Language Games, Natural and Artificial

John F. Sowa

VivoMind Intelligence, Inc.

Presented at FLAIRS-06 Conference, Melbourne, Florida

11 May 2006

Dilemma

Problem:

> **Natural languages are vague, ambiguous, and hard to process by computer.**

Solution proposed by Frege, Russell, and Carnap:

> **Replace NLs with precise, unambiguous formal languages.**

Result:

> **People find formal languages hard to learn.**
>
> **And when people use those formal languages, what they say so precisely often has no relationship to what they intended.**

Why is language ambiguous?

Three requirements for natural languages:

1. **Learnable by infants without any formal training.**
   ∴ **Finite vocabulary**
2. **Express anything anyone would ever need to say.**
   ∴ **Infinite extensibility**
3. **Accommodate human breathing rate and short-term memory.**
   ∴ **Short phrases**

Result:

The same stock of words is reused for everything.

And lexical ambiguity is inevitable.

Why is language vague?

Lord Kelvin:

Better a rough answer to the right question, than an exact answer to the wrong question.

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:

Frege compares a concept to an area and says that an area with vague boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. — But is it senseless to say: "Stand roughly there"?

Convergence of Opinions

Peirce:

> **"It is not so difficult to be pretty precise and fairly certain at once about a very narrow subject."**

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.**

In effect, a semantically restricted sublanguage or language game is a formal language that happens to use the syntax and vocabulary of a natural language.

The challenge is to find a way of reconciling the need for flexibility, which leads to ambiguity and vagueness, with the need for precision in computer implementations.

Microsenses

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

Most microsenses result from the complexity of the world, not from the nature of natural languages.

The word "car" —

- **1900: A horeseless 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: A Computerized Autonomous Rover.**

The word "file" in the area of operating systems —

- **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.**

Even for computer systems, the formal definitions of the terminology may change with every release and patch to the hardware or software.

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.”**

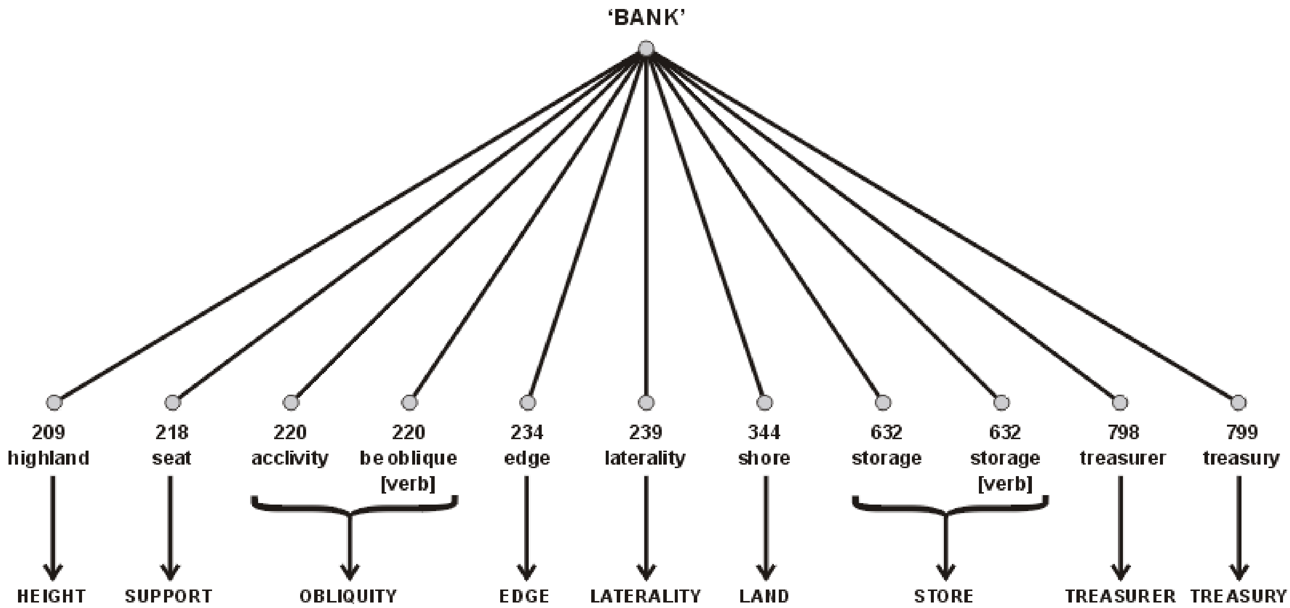
- **Developed a context-dependent method of analysis:**

  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.**

- **Identified areas of the thesaurus with different language games.**
- **Recognized that analogy and metaphor are fundamental to the creation of novel uses of language, especially in science**

Masterman developed these systems in the late 1950s and '60s, and for the earliest systems, her only computing facilities were punched cards and sorting machines. Yet her methods demonstrate the feasibility of a semantic approach that is in sharp contrast with the purely syntactic approach that Chomsky developed at about the same time.

