# A rule-to-rule semantics for a simple language

Jeff Speaks
PHIL 43916

September 10, 2014



## 1 Sentences, names, and $V_i$'s

What we want to our theory to tell us the truth conditions of sentences; that is, the conditions $v$ under which $[\![S]\!]^v$=1. Given this, we know that $[\![S]\!]$ is a truth-value – 1 or 0. But what are, for example, $[\![N]\!]$ and $[\![V_i]\!]$?

Keep in mind: what we really want for names and intransitive verbs, just as for sentences, is not their semantic values, but their semantic values relative to some state of the world. (Can you see why we should want this, given our motivation for pursuing a semantic theory in the first place?) But it will ease exposition to simplify by ignoring this fact for the moment, and imagining that we are just trying to derive the semantic values of sentences. We un-simplify in §5 below.

The semantic value of a name is the object for which the name stands. So, for example, $[\![$Pavarotti$]\!]$=Pavarotti, $[\![$Sophia Loren$]\!]$=Sophia Loren, and so on.

The semantic value of an intransitive verb will be a set of individuals. So $[\![$is boring$]\!]$ will be the set of individuals that are boring, $[\![$is cute$]\!]$ will be the set of cute individuals, and so on. We will refer to the set of boring individuals using the notation

$\{x$: $x$ is boring$\}$

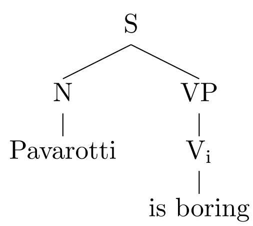
Intuitively: the set of all the $x$'s which are such that $x$ is boring.

Note that so far we have said what the semantic values of sentences, names, and intransitive verbs are, but have not provided a rule for determining the semantic value of a sentence consisting of a N and a $V_i$ on the basis of the semantic values of the latter. That we can do using (to follow the numbering in the text) rules 31 (a) and (e). Let's start with the first of these:

> (a) $\llbracket[_S \text{ N VP}]\rrbracket = 1$ iff $\llbracket \text{N} \rrbracket \in \llbracket \text{VP} \rrbracket$ and 0 otherwise

Let's pause on this rule for a second. What does it say? Consider some examples of sentences consisting of names and $V_i$'s. What does it indicate about the truth conditions of sentences of this sort?

But note that rule (a) by itself, plus the above remarks about the semantic values of names, sentences, and $V_i$'s, does not tell us how to derive the truth or falsity of any sentence. Consider our tree diagram for 'Pavarotti is boring':




We know what the semantic values of the N and the $V_i$ are, and we know how to figure out the semantic value of the S once we have the semantic value of the N and the VP; but so far we have no way of determining the semantic value of the VP.

You might think that this is pretty obvious: surely, in this case, $\llbracket \text{VP} \rrbracket = \llbracket V_i \rrbracket$. And this is correct. But we need some rule to state clearly when we're allowed to 'pass up' the semantic value of an expression from a child node to its parent. That is the point of rule 31 (e):

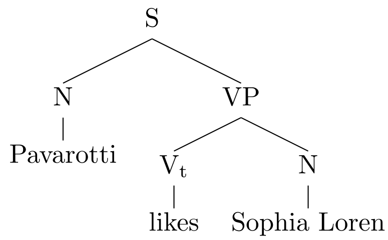
> (e) If A is a category and a is a lexical entry or category and $\Delta = [_A \text{ a}]$, then $\llbracket \Delta \rrbracket = \llbracket \text{a} \rrbracket$.

Here we use '$[_A$ a]' to mean 'the tree dominated by A, whose only child is a.' More generally, '$[_A$ b c]' means 'the tree dominated by A, whose only children are b and c.'

We can think of the process of determining the semantic value – i.e. truth value – of a sentence as working in steps. First, enter the semantic values of the leaves. Then, we consult the rules of our semantics to determine the semantic values of the parents of the leaf nodes, continuing to work from child to parent until we have assigned a semantic value – 1 or 0 – to the S.

