The Logic Of Semantic Content

Simon D'Alfonso

March 16, 2010

If σ is an instance of semantic content as understood in this paper, then:

1. σ consists of one or more data;
2. the data in σ are well-formed;
3. the well-formed data in σ are meaningful.

So data are the stuff of which semantic content is made; semantic content cannot be dataless, but, in the simplest case, it can consist of a single datum. A general definition of a datum is:

> A datum is a putative fact regarding some difference or lack of uniformity within some context.[1]

Some examples will help to clarify the gist of this definition. Take a single sheet of unmarked white paper. It is an example of complete uniformity; each unit of the paper's surface is the same as every other unit.[2] As it is, there is no datum associated with this sheet. If a black marker were used to place a black dot in the middle of the sheet, then there would be a lack of uniformity. The white background plus the black dot would constitute the datum.[3]

Or as another example, consider a unary alphabet, consisting of the symbol 0. Any source that continuously emits symbols from this alphabet is not emitting data, for there is no lack of uniformity in its output. However, if the alphabet were expanded to include the symbol 1 as well as the symbol 0, then it would be possible for the source to emit data, by using both instances of the 0 symbol and instances of the 1 symbol.

With condition 2, 'well-formed' means that the data are composed according to the rules (syntax) governing the chosen system, code or language being analysed. Syntax here is to be understood generally, not just linguistically, as what determines the form, construction, composition or structuring of something. The string 'the an two green four cat !?down downx' is not well-formed in accordance with the rules of the English language, so therefore cannot be an instance of semantic content in the English language. Or, to

[1]See http://plato.stanford.edu/entries/information-semantic/#1.3 for discussion of this definition

[2]Whatever a unit might be measured in, pixels, millimetres, etc. Also, when comparing units, the attribution of sameness is based only on a certain property, namely that each unit is in its original state of unmarked whiteness. In certain ways each unit might differ. For example, each unit is in a different part of the sheet of paper, some units might be smoother than others. In this case, we are talking about the state of each unit in terms of its marking.

[3]This involves the notion of Taxonomic Neutrality. A datum is a relational entity. Neither of these two *relata*, the black dot or the white background, is the datum. Rather both, along with the fundamental relation of inequality between the dot and the background constitute the datum.

take another example, the string '$A \neg B$' is not well-formed in accordance with the rules of the language of propositional logic, so therefore cannot be an instance of semantic content in propositional logic.

With condition 3, 'meaningful' means that the well-formed data must comply with the meanings (semantics) of the chosen system, code or language in question. For example, the well-formed string 'Colourless green ideas sleep furiously' cannot be semantic content in the English language because we may say (without getting into a debate about theories of meaning) that it is meaningless; it does not correspond to anything. Finally, an example of a string which fulfills conditions 1, 2 and 3 is 'The native grass grew nicely in spring'. Following are some cases of semantic content.

- A map of Europe contains the factual information that Germany is north of Italy, in the language of cartography. The data that this information is made of is identified with the sheet of paper on which the map is printed plus the various markings on the page. This data is well-formed; among other things, the North-South-East-West coordinates are correctly positioned and no countries are marked as overlapping each other. Finally, this data is meaningful. Each part of the paper, contained in a thick black line and shaded in a certain colour corresponds or refers to a country. Thin blue lines mean rivers, etc.

- A person's nod contains the factual information that they are in agreement, in certain human body languages. The data that this information is made of is indentified with the variation in head position. This data is well-formed; head movement is a legitimate expression in the language. This data is also meaningful; this particular expression means 'yes' or 'positive'.

- The content of an Encyclopaedia Britannica entry on Italy will contain the information that Rome is the capital of Italy, in the language of English. The data that this information is made of is indentified with the varied string of English alphabet symbols that constitute the entry. This data is well-formed as it accords with the syntax of the English language. It is also meaningful to an English language reader.

- The content of a book which says that the earth is flat is false semantic content or misinformation. The data that this misinformation is made of is indentified with the varied string of English alphabet symbols that constitute the content. his data is well-formed as it accords with the syntax of the English language. It is also meaningful to an English language reader.

