13 Three Meanings of "Semantics"

Drew McDermott drew.mcdermott@yale.edu

2015-11-20, 11-30, 2016-10-28 — Yale CS470/570

Ironically, "semantics" is ambiguous, and often trips people up. There are three[1] possible meanings:

1. The relation between representations and the "world." Sometimes, as in mathematical logic, the world is abstract, but in AI we're interested in the relation between representations and the actual world, inhabited by people, buildings, substances, institutions, currency, contracts, etc.

2. The relation between representations of formal statements in mathematical logic and representations of alternative "worlds" (call these *R-worlds*). It is often useful for algorithms to manipulate R-worlds. Such methods are often called *semantic*, even though R-worlds are just as syntactic as the formulas linked to them.

3. The relation between a natural-language sentence and its "internal representation." This is the hypothetical object that (a) represents a reading (or set of readings) of the sentence, and (b) is or might be useful computationally in the human mind. Many linguists accept the existence of the *logical form* of a sentence, which fills role (a) and explains linguistic data. Fewer of them are willing to entertain the idea that logical form fills role (b).

As far as meaning 1 is concerned, introductory textbooks, and these lectures, talk mainly about the *formal structure* of mappings, or *interpretations* of symbol structures as entities in the world. One is entitled to ask, *How* do these structures in the computer or the brain "reach out and touch" the real-world objects they denote?

The paradigm case is how a robot keeps track of the objects in its vicinity. Its sensory systems extract object descriptions from sensor data (e.g., shapes from images), and match them to descriptions of objects that the robot is

[1] Sometimes, especially in older papers, one sees the adjective "semantic" used in a fourth way. The author's program or algorithm is described as "semantic," in contrast to previous efforts, which have been "syntactic," meaning the new approach is clever, deep, and expensive where the old one is stupid, shallow, and cheap. Computational cycles are always dropping in price, so paying a bit more to be clever seems like an obvious win. However, over time cycles have gotten to be so cheap that being stupid and shallow — but uniform — may beat approaches that do something clever but harder to parallelize. So "semantic" is less often seen as a term of praise.

tracking (because it has some "interest" in them, or reason to keep track of them; for example, to collect them, or keep from bumping into them). The idea is that an internal name, `object81047`, is generated when an object is first encountered, and, if all goes well, is continually reused as the object is tracked through a series of sensory events. When the object is out of range, the name may be retired, or, if this is a really advanced robot, be reused if the object is encountered again.

This is the sort of machinery that lies behind semantic equations of the sort we casually write. For example, we might write $I(\texttt{object81047}) =$ "desk in room 508"; this is likely to be true only if the robot's software works as described most of the time, which is not easy to insure, as decades of robotics research attests.

Starting from this easy case, we can work our way up to some really hard cases.

- Internal representations of people you've never met and objects you've never seen, such as (the internal equivalents of) "Amy Adams" or the "Taj Mahal." You do encounter *pictures* (sometimes moving pictures) of these entities, and textual or spoken information about them. Your beliefs regarding these people and things may be true or false, but they really are *about* the people. If you believe "Amy Adams secretly loves me," you are probably mentally ill, but nonetheless your semantic machinery is working properly, unless you have Amy Adams confused with Abigail Adams.

- Representations of people who died long ago, such as Abigail Adams. If you are thinking of the wife of the second President of the United States, *how* are you doing that? No one alive has ever met this woman, and yet many people succeed in having an internal symbol structure that tracks her. Of course, we also succeed in referring to plenty of non-human objects that have ceased to exist, such as the Colossus of Rhodes.

- Possible objects that may never exist, such as the first of your descendents to be chosen as Secretary-General of the United Nations.

- Reference to predicates and categories, such as "polka-dotted" and "ostrich." A robot doesn't track a predicate the way it does an object, and yet the paradigm cases are easy to describe: when the robot gets close enough to an object satisfying the predicate, it classifies it as satisfying it, and it does so reliably, rarely classifying a non-P as

a P or vice versa. Or perhaps it waits until it has a reason to decide whether an object satisfies the predicate and then investigates the matter, possibly moving around or deploying sensors in a different configuration. As predicates get more abstract, they get further from the paradigm. Here are some examples, in no particular order: "flustered," "trustworthy," "flexible," "crystalline," "stable," "true," "provable."

(What follows is a couple of issues that really belong to "knowledge representation" (KR), but they have psychosemantic repercussions; there's no way for a program to have beliefs about an aspect of the world unless its representations are sensitive to that aspect.[2]

- Representations of substances, especially liquids, and "objects" made out of liquids, such as the beer in my glass or the Mississippi River.

- Representations of events and event types, such as "World War II" (particular event) and "financial panic" (event type). Some of these are in the future, like "0.7m rise in sea level (due to global warming and melting of glaciers)."[3] Should we refer to the rise in sea level as an event type, which could occur differently in different futures, or is it recognizably the same Event in all of them? In general, this is called the "cross-world identification problem." If I want to test a statement about an object with respect to some possible world or set of possible worlds, how do I "find" that object? What if there are several entities with an equal claim to be that object?

That's all we're going to say about psychosemantics in CS470/570. It's a thorny philosophical issue, but not really a technical one. We're going to need the other two senses of "semantics," the second in connection with inference procedures, the third as our default meaning for the phrase "natural-language semantics."

So imagine we're translating into an internal language. SLP does a good job of covering the basic techniques that have been developed for building such languages.

[2] Until Google comes up with some deep neural net tha captures the aspect in a vector space of some kind.

[3] Okay, perhaps this one is already underway; which raises its own problems. Suppose I'm halfway down the hall to a meeting, so "I go to the meeting" is underway. Then I change my mind and decide not to attend, turn around, and go back to my office. Was that event underway, even though it never happened? But I suppose the chances of humanity reversing course on climate change are very low.