A Word Fan for "BANK"



Numbers and labels represent areas of word usage in Roget's *Thesaurus*.

Masterman identified the areas of the thesaurus with Wittgenstein's language games.

By combining the fans for various words in a phrase, Masterman's method could select the common area of use in order to resolve ambiguities.

Method of Disambiguation

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

- **All the words except “*the*” have similar fans.**
- **Combinations of fans resolve the ambiguities by retaining only the spokes of the fans 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'.**

This approach should be distinguished from the Katz-Fodor method of using artificial features plus an informal phrase called a “distinguisher”.

Masterman's technique uses only words and families of words that actually occur in English — not abstract or artificial markers, features, or categories.

The avoidance of abstract features makes this approach a likely candidate for a cognitive theory, because it does not require an infant to learn anything other than words and patterns of words that are actually spoken. There is no need to assume any kind of innate ideas or syntactic categories, as in Chomsky's theory of Universal Grammar.

However, more work is needed to show how any such method could be linked to formal logic, if it is used as a front-end to a computer system.

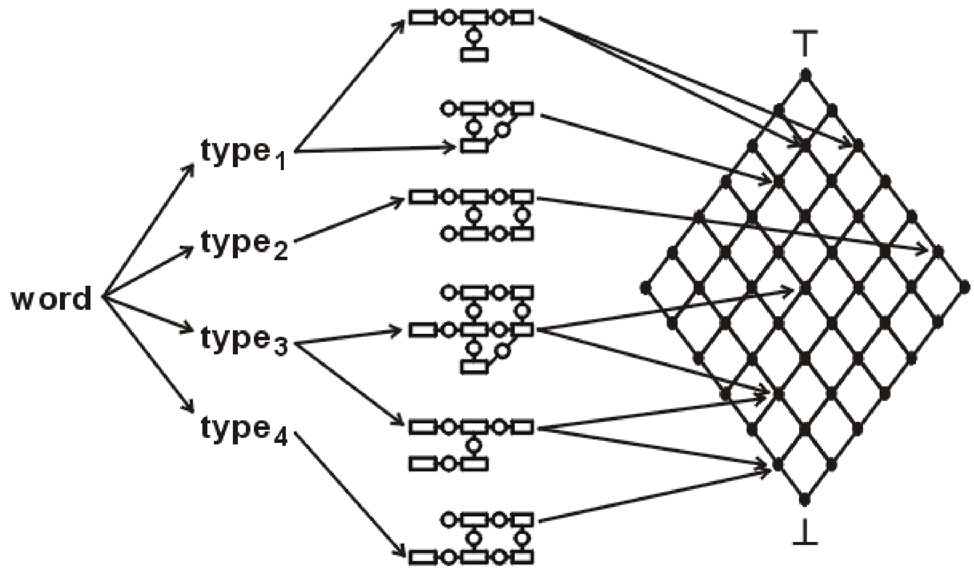
Relating Natural Language to Formalisms

- If we can't force natural languages into a computer mold, how can we adapt computer systems to NLs?
- First abandon the goal of a single universal ontology.
- Lenat tried that for Cyc, but they switched to a system with several thousand *microtheories*.
- Develop a dynamic method for finding, or adapting an appropriate theory for each language game.

Recommendation:

- Group the microsenses into general concept types.
- Define a theory or microtheory to distinguish different subtypes for each microsense.
- The collection of all microtheories corresponds to Wittgenstein's totality of all the language games that constitute a language.

Mapping Word Fans to a Lattice of Theories



words → concept types → canonical graphs → theories

On the left is a word fan linked to four concept types, which represent four different areas of use.

Each type is linked to one or more canonical graphs, each of which represents a characteristic pattern of concepts and relations in which that type may be used.

