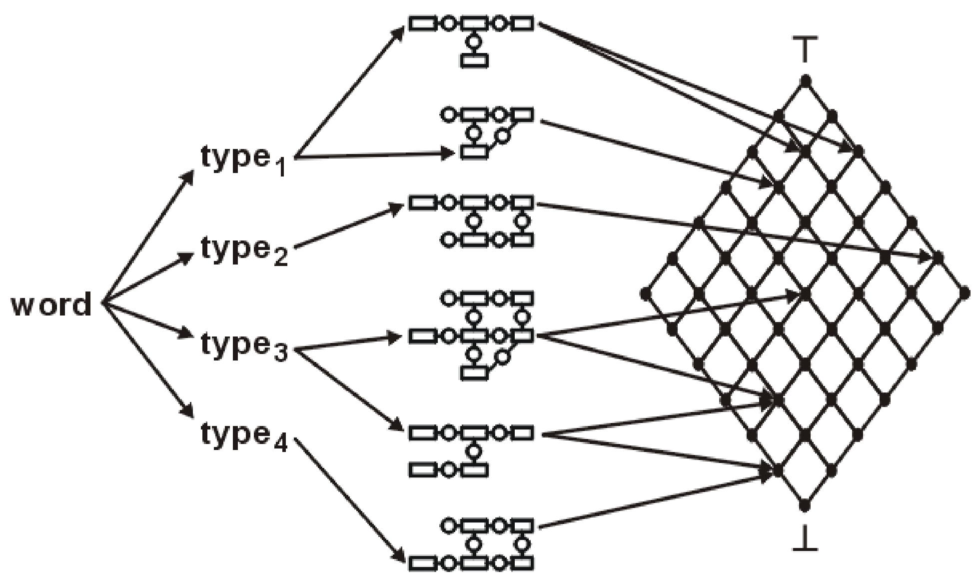
- **All the words except** ***the*** **have similar fans.**
- **Combinations of fans “pare down” the ambiguities “by retaining only the spokes that retain ideas which occur in each.”**
- **For this example,**
  - **— OBLIQUITY 220 is common to 'STEEP' and 'BANK'.**
  - **— STORE 632 and TREASURY 799 are common to 'SAVINGS' and 'BANK'.**

Formalizing Wittgenstein´s Language Games

Organize the ontology as a lattice of all possible theories that may be expressed with the chosen version(s) of logic:

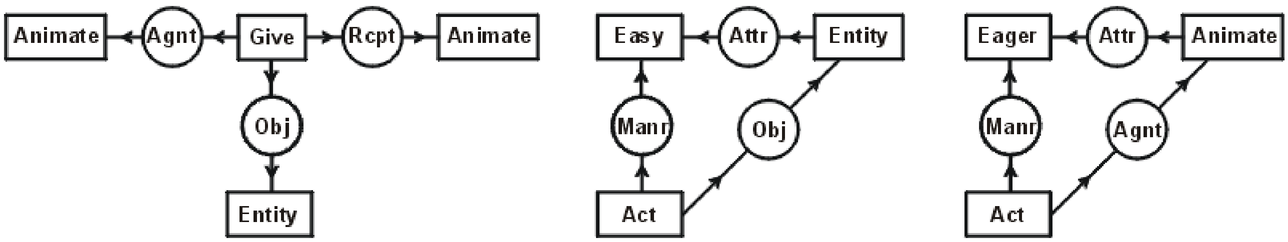
- **The universal theory at the top contains nothing but tautologies, which are true in all possible language games.**
- **General theories near the top are “sufficiently vague” to characterize a wide range of language games.**
- **Specialized theories at lower levels are sufficiently “narrow” to be “pretty precise and fairly certain” for more specialized games.**
- **Map the words of language to types of entities and relations.**
- **Map the types to *canonical graphs*, which show the expected patterns of usage.**
- **Map the canonical graphs to the theories in which they occur.**