## 2 Transitive verbs

Now consider a sentence like our example from last time,




This is still a sentence of the form $[_{\text{S}}$ N VP$]$, so rule (a) above should apply. This means that, as above, the semantic value of our VP must be a set – in this case, it will be

$$\{\text{x: x likes Sophia Loren}\}$$

One idea would be to simply add this fact about $[\![\text{likes Sophia Loren}]\!]$ to our semantic theory. Why might this be a bad idea?

Better would be to give a rule for determining $[\![\text{likes Sophia Loren}]\!]$ on the basis of $[\![\text{likes}]\!]$ and $[\![\text{Sophia Loren}]\!]$. We already know that $[\![\text{Sophia Loren}]\!]$=Sophia Loren. So our question is: what is $[\![\text{likes}]\!]$?

Remember that the semantic value of 'boring' was the set of boring things. So one might think, by extension, that the semantic value of 'likes' is a set of sets: the set of sets of things which are such that one likes the other:

$$[\![\text{likes}]\!] = \{\{\text{x,y}\}\text{: x likes y}\}$$

What would be wrong with this? (Keep in mind that if sets S1, S2 have the same members, then S1=S2; so, in particular, {a,b}={b,a}.)

Better to take the semantic value of an intransitive verb to be a set of ordered pairs, namely

$$[\![\text{likes}]\!] = \{\langle\text{x,y}\rangle\text{: x likes y}\}$$

This leaves open the possibility that ⟨Pavarotti, Sophia Loren⟩ will be an element of $[\![\text{likes}]\!]$, whereas ⟨Sophia Loren, Pavarotti⟩ will not.

But now we are in a situation like the one above: we have an assignment of semantic values to 'likes' and 'Sophia Loren,' but we need an extra rule to tell us how to get from these semantic values to the semantic value of the complex VP 'likes Sophia Loren.'

That is the point of rule 31 (d):

$$\text{(d) } [\![[_{\text{VP}}\ \text{V}_{\text{t}}\ \text{N}]]\!] = \{\text{x: } \langle \text{x, } [\![\text{N}]\!]\rangle \in [\![\text{V}_t]\!]\}$$

What does this say? What does it imply about the case of 'likes Sophia Loren'?

## 3 Sentence operators and connectives

We're almost done with the semantics for our simple language: all that's left is to explain the semantic values of sentence operators and connectives.

Consider first our lone sentence operator – our lone member of the category neg – 'it is not the case that.' We know that our languages permits sentences of the form $[_S$ neg S], so $[\![$neg$]\!]$ must be something which combines with a truth-value – which is $[\![$S$]\!]$ – to give us a truth-value.

A natural choice for $[\![$neg$]\!]$ is a *function*. A function is a relation between a set of inputs – the function's arguments – and a set of outputs – its values – which has the property that any argument is related to exactly one value.

A familiar example of a function is addition. Its arguments are pairs of numbers, and its values are individual numbers – the sum of the arguments. Addition is a function, rather than some other sort of relation, because it is never the case that, for any a, b, a+b=c and a+b=d for c≠d.

What sorts of things should the arguments and values of $[\![$it is not the case that$]\!]$ be? Which arguments should get mapped to which values?

We write this as:

$[\![$it is not the case that$]\!]$ = $[1 \rightarrow 0, 0 \rightarrow 1]$

Now, as before, in addition to specifying the semantic value of $[\![$it is not the case that$]\!]$, we need an extra rule telling us how to compute the semantic value of $[_S$ It is not the case that S] – or, more generally, $[_S$ neg S] – on the basis of $[\![$neg$]\!]$ and $[\![$S$]\!]$. That is rule 31 (c):

(c) $[\![[_S$ neg S$]]\!]$ = $[\![$neg$]\!]$ ($[\![$S$]\!]$)

This follows the standard notation for functions, where we express the claim that 'function f applied to argument a has value v' as 'f(a)=v' – as in '+(2,3)=5.'

How would you extend this treatment of 'it is not the case that' to our two sentence connectives, 'and' and 'or'?