Ultimately, these various forms of semantic information are reducible to propositional form, or propositional expression. If p is factual information, then it can be expressed in the form 'the information that p'. This leads to an identification of information with propositions.

Truth and falsity supervene on semantic content, which is neutral with regards to these alethic properties. Basically, semantic content can be either true or false. Since semantic content is propositional it can be dealt with using a propositional logic. For example, if p is true semantic content and q is true semantic content, then $p \wedge q$ is true semantic content. Or if p is false semantic content, then $p \wedge q$ is false semantic content.

As well as these 'internal' alethic valuations, we can also say things like if p is semantic content and q is semantic content, then

- it is true that $p \wedge q$ is semantic content
- it is true that $p \vee q$ is semantic content
- it is true that $\neg p$ is semantic content

But what happens when semantic content is connected with data that is not semantic content? For example, what is the status of the conjunction:

'Colourless green ideas sleep furiously' $\wedge$ 'The native grass grew nicely in spring'

I shall now turn to discussion of a logic which can formally reason about these things.

1 Bochvar's 3-valued logic

Bochvar's 3-valued logic can be appropriated as a logic to reason about semantic content. This system introduces the intermediate third value * in addition to the classical values t (true) and f (false). The idea of this logic is to "avoid logical paradoxes such as Russell's and Grelling's by declaring the crucial sentences involving them to be meaningless" [1, p. 75]. The truth tables for the *internal* connectives of this logic are[4]:

$f_\neg$	
t	f
$*$	$*$
f	t

$f_\wedge$	t	$*$	f
t	t	$*$	f
$*$	$*$	$*$	$*$
f	f	$*$	f

$f_\vee$	t	$*$	f
t	t	$*$	t
$*$	$*$	$*$	$*$
f	t	$*$	f

$f_\supset$	t	$*$	f
t	t	$*$	f
$*$	$*$	$*$	$*$
f	t	$*$	t

$f_\equiv$	t	$*$	f
t	t	$*$	f
$*$	$*$	$*$	$*$
f	f	$*$	t

As can be seen, when only the values t and t are involved, these connectives are the same as their classical counterparts. When a * is involved, the result yields a valuation of *; a meaningless constituent proposition 'infects' the compound proposition of which it is a part. With t as the only designated value, this system has no tautologies. An 'assertion operator' Ap can be added to represent the 'external assertion' of a proposition p. Ap is the assertion 'p is true' in a two-valued metalanguage. So Ap is t if p is true, otherwise it is f:

$f_\neg$	
t	t
$*$	f
f	f

[4] these are the same as the Weak Kleene Logic

Using the assertion operator, one can define the 'external connectives', which always take the values t or f. For each external connective, they are defined by applying the assertion operator to the arguments of an internal connective. So

- the external negation $\sim$ is defined as $\neg$Ap
- the external conjunction $p \otimes q$ is defined as $Ap \wedge Aq$
- the external conjunction $p \oplus q$ is defined as $Ap \vee Aq$
- the external conditional $p \rightarrow q$ is defined as $Ap \supset Aq$
- the external biconditional $p \leftrightarrow q$ is defined as $Ap \equiv Aq$

$f_\sim$	
t	f
$*$	t
f	t

$f_\otimes$	t	$*$	f
t	t	f	f
$*$	f	f	f
f	f	f	f

$f_\oplus$	t	$*$	f
t	t	t	t
$*$	t	f	f
f	t	f	f

$f_\rightarrow$	t	$*$	f
t	t	f	f
$*$	t	t	t
f	t	t	t

$f_\leftrightarrow$	t	$*$	f
t	t	f	f
$*$	f	t	t
f	f	t	t

2 Reasoning About Semantic Content

So how can this system help us in reasoning about semantic content? We start off by letting the usual propositional variables range over data. We shall now call these data variables. If a data variable is assigned the value t or the value f, then it qualifies as semantic content, since only data that is also semantic content can be alethically qualifiable. If a data variable is assigned the value $*$, then it fails to be semantic content; at the least this is because it is meaningless and at the most this is because it is not well-formed.