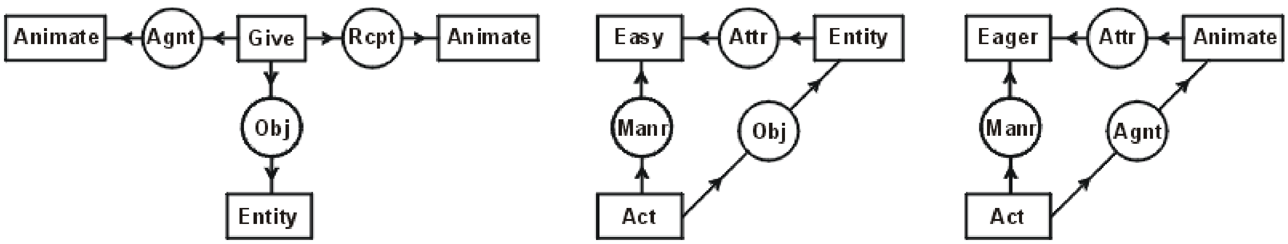
Each canonical graph is linked to one or more theories, each of which specifies the details of some specially defined microsense.

On the right is a lattice of all possible theories expressible in some given version of logic.

Since the lattice is infinite, only finitely many theories are ever used, but there is no shortage of possibilities in the lattice.

The words may be ambiguous or vague, but the theory for any sense or microsense can be as precise as necessary for the associated language game.

Canonical Graphs



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

For each type or subtype, the canonical graphs encode the expected patterns of concepts and relations and select appropriate theories.

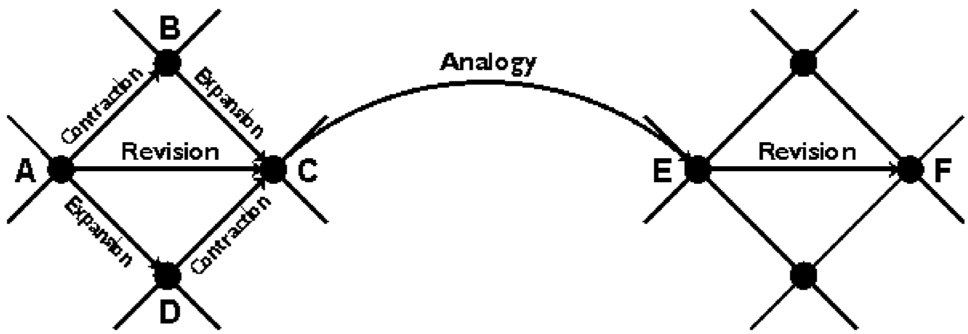
Note that that canonical graph for Give represents the kind of information that would be typical of a case frame: it says that the agent (Agnt) must be Animate, the recipient (Rcpt) must also be Animate, but the object (Obj) may be any kind of Entity – it might, in fact, be an action, as in giving a kiss, a kick, or a bath.

But the canonical graphs for Easy and Eager show the advantages of graphs over frames: they can include cross links, cycles, and even nested contexts.

The graph in the middle shows that the Entity whose attribute (Attr) is Easy is the object of the Act, whose manner (Manr) is Easy, but the graph on the right shows that the Animate being whose attribute is Eager is the agent of the Act whose manner is Easy.

These graphs can distinguish the relations for the sentences “*Bob is easy to please*” vs. “*Bob is eager to please.*” They also permit the phrase “easy to read book,” but not “*eager to read book*.” Canonical graphs for micro-senses may contain any details necessary to distinguish phrases such as “*easy to please person”* vs. “*easy to read book”* vs. “*easy to drive car”*.

Navigating the Lattice of Theories



Three AGM operators for theory revision: contraction, expansion, and revision.

Analogy as a method of systematically relabeling concepts and relations.

Define paths through the lattice to accommodate all major or minor revisions, combinations, or transformations of theories.

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."

Related Readings

A paper, “The Challenge of Knowledge Soup”:
    http://www.jfsowa.com/pubs/challenge.pdf

A paper, “Ontology, Metadata, and Semiotics”:
    http://www.jfsowa.com/ontology/ontometa.htm

A paper, “Categorization in Cognitive Computer Science”:
    http://www.jfsowa.com/pubs/cogcat.htm

A paper, “Concepts in the Lexicon”:
    http://www.jfsowa.com/ontology/lexicon.htm

A lecture, “Concept Mapping”:
    http://www.jfsowa.com/talks/cmapping.pdf

A tutorial on math and logic:
    http://www.jfsowa.com/logic/math.htm

The web site for the draft ISO standard for Common Logic:
    http://cl.tamu.edu