Since these two members of the category conj combine with two sentences to form a sentence, it is natural to treat them as functions from pairs of truth-values to truth-values. In particular:

$[\![$and$]\!]$ = $[\langle 1,1 \rangle \rightarrow 1$
$\langle 1,0 \rangle \rightarrow 0$
$\langle 0,1 \rangle \rightarrow 0$
$\langle 0,0 \rangle \rightarrow 0]$

$[\![\text{or}]\!] = [\langle 1,1\rangle \to 1$
$\langle 1,0\rangle \to 1$
$\langle 0,1\rangle \to 1$
$\langle 0,0\rangle \to 0]$

and we derive the semantic values of sentences involving a conj using rule 31 (b):

(b) $[\![[_S \text{ S1 conj S2}]]\!] = [\![\text{conj}]\!](\langle [\![\text{S1}]\!], [\![\text{S2}]\!]\rangle)$

It is worth pausing for a moment over the case of 'or.' It might seem that, whatever is true of our simple language, the semantic value given to 'or' can't possibly be the semantic value of the English word 'or.' For consider a sentence like 'Jim will go to bed early or Jim will fail the exam.' Surely this means that exactly one – not at least one – of the two sentences connected by 'or' is true.

This is a good case to bring up the distinction between what sentences mean and what speakers mean by using those sentences. What we're trying to capture is, in the first instance, facts about sentence meaning.

Some evidence that 'or' in English has $[\![\text{or}]\!]$ as its semantic value is given by the way that 'or' sentences behave as a part of more complex discourses. There is, for example, no contradiction in saying

> Jim will go to bed early or Jim will fail the exam – indeed, he's not very bright, so he might well do both.

And the sentence

> It is not the case that Jim will go to bed early or Jim will fail the exam.

seems to be false, not true, if Jim does both.

## 4 Some examples

Let's work through some examples, and try to derive the truth-values of some sentences using the rules of our semantic theory. To do this we will have to be clear about exactly what the semantic values of our $\text{V}_\text{i}$'s and $\text{V}_\text{t}$'s are; we know that $[\![\text{is boring}]\!]$ = the set of boring things, but we don't know what things are in that set. So let's suppose that:

$[\![\text{is boring}]\!] = \{\text{James Bond, Pavarotti}\}$
$[\![\text{is cute}]\!] = \{\text{Pavarotti}\}$
$[\![\text{likes}]\!] = \{\langle\text{Sophia Loren, Pavarotti}\rangle, \langle\text{James Bond, Pavarotti}\rangle\}$

And consider the following sentences:

Pavarotti is cute.
Sophia Loren is boring or James Bond is cute.
It is not the case that James Bond is boring and Pavarotti is cute.

In each case, our theory allows us to derive semantic values – truth-values – for the sentences on the basis of the semantic values of the simple terms, plus facts about how they are combined, plus our semantic rules for combining expressions to form complex expressions.

## 5 Relativizing to circumstances

So far, in introducing our semantic theory, I've suppressed the need to relativize semantic values to different circumstances of evaluation; I've been talking, e.g., about $[\![\text{Pavarotti is cute}]\!]$ but not $[\![\text{Pavarotti is cute}]\!]^v$.

It's now time to re-introduce this. For some expressions we've discussed, this makes no difference. Ignoring some complications to which we will return later, the semantic values of names and of connectives will be the same with respect to every circumstance of evaluation; for any v, $[\![\text{Pavarotti}]\!]^v$ = Pavarotti, and $[\![\text{neg}]\!]^v = [1{\rightarrow}0,\ 0{\rightarrow}1]$.

But this is not true of our $\text{V}_\text{i}$'s and $\text{V}_\text{t}$'s – can you see why?

However, the modification of their semantic values which this requires is, in one sense, not so great. Rather than the simple

$$[\![\text{is boring}]\!] = \{\text{x: x is boring}\}$$

we will now have

$$[\![\text{is boring}]\!]^v = \{\text{x: x is boring in v}\}$$

The important thing about this change, for our purposes, is that it now allows us to derive not just the truth-values of sentences of our language, but their truth-conditions – i.e., their truth-value with respect to different circumstances of evaluation. And this is important because, plausibly, this is what competent language users know about sentences they understand – not whether they are true or false, but the conditions under which they would be true or false.

This requires a modification of our semantic rules 31 (a)-(e) – in each case, simply replace every reference to a semantic value $[\![\text{x}]\!]$ with a relativized $[\![\text{x}]\!]^v$ and everything else remains the same. The relativized rules – which are the versions in the text – are:

(a) $[\![[_\text{S}\ \text{N VP}]]\!]^v = 1$ iff $[\![\text{N}]\!]^v \in [\![\text{VP}]\!]^v$ and 0 otherwise
(b) $[\![[_\text{S}\ \text{S1 conj S2}]]\!]^v = [\![\text{conj}]\!]^v(\langle[\![\text{S1}]\!]^v,\ [\![\text{S2}]\!]^v\rangle)$
(c) $[\![[_\text{S}\ \text{neg S}]]\!]^v = [\![\text{neg}]\!]^v\ ([\![\text{S}]\!]^v)$

(d) $[\![[_{\mathrm{VP}}\ \mathrm{V_t}\ \mathrm{N}]]\!]^v = \{\mathrm{x}\!: \langle \mathrm{x}, [\![\mathrm{N}]\!]^v\rangle \in [\![\mathrm{V_t}]\!]^v\}$

(e) If A is a category and a is a lexical entry or category and $\Delta = [_{\mathrm{A}}\ \mathrm{a}]$, then $[\![\Delta]\!]^v = [\![\mathrm{a}]\!]^v$.

Using these rules, let's derive $[\![\text{Pavarotti is boring}]\!]^v$.

Looking at 31 (a)-(e), you can see why in the text this theory is referred to as an example of 'rule-to-rule' semantics. 31 (a)-(e) mirror the syntactic rules 21 (a)-(e) of our language. Each of those syntactic rules gives one type of case when it is is possible in our language to grammatically combine expressions of two types. For any such case, we then need to add to our semantics a rule which tells us how, in cases of that type, the relevant semantic values combine to give us the semantic value of the complex expression. For each syntactic rule, we have a corresponding semantic rule.

## 6 Entailment and contradiction

Another benefit of our relativization of semantic values to circumstances is that it enables us to define entailment.

To a first approximation, one sentence S1 entails another sentence S2 if and only if, necessarily, if S1 is true, then S2 is true – or, to put the same point another way, S1 entails S2 if and only if the truth of S1 guarantees the truth of S2.

Often, just on the basis of understanding sentences, and without knowing whether either is true, we can see that one sentence entails another. For example, many have claimed that any competent speaker can see that if

> Pavarotti is boring and James Bond is cute.

is true, so must be

> Pavarotti is boring.

If we relativize semantic values to circumstances, then we can define entailment as a relation between individual sentences as follows:

> S1 entails S2 iff for all $v$, if $[\![\mathrm{S1}]\!]^v$=1, then $[\![\mathrm{S2}]\!]^v$=1.

In a related way, we can define the relation of contradiction between sentences:

> S1 contradicts S2 iff for all $v$, if $[\![\mathrm{S1}]\!]^v$=1, then $[\![\mathrm{S2}]\!]^v$=0.

And in many cases we can use our semantic theory to prove that one sentence entails (or contradicts) another. Consider the example above, about Pavarotti and James Bond. How, using our semantic rules, could you prove that the first of these sentences entails the other?

Next, try to prove that

> It is not the case that Pavarotti is boring or James Bond is cute.

(on one interpretation) contradicts

> Pavarotti is boring.

These facts about entailment (and contradiction) are connected to the question of how we can tell that whether a semantic theory for a language like English is correct. What, in semantics, is supposed to play the role of experimental results in physics? Many have thought that the answer is, at least in part, given by the following two tests:

- Competent speakers of a language know the truth conditions of sentences of their own language. The correct semantic theory should therefore assign truth conditions to those sentences which fit the beliefs of competent speakers.
- Competent speakers of a language know when one sentence of their language entails (or contradicts) another. A semantic theory should explain this ability by providing an explanation, in something like the above way, of these entailment relations.

In the end, we'll see that, plausibly, no theory can quite meet these tests – every theory makes some surprising claims about truth conditions, and no theory can explain every entailment. But these at least provide reasonable starting points for evaluating semantic theories.