Mapping Word Fans to a Lattice of Theories



words → types → canonical graphs → theories

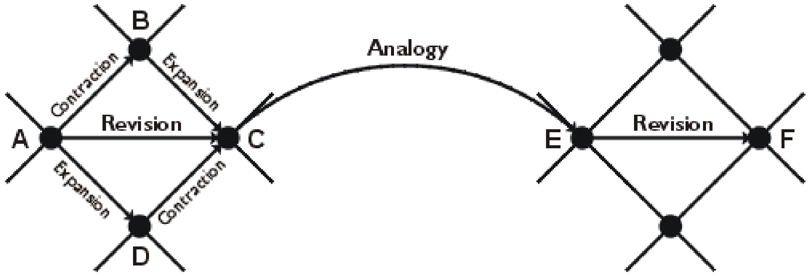
Canonical Graphs



Canonical graphs for the types Give, Easy, and Eager.

- **Encode the expected patterns of concepts and relations.**
- **Determine appropriate theories for each type.**
- **But canonical graphs for microsenses may become arbitrarily complex.**
- **Consider "*easy to please person*" vs. "*easy to read book*" vs. "*easy to drive car*".**
- **Specialized microsenses may add arbitrary amounts of detail to the canonical patterns.**

Navigating the Lattice of Theories



All methods of nonmonotonic reasoning may be viewed as strategies for finding a path through the lattice to a preferred theory.

- **Three AGM operators for theory revision: contraction, expansion, and revision.**
- **Analogy as a method of systematically relabeling concepts and relations.**
- **In learning, new information triggers a search through the lattice to find a new theory that is simpler or more accurate.**

Why is the Lattice Infinite?

Because it is open ended: there is always room for one more.

Like the integers, you only "implement" as many or as few of the members as you need or want:

- **The lattice is a set with an associated algebra of operations.**
- **It guarantees that whenever a new theory is defined, learned, or discovered, there will be a place for it.**
- **All the microtheories of Cyc have a place in the lattice.**
- **All the ontologies that have ever been defined or ever will be defined, whether good, bad, or indifferent, have a place in the lattice.**
- **The position in the lattice shows exactly how any new theory is related to any other theory as a generalization, a specialization, a sibling, or a distant cousin.**
- **The four theory-revision operators define paths for moving through the lattice during any process of learning or reasoning.**

The lattice is a mathematical structure that unifies many reasoning methods, shows how they are related to one another, and suggests new methods that may be simpler, more general, or more systematic than the older methods.

For a brief introduction to lattice theory, see
http://www.jfsowa.com/logic/math.htm#Lattice

Learning a New Theory

Initial facts form a very specialized theory with six axioms:

Tweety is a bird. **Tweety flies.**
Daffy is a bird. **Daffy flies.**
Hooty is a bird. **Hooty flies.**

Three possible generalizations that can reduce the number of axioms from six to four:

Every bird flies.

Every flying thing is a bird.

For every *x*, *x* is a bird if and only if *x* flies.

Heuristics give a slight preference for "Every bird flies."

But without more information, the other options cannot be ruled out.

New Information Triggers Theory Revision

Observation:

Vampy is not a bird.   Vampy flies.

This observation rules out two possible generalizations, leaving

Every bird flies.

Another observation:

Tux is a penguin.   Tux is a bird.   Tux does not fly.

This observation restricts the universal quantifier:

Every bird that is not a penguin flies.

Conclusions

Natural languages can express anything from a vague initial idea to a precise, final specification.

Formal theories can never be vague.

- **But they can be underspecified,**
- **And they can be organized to facilitate revision and reuse.**

Recommendations:

- **Emulate the flexibility of natural languages.**
- **Design formal systems to support multiple language games.**
- **Emphasize interoperability on local, task-oriented approaches.**

Alfred North Whitehead: "We must be systematic, but we should keep our systems open."

Suggested Readings

Two talks that discuss issues of ambiguity and vagueness from different, but related points of view:

http://www.jfsowa.com/talks/cmapping.pdf

http://www.jfsowa.com/talks/lgames.pdf

A collection of papers for a "Guided Tour of Ontology":

http://www.jfsowa.com/ontology/guided.htm

Two papers that elaborate issues discussed in this talk:

http://www.jfsowa.com/pubs/dynonto.htm

http://www.jfsowa.com/pubs/lgsema.pdf