Just like propositions can be used as arguments of connectives to form compound propositions, data can also be built up. We shall refer to the equivalent of an atomic proposition as a datum. So instead of saying 'the atomic proposition p', we shall say 'the datum p'. Desirable properties of semantic in relation to the connectives are given in the following list:

- A is semantic content iff $\neg A$ is semantic content
- A is semantic content and B is semantic content iff $A \wedge B$ is semantic content
- A is semantic content and B is semantic content iff $A \wedge B$ is semantic content
- A is semantic content and B is semantic content iff $A \supset B$ is semantic content

- A is semantic content and B is semantic content iff $A \equiv B$ is semantic content

Take the following items of data based on strings composed of characters in the English language. The data variables that they will be represented by in examples are on the left of the $=_{df}$ and the strings of data on the right:

- $p =_{df}$ kangaroos are the largest surviving marsupial
- $q =_{df}$ the Eastern Grey Kangaroo is the largest type of kangaroo
- $r =_{df}$ furiously sleep ideas kangaroo colourless

Now, $v(p) = t$, because kangaroos are the largest surviving marsupial. $v(q) = f$ because it is actually the Red Kangaroo that is the largest type of kangaroo. $v(r) = *$, because this is a meaningless proposition, or meaningless datum. We can plug these data variables into some connectives:

- $v(p \wedge q) = f$ - the data $p \wedge q$ is false semantic content.
- $v(p \vee q) = t$ - the data $p \vee q$ is true semantic content.
- $v(p \vee r) = *$ - the data $p \vee r$ is meaningless.

The basic Bochvar connectives capture this reasoning. Further to this, 3 external one-place operators can be added.

The first one, I, is such that Ip is to be read as 'p is a piece of information'. If p is information, then it is true semantic content. If p is false semantic content or meaningless data, then it is not information. This essentially gives an operator that is the same as Bochvar's assertion operator. The second operator, M, is such that Mp is to be read as 'p is a piece of misinformation'. If p is misinformation, then it is false semantic content false. If p is true semantic content or meaningless data, then it is not misinformation. The third operator, S, is such that Sp is to be read as 'p is a piece of semantic content'. If p is semantic content, then it is true or false. If p is meaningless, then it is not semantic content. All this gives the following:

f_{I}	
t	t
$*$	f
f	f

f_{M}	
t	f
$*$	f
f	t

f_{S}	
t	t
$*$	f
f	t

From here we can easily see that the S operator formally captures the above list of desirable properties regarding semantic content and connectives.

A	B	SA	SB	S$\neg A$	S$\neg A \leftrightarrow$ SA	S$(A \wedge B)$	S$(A \vee B)$	S$(A \supset B)$	S$(A \equiv B)$
t	t	t	t	t	t	t	t	t	t
t	$*$	t	f	t	t	f	f	f	f
t	f	t	t	t	t	t	t	t	t
$*$	t	f	t	f	t	f	f	f	f
$*$	$*$	f	f	f	t	f	f	f	f
$*$	f	f	t	f	t	f	f	f	f
f	t	t	t	t	t	t	t	t	t
f	$*$	t	f	t	t	f	f	f	f
f	f	t	t	t	t	t	t	t	t

Further to this, here is a list of other valid and invalid properties:

- $\vdash \mathrm{I}p \equiv p$
- $\vdash \mathrm{M}p \equiv \neg p$
- $\vdash \mathrm{I}\neg p \equiv \mathrm{M}p$
- $\vdash \mathrm{M}\neg p \equiv \mathrm{I}p$
- $\vdash \mathrm{S}p \equiv \mathrm{I}p \vee \mathrm{M}p$
- $\vdash \mathrm{I}p \equiv \mathrm{II}p$
- $\nvdash \mathrm{M}p \rightarrow \mathrm{MM}p$
- $\nvdash \mathrm{MM}p \rightarrow \mathrm{M}p$

References

[1] Urquhart, Alasdair. 'Many-Valued Logic'. In D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic (Volume III)*. D. Reidel Publishing Company, 1986, pp. 